Sequential Cooperative Bayesian Inference

Junqi Wang [1] **Pei Wang** [1] **Patrick Shafto** [1]

Abstract

Cooperation is often implicitly assumed when learning from other agents. Cooperation implies that the agent selecting the data, and the agent learning from the data, have the same goal, that the learner infer the intended hypothesis. Recent models in human and machine learning have demonstrated the possibility of cooperation. We seek foundational theoretical results for cooperative inference by Bayesian agents through sequential data. We develop novel approaches analyzing consistency, rate of convergence and stability of Sequential Cooperative Bayesian Inference (SCBI). Our analysis of the effectiveness, sample efficiency and robustness show that cooperation is not only possible in specific instances but theoretically well-founded in general. We discuss implications for human-human and human-machine cooperation.

1. Introduction

Learning often occurs sequentially, as opposed to in batch, and from data provided by other agents, as opposed to from a fixed random sampling process. The canonical example of sequential learning from an agent occurs in educational contexts where the other agent is a teacher whose goal is to help the learner. However, instances appear across a wide range of contexts including informal learning, language, and robotics. In contrast with typical contexts considered in machine learning, it is reasonable to expect the cooperative agent to adapt their sampling process after each trial, consistent with the goal of helping the learner learn more quickly. It is also reasonable to expect that learners, in dealing with such cooperative agents, would know the other agent intends to cooperate and incorporate that knowledge when updating their beliefs. In this paper, we analyze basic statistical properties of such sequential cooperative inferences.

Large behavioral and computational literatures highlight the importance cooperation for learning. Across behavioral sciences, cooperative information sharing is believed to be a core feature of human cognition. Education, where a teacher selects examples for a learner, is perhaps the most obvious case. Other examples appear in linguistic pragmatics (Frank & Goodman, 2012), in speech directed to infants (Eaves Jr et al., 2016), and children's learning from demonstrations (Bonawitz et al., 2011). Indeed, theorists have posited that the ability to select data for and learn cooperatively from others explains humans' ability to learning quickly in childhood and accumulate knowledge over generations (Tomasello, 1999; Csibra & Gergely, 2009).

Across computational literatures, cooperative information sharing is also believed to be central to human-machine interaction. Examples include pedagogic-pragmatic value alignment in robotics (Fisac et al., 2017), cooperative inverse reinforcement learning (Hadfield-Menell et al., 2016), machine teaching (Zhu, 2013), and Bayesian teaching (Eaves Jr et al., 2016) in machine learning, and Teaching dimension in learning theory (Zilles et al., 2008; Doliwa et al., 2014). Indeed, rather than building in knowledge or training on massive amounts of data, cooperative learning from humans is a strong candidate for advancing machine learning theory and improving human-machine teaming more generally.

While behavioral and computational research makes clear the importance of cooperation for learning, we lack mathematical results that would establish statistical soundness. In the development of probability theory, proofs of consistency and rate of convergence were celebrated results that put Bayesian inference on strong mathematical footing (Doob, 1949). Moreover, establishment of stability with respect to mis-specification ensured that theoretical results could apply despite the small differences between the model and reality (Kadane et al., 1978; Berger et al., 1994). Proofs of consistency, convergence, and stability ensure that intuitions regarding probabilistic inference were formalized in ways that satisfied basic desiderata.

Our goal is to provide a comparable foundation for sequential Cooperative Bayesian Inference as statistical inference for understanding the strengths, limitations, and behavior of cooperating agents. Grounded strongly in machine learning (Murphy, 2012; Ghahramani, 2015) and human learning

[1] CoDaS Lab, Department of Math & CS, Rutgers University at Newark, New Jersey, USA. Correspondence to: Junqi Wang <junqi.wang@rutgers.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020. Copyright 2020 by the author(s).

(Tenenbaum et al., 2011), we adopt a probabilistic approach. We approach consistency, convergence, and stability using a combination of new analytical and empirical methods. The result will be a model agnostic understanding of whether and under what conditions sequential cooperative interactions result in effective and efficient learning.

Notations are introduced at the end of this section. Section 2 introduces the model of sequential cooperative Bayesian inference (SCBI), and Bayesian inference (BI) as the comparison. Section 3 presents a new analysis approach which we apply to understanding consistency of SCBI. Section 4 presents empirical results analyzing the sample efficiency of SCBI versus BI, showing convergence of SCBI is considerably faster. Section 5 presents the empirical results testing robustness of SCBI to perturbations. Section 6 introduces an application of SCBI in Grid world model. Section 7 describes our contributions in the context of related work, and Section 8 discusses implications for machine learning and human learning.

Preliminaries. Throughout this paper, for a vector θ, we denote its i-th entry by θ_i or $\theta(i)$. Similarly, for a matrix $\mathbf{M}$, we denote the vector of i-th row by $\mathbf{M}_{(i,\_)}$, the vector of j-th column by $\mathbf{M}_{(\_,j)}$, and the entry of i-th row and j-th column by $\mathbf{M}_{(i,j)}$ or simply $\mathbf{M}_{ij}$. Further, let $\mathbf{r}, \mathbf{c}$ be the column vectors representing the row and column marginals (sums along row/column) of $\mathbf{M}$. Let $\mathbf{e}_n$ or simply $\mathbf{e}$ be the vector of ones. The symbol $\mathscr{N}_{\text{vec}}(\theta, s)$ is used to denote the normalization of a non-negative vector θ, i.e., $\mathscr{N}_{\text{vec}}(\theta, s) = \frac{s}{\sum \theta_i}\theta$ with $s = 1$ if absent. Similarly, the normalization of matrices are denoted by $\mathscr{N}_{\text{col}}(\mathbf{M}, \theta)$, with "col" indicating column normalization (for row normalization, write "row" instead), and θ denotes to which vector of sums the matrix is normalized. The set of probability distributions on a finite set $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$, we do not distinguish it with the simplex $\Delta^{|\mathcal{X}|-1}$. The language of statistical models and estimators follows the notations of the book (Miescke & Liese, 2008).

2. The Construction

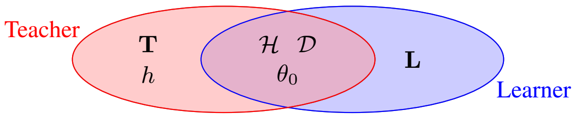

Figure 1. Two agents and their knowledge before starting.

In this paper, we consider cooperative communication models with two agents, which we call a teacher and a learner. Let $\mathcal{H} = \{1, \dots, m\}$ be the set of m hypotheses, i.e., concepts to teach. The shared **goal** is for the learner to infer the correct hypothesis $h \in \mathcal{H}$ which is only known by the teacher at the beginning. To facilitate learning, the teacher passes one element from a finite data set $\mathcal{D} = \{1, \dots, n\}$ sequentially. Each agent has knowledge about the relation between $\mathcal{H}$ and $\mathcal{D}$, in terms of a positive matrix whose normalization can be treated as the likelihood matrix in a Bayesian sense. Let $\mathbf{T}, \mathbf{L} \in \text{Mat}_{n \times m}(\mathbb{R}^+)$ be the matrices for teacher and learner, respectively.

In order to construct a Bayesian theory, the learner has an initial prior θ_0 on $\mathcal{H}$, which, along with posteriors $\theta_k (k \geq 1)$, are elements in $\mathcal{P}(\mathcal{H}) = \Delta^{m-1} := \{\theta \in \mathbb{R}^m : \sum_{i=1}^m \theta(i) = 1\}$. Privately, the teacher knows the true hypothesis $h \in \mathcal{H}$ to teach. To measure how well a posterior θ_k performs, we may view h as a distribution on $\mathcal{H}$, namely $\widehat{\theta} = \delta_h \in \mathcal{P}(\mathcal{H})$, and calculate the L^1-distance $||\theta_k - \widehat{\theta}||_1$ on $\mathcal{P}(\mathcal{H}) = \Delta^{m-1} \subseteq \mathbb{R}^m$.

We assume that $\mathcal{H}$, $\mathcal{D}$, $\mathbf{T}$, $\mathbf{L}$ and θ_0 satisfy:

(i) There are no fewer data than hypotheses ($n \geq m$).

(ii) The hypotheses are distinguishable, i.e., there is no $\lambda \in \mathbb{R}$ such that $\mathbf{T}_{(\_,i)} = \lambda \mathbf{T}_{(\_,j)}$ for any $i \neq j$, and so is $\mathbf{L}$.

(iii) $\mathbf{T}$ is a *scaled matrix* of $\mathbf{L}$, i.e., there exist invertible diagonal matrices $\mathbf{E_1}$ and $\mathbf{E_2}$ such that $\mathbf{T} = \mathbf{E_1 L E_2}$. (Both agents aware this assumption, though possibly neither know the other's matrix.)

(iv) θ_0 is known by the teacher.

Our model is constructed and studied under these assumptions (Sec. 3 and Sec. 4). We also studied stability under violations of (iii) and (iv), where we assume that $\mathbf{T}$ and teacher's knowledge θ_0^T about θ_0 is slightly different from (some scaled matrix of) $\mathbf{L}$ and θ_0 (Sec. 5). Assumption (iii) is a relaxation of the assumption of Bayesian inference that $\mathbf{T} = \mathbf{L} = \mathbf{M}$ is the likelihood matrix. Practically, we may achieve (iii) by adding to the common ground a shared matrix $\mathbf{M}$ (e.g. joint distribution on $\mathcal{D}$ and $\mathcal{H}$) and scaling it to $\mathbf{T}$ and $\mathbf{L}$. We may obtain $\mathbf{M}$ by taking the same ground model or using the same statistical data (e.g. historical statistical records). In fact, with (iii), it does not affect the process of inference whether $\mathbf{M}$ is accessible to agents.

In SCBI (see details in later this section), thanks to the property that a matrix and its scaled matrices behave the same in Sinkhorn scaling (Hershkowitz et al., 1988), the pre-processings of $\mathbf{T}$ and of $\mathbf{L}$ lead to the same results under (iii) and (iv). Thus assumption (iii) is equivalent to:

(iii') $\mathbf{T} = \mathbf{L} = \mathbf{M}$ where $\mathbf{M}$ is a column-stochastic matrix.

We assume (iii') is valid until we discuss stability.

In our setup, the teacher teaches in sequence. At each round the teacher chooses a data point from $\mathcal{D}$ by sampling according to a distribution. And the learner learns by maintaining a posterior distribution on $\mathcal{H}$ through Bayesian inference with likelihood matrices not necessarily fixed.

Formally, the teacher's job is to select a sequence of data $(d_k)_{k\in\mathbb{N}}$ by first constructing a sequence of random variables $(\mathrm{D}_k)_{k\in\mathbb{N}}$, then sampling each d_k as a realization of D_k. Each d_k is given to the learner at round k. And the learner's job is to produce a sequence of posteriors $(\theta_k)_{k\in\mathbb{N}}$ on $\mathcal{P}(\mathcal{H})$. To calculate θ_k, learner can use the matrix $\mathbf{L}$, the initial prior θ_0 which is common knowledge, and the sequence of data $(d_i)_{i\le k}$ which is visible at round k. The learner find each posterior by giving a function $S_k((d_i)_{i\le k}; \mathbf{L}, \theta_0)$ [1] for $k > 0$. We may further define $S_0(\varnothing; \mathbf{L}, \theta_0) = \theta_0$.

Since $(d_k)_{k\in\mathbb{N}}$ is generated by a sequence of random variables $(\mathrm{D}_k)_{k\in\mathbb{N}}$, the function S_k can be treated as a function taking $(\mathrm{D}_i)_{i\le k}$ as inputs and producing a random variable Θ_k as output. We call the distribution of Θ_k by $\mu_k \in \mathcal{P}(\mathcal{P}(\mathcal{H})) = \mathcal{P}(\Delta^{m-1})$. The S_k's as functions of random variables are called *estimators*.

Being a special case of the above framework, Bayesian inference dealing with sequential data is a well-studied model. However, there is no cooperation in Bayesian inference since the teaching distribution and learning likelihood are constant on time (the teacher side is typically left implicit). To introduce cooperation following cooperative inference (Yang et al., 2018), we propose Sequential Cooperative Bayesian Inference (SCBI), which is a sequential version of the cooperative inference.

2.1. Sequential Cooperative Bayesian Inference

Sequential Cooperative Bayesian Inference (SCBI) assumes that the two agents—a teacher and a learner—cooperate to facilitate learning. Prior research has formalized this cooperation (in a single-round game) as a system of two interrelated equations in which the teacher's choice of data depends on the learner's inference, and the learner's inference depends on reasoning about the teacher's choice. This prior research into such Cooperative Inference has focused on batch selection of data (Yang et al., 2018; Wang et al., 2019a), and has been shown to be formally equivalent to Sinkhorn scaling (Wang et al., 2019b). Following this principle, we propose a new *sequential* setting in which the teacher chooses data sequentially, and both agents update the likelihood at each round to optimize learning.

Cooperative Inference. Let $P_{L_0}(h)$ be the learner's prior of hypothesis $h \in \mathcal{H}$, $P_{T_0}(d)$ be the teacher's prior of selecting data $d \in \mathcal{D}$. Let $P_T(d|h)$ be the teacher's likelihood of selecting d to convey h and $P_L(h|d)$ be the learner's posterior for h given d. **Cooperative inference** is then a system of two equations shown below, with $P_L(d)$ and $P_T(h)$ the normalizing constants:

$$P_L(h|d) = \frac{P_T(d|h)\, P_{L_0}(h)}{P_L(d)},\ P_T(d|h) = \frac{P_L(h|d)\, P_{T_0}(d)}{P_T(h)}. \quad (1)$$

Algorithm 1 SCBI, without assumption (iii')

```
== Teacher's Part: ==
Input: $\mathbf{T} \in \mathrm{Mat}_{n\times m}(\mathbb{R}^+)$, $\theta_0$, $h \in \mathcal{H}$, $(\widehat{\theta} = \delta_h)$
Output: Share $(d_1, d_2, \dots)$ to learner
for all $i \ge 1$ do
    sample $d_i \sim \mathscr{N}_{\text{vec}}\left(\mathbf{T}^{\langle n\theta_{i-1}\rangle}_{(\_,h)}, 1\right)$.
    $\theta_i \leftarrow \mathbf{T}^{\langle n\theta_{i-1}\rangle}_{(d_i,\_)}$ estimation of learner's posterior
end for
== Learner's Part: ==
Input: $\mathbf{L} \in \mathrm{Mat}_{n\times m}(\mathbb{R}^+)$, $\theta_0$, $(d_1, d_2, \dots)$
Output: $(\theta_0, \theta_1, \theta_2, \dots)$ posteriors
for all $i \ge 1$ do
    $\theta_i \leftarrow \mathbf{L}^{\langle n\theta_{i-1}\rangle}_{(d_i,\_)}$
end for
```

Note: $\mathbf{T}^{\langle n\theta_{i-1}\rangle}$, $\mathbf{L}^{\langle n\theta_{i-1}\rangle}$ are the $\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}$ in the text.

It is shown (Wang et al., 2019a;b) that Eq. (1) can be solved using Sinkhorn scaling, where $(\mathbf{r}, \mathbf{c})$-**Sinkhorn scaling** of a matrix $\mathbf{M}$ is simply the iterated alternation of row normalization of $\mathbf{M}$ with respect to $\mathbf{r}$ and column normalization of $\mathbf{M}$ with respect to $\mathbf{c}$. The limit of such iterations exist if the sums of elements in $\mathbf{r}$ and $\mathbf{c}$ are the same (Schneider, 1989).

Sequential Cooperation. SCBI allows multiple rounds of teaching and requires each choice of data to be generated based on cooperative inference, with the learner updating their beliefs between each round. In each round, based on the data being taught and the learner's initial prior on $\mathcal{H}$ as common knowledge, the teacher and learner update their common likelihood matrix according to cooperative inference (using Sinkhorn scaling), then the data selection and inference proceed based on the updated likelihood matrix.

Precisely, starting from learner's prior $S_0 = \theta_0 \in \Delta^{m-1}$, let the data been taught up to round k be $(d_1, \dots, d_{k-1})$ and the posterior of the learner after round $k-1$ be $\theta_{k-1} = S_{k-1}(d_1, \dots, d_{k-1}; \theta_0) \in \mathcal{P}(\mathcal{H})$, which is actually predictable for both agents (obvious for $k = 1$ and inductively correct for $k > 1$ by later argument). To teach, the teacher calculates the Sinkhorn scaling of $\mathbf{M}$ given the uniform row sums $\mathbf{r}_{k-1} = \mathbf{e}_n = (1, 1, \dots, 1)^\top$ and column sums $\mathbf{c}_{k-1} = n\theta_{k-1}$ (to make the sum of $\mathbf{r}_{k-1}$ equal that of $\mathbf{c}_{k-1}$, which guarantees the existence of the limit in Sinkhorn scaling), denoted by $\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}$. The teacher's data selection is proportional to columns of $\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}$. Thus let $\mathbf{M}_k$ be the column normalization of $\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}$ by $\mathbf{e}_m$, i.e., $\mathbf{M}_k = \mathscr{N}_{\text{col}}\left(\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}, \mathbf{e}_m\right)$. Then the teacher defines D_k using distribution $(\mathbf{M}_k)_{(\_,h)}$ on set $\mathcal{D}$ and samples $d_k \sim \mathrm{D}_k$, then passes d_k to the learner.

On learner's side, the learner obtains the likelihood matrix $\mathbf{M}_k$ in the same way as above and applies normal Bayesian inference with datum d_k past from the teacher. First, learner

[1] we may omit $\mathbf{L}$ and (or) θ_0 when there is no ambiguity.

takes the prior to be the posterior of the last round, $\theta_{k-1} = \frac{1}{n}\mathbf{c}_{k-1}$, then multiply it by the likelihood of selecting d_k — the row of $\mathbf{M}_k$ corresponding to d_k, which results $\mathring{\theta}_k = (\mathbf{M}_k)_{(d_k,\text{-})}\text{diag}(\theta_{k-1})$. Then the posterior θ_k is obtained by row normalizing $\mathring{\theta}_k$. Inductively, in the next round, the learner will start with θ_k and $\mathbf{c}_k = n\theta_k$. The learner's calculation in round k can be simulated by the teacher, so the teacher can predict θ_k, which inductively shows the assumption (teacher knows θ_{k-1}) in previous paragraph.

The calculation can be simplified. Consider that the vector $\mathbf{c}_{k-1}$, being proportional to the prior, is used in $\mathbf{M}_k = \mathscr{N}_{\text{col}}(\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}, \mathbf{e}_m) = \mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}\left(\text{diag}(n\theta_{k-1})\right)^{-1}$, then $\mathring{\theta}_k = \left(\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}\left(\text{diag}(n\theta_{k-1})\right)^{-1}\text{diag}(\theta_{k-1})\right)_{(d_k,\text{-})}$ $= \frac{1}{n}\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}_{(d_k,\text{-})}$. Furthermore, since $\mathbf{M}^{\langle \mathbf{c}_{k-1}\rangle}$ is row normalized to $\mathbf{e}_m$, each row of it is a probability distribution on $\mathcal{H}$. Thus $S_k(d_1, \dots, d_k) = \theta_k = n\mathring{\theta}_{k-1} = \mathbf{M}^{\langle \mathbf{c}_k\rangle}_{(d_k,\text{-})}$. [2]

The simplified version of SCBI algorithm is given in Algorithm 1.

2.2. Bayesian Inference: the Control

In order to test the performance of SCBI, we recall the classical Bayesian inference (BI). In BI, a fixed likelihood matrix $\mathbf{M}$ is used throughout the communication process. Bayes' rule requires $\mathbf{M}$ to be the conditional distribution on the set of data given each hypothesis, thus $\mathbf{M} = \mathbf{T} = \mathbf{L}$ is column-stochastic as in assumption (iii').

For the teacher, given $h \in \mathcal{H}$, the teaching distribution is the column vector $P_h = \mathbf{M}_{(\text{-},h)} \in \mathcal{P}(\mathcal{D})$. This defines random variable D_k. Then the teacher selects data via i.i.d. sampling according to P_h. The random variables $(\mathrm{D}_k)_{k\geq 1}$ are identical.

The learner first chooses a prior $\theta_0 \in \mathcal{P}(\mathcal{H})$ ($\theta_0 = S_0$ is part of the model, usually the uniform distribution), then uses Bayes' rule with likelihood $\mathbf{M}$ to update the posterior distribution repeatedly. Given taught datum d, the map from the prior θ to the posterior distribution is denoted by $B_d(\theta) = \mathscr{N}_{\text{vec}}\left(\mathbf{M}_{(d,\text{-})}\text{diag}(\theta), 1\right)$. Thus the learner's estimation over $\mathcal{H}$ given a sequential data $(d_1, \dots, d_k)$ can be written recursively by $S_0 = \theta_0$, and $S_k(d_1, \dots, d_k) = B_{d_k}(S_{k-1}(d_1, \dots, d_{k-1}))$. Thus, by induction, $S_k(d_1, \dots, d_k) = (B_{d_k} \circ B_{d_{k-1}} \circ \dots \circ B_{d_1})(S_0)$.

3. Consistency

We investigate the effectiveness of the estimators in both BI and SCBI by testing their *consistency*: setting the true hypothesis $h \in \mathcal{H}$, given (D_k), (S_k) and θ_0, we examine the convergence (using the L^1-distance on $\mathcal{P}(\mathcal{H})$) of the posterior sequence $(\Theta_k) = (S_k(\mathrm{D}_1, \dots, \mathrm{D}_k))$ as sequence of random variables and check whether the limit is $\widehat{\theta}$ as a constant random variable.

3.1. BI and KL Divergence

The consistency of BI has been well studied since Bernstein and von Mises and Doob (Doob, 1949). In this section, we state it in our situation and derive a formula for the rate of convergence, as a baseline for the cooperative theory. Derivations and proofs can be found in the Supplementary Material.

Theorem 3.1. *[(Miescke & Liese, 2008, Theorem 7.115)] In BI, the sequence of posteriors (S_k) is strongly consistent at $\widehat{\theta} = \delta_h$ for each $h \in \mathcal{H}$, with arbitrary choice of an interior point $\theta_0 \in (\mathcal{P}(\mathcal{H}))^\circ$ (i.e. $\theta_0(h) > 0$ for all $h \in \mathcal{H}$) as prior.*

Remark 1. For a fixed true distribution $\widehat{\theta}$, *strong consistency* of $(S_k)_{k\in\mathbb{N}}$ is defined to be: the sequence of posteriors Θ_k given by the estimator S_k, as a sequence of random variables, converges to $\widehat{\theta}$ (as a constant random variable) almost surely according to random variables $(\mathrm{D}_k)_{k\in\mathbb{N}}$ that the teacher samples from. If the convergence is in probability, the sequence of estimators is said to be *consistent*.

Remark 2. Theorem 3.1 also assumes that hypotheses are distinguishable (Section 2). In a general theory of statistical models, $\widehat{\theta}$ is not necessarily δ_h for some $h \in \mathcal{H}$. However, in BI, it is critical to have $\widehat{\theta} = \delta_h$, since BI with a general $\widehat{\theta} \in \mathcal{P}(\mathcal{H})$ is almost never consistent or strongly consistent.

Consistency—independent of the choice of prior θ_0 interior of $\mathcal{P}(\mathcal{H})$—guarantees that BI is always effective.

Rate of Convergence. After effectiveness, we provide the efficiency of BI in terms of asymptotic rate of convergence.

Theorem 3.2. *In BI, with $\widehat{\theta} = \delta_h$ for some $h \in \mathcal{H}$, let $\Theta_k(h)(\mathrm{D}_1, \dots, \mathrm{D}_k) := S_k(h|\mathrm{D}_1, \dots, \mathrm{D}_k)$ be the h-component of posterior given $\mathrm{D}_1, \dots, \mathrm{D}_k$ as random variables valued in $\mathcal{D}$. Then $\dfrac{1}{k}\log\left(\dfrac{\Theta_k(h)}{1-\Theta_k(h)}\right)$ converges to a constant $\min_{h'\neq h}\left\{\text{KL}(\mathbf{M}_{(\text{-},h)}, \mathbf{M}_{(\text{-},h')})\right\}$ almost surely.*

Remark 3. We call $\min_{h'\neq h}\left\{\text{KL}(\mathbf{M}_{(\text{-},h)}, \mathbf{M}_{(\text{-},h')})\right\}$ the *asymptotic rate of convergence* (RoC) of BI, denoted by $\mathfrak{R}^{\text{b}}(\mathbf{M}; h)$.

3.2. SCBI as a Markov Chain

From the proof of Theorem 3.1, the pivotal property is that the variables $\mathrm{D}_1, \mathrm{D}_2, \dots$ are commutative in posteriors (the variables can occur in any order without affecting the posterior) thanks to commutativity of multiplication. However, in SCBI, the commutativity does not hold, since the likelihood matrix depends on previous outcome. Thus the method used in BI analysis no longer works here.

[2] See Supplementary Material for detailed examples.

Because the likelihood matrix $\mathbf{M}_k = \mathbf{M}^{\langle \mathbf{c}_{k-1} \rangle}$ depends on the predecessive state only, the process is in fact Markov, we may analyze the model as a Markov chain on the continuous state space $\mathcal{P}(\mathcal{H})$.

To describe this process, let $\mathcal{P}(\mathcal{H}) = \Delta^{m-1}$ be the space of states, and let $h \in \mathcal{H}$ be the true hypothesis to teach ($\widehat{\theta} = \delta_h$), let learner's prior be $S_0 = \theta_0 \in \mathcal{P}(\mathcal{H})$, or say, the distribution of learner's initial state is $\mu_0 = \delta_{\theta_0} \in \mathcal{P}(\mathcal{P}(\mathcal{H}))$.

The operator Ψ. In the Markov chain, in each round, the transition operator maps the prior as a probability distribution on state space $\mathcal{P}(\mathcal{H}) = \Delta^{m-1}$ to the posterior as another, i.e., $\Psi(h) : \mathcal{P}(\mathcal{P}(\mathcal{H})) \to \mathcal{P}(\mathcal{P}(\mathcal{H}))$.

To make the formal definition of $\Psi(h)$ simpler, we need to define some maps. For any $d \in \mathcal{D}$, let $T_d : \Delta^{m-1} \to \Delta^{m-1}$ be the map bringing the learner's prior to posterior when data d is chosen by the teacher, that is, T_d sends each normalized vector θ to $T_d(\theta) = \mathbf{M}^{\langle n\theta \rangle}_{(d,\text{-})}$ according to SCBI. Each T_d is a bijection based on the uniqueness of Sinkhorn scaling limits of $\mathbf{M}$, shown in (Hershkowitz et al., 1988). Further, the map T_d is continuous on Δ^{m-1} and smooth in its interior according to (Wang et al., 2019b). Continuity and smoothness of T_d make it natural to induce a push-forward $T_{d*} : \mathcal{P}(\Delta^{m-1}) \to \mathcal{P}(\Delta^{m-1})$ on Borel measures. Explicitly, $(T_{d*}(\mu))(E) = \mu(T_d^{-1}(E))$ for each Borel measure $\mu \in \mathcal{P}(\Delta^{m-1})$ and each Borel measurable set $E \subseteq \Delta^{m-1}$. Let $\tau : \mathcal{P}(\mathcal{H}) \to \mathcal{P}(\mathcal{D})$ be the map of teacher's adjusting sample distribution based on the learner's prior, that is, given a learner's prior $\theta \in \Delta^{m-1}$, by definition of SCBI, the distribution of the teacher is adjusted to $\tau(\theta) = \frac{\mathbf{M}^{\langle n\theta \rangle}_{(\text{-},h)}}{n\theta(h)} = (n\theta(h))^{-1}(T_1(\theta)(h), \ldots, T_n(\theta)(h))$. Each component d of τ is denoted by τ_d. We can use τ only for $\theta_0 = \delta_h$ in which case teacher can trace learner's state. Now we can define $\Psi(h)$ formally.

Definition 3.3. Given a fixed hypothesis $h \in \mathcal{H}$, or say $\delta_h \in \mathcal{P}(\mathcal{H})$, the operator $\Psi(h) : \mathcal{P}(\Delta^{m-1}) \to \mathcal{P}(\Delta^{m-1})$ translating a prior as a Borel measure μ to the posterior distribution $\Psi(h)(\mu)$ according to one round of SCBI is given below, for any Borel measurable set $E \subset \Delta^{m-1}$.

$$\left(\Psi(h)(\mu)\right)(E) := \int_E \sum_{d \in \mathcal{D}} \tau_d(T_d^{-1}(\theta)) \mathrm{d}\left(T_{d*}(\mu)\right)(\theta). \qquad (2)$$

In our case, we start with a distribution δ_θ where $\theta \in \mathcal{P}(\mathcal{H})$ is the prior of the learner on the set of hypotheses. In each round of inference, there are n different possibilities according to the data taught. Thus in any finite round k, the distribution of the posterior is the sum of at most n^k atoms (actually, we can prove n^k is exact). Thus in the following discussions, we assume that μ is atomic. The Ψ action on an atomic distribution is determined by that of an atom:

$$\Psi(h)(\delta_\theta) = \sum_{i=1}^{n} \frac{\mathbf{M}^{\langle n\theta \rangle}_{(i,h)}}{n\theta(h)} \delta_{\left(\mathbf{M}^{\langle n\theta \rangle}_{(i,\text{-})}\right)}. \qquad (3)$$

Moreover, since the SCBI behavior depends only on the prior (with fixed $\mathbf{M}$ and h) as a random variable, the same operator $\Psi(h)$ applies to every round in SCBI. Thus we can conclude that the following proposition is valid:

Proposition 3.4. *Given $h \in \mathcal{H}$, let $\widehat{\theta} = \delta_h$, the sequence of estimators $(S_k)_{k \in \mathbb{N}}$ in SCBI forms a time-homogeneous Markov chain on state space $\mathcal{P}(\mathcal{H})$ with transition operator $\Psi(h)$ characterized by Eq. (2) and Eq. (3).*

Thanks to the fact that the SCBI is a time homogeneous Markov process, we can further show the consistency.

Theorem 3.5 (Consistency)**.** *In SCBI, let $\mathbf{M}$ be a positive matrix. If the teacher is teaching one hypothesis h (i.e., $\widehat{\theta} = \delta_h \in \mathcal{P}(\mathcal{H})$), and the prior distribution $\mu_0 \in \mathcal{P}(\Delta^{m-1})$ satisfies $\mu_0 = \delta_{\theta_0}$ with $\theta_0(h) > 0$, then the estimator sequence (S_k) is consistent, for each $h \in \mathcal{H}$, i.e., the posterior random variables $(\Theta_k)_{k \in \mathbb{N}}$ converge to the constant random variable $\widehat{\theta}$ in probability.*

Remark 4. The assumption in Theorem 3.5 that $\theta_0(h) > 0$ is necessary in any type of Bayesian inference since it is impossible to get the correct answer in posterior by Bayes' rule, if it is excluded in the prior at the beginning. In practice, the prior distribution is usually chosen to be $\mu_0 = \delta_{\mathbf{u}}$ with the uniform distribution vector in $\mathcal{P}(\mathcal{H})$, i.e., $\mathbf{u} = \frac{1}{m}(1, \ldots, 1)^\top \in \Delta^{m-1}$.

Rate of Convergence. Thanks to consistency, we can calculate the asymptotic rate of convergence for SCBI.

Theorem 3.6. *With matrix $\mathbf{M}$, hypothesis $h \in \mathcal{H}$, and a prior $\mu_0 = \delta_{\theta_0} \in \mathcal{P}(\Delta^{m-1})$ same as in Theorem. 3.5, let θ_k denote a sample value of the posterior Θ_k after k rounds of SCBI, then*

$$\lim_{k \to \infty} \mathbb{E}_{\mu_k}\left[\frac{1}{k}\log\left(\frac{\theta_k(h)}{1 - \theta_k(h)}\right)\right] = \mathfrak{R}^{\mathrm{s}}(\mathbf{M}; h) \qquad (4)$$

where $\mathfrak{R}^{\mathrm{s}}(\mathbf{M}; h) := \min_{h \neq h'} \mathrm{KL}\left(\mathbf{M}^{\sharp}_{(\text{-},h)}, \mathbf{M}^{\sharp}_{(\text{-},h')}\right)$ with $\mathbf{M}^{\sharp} = \mathscr{N}_{col}(\mathrm{diag}(\mathbf{M}_{(\text{-},h)})^{-1}\mathbf{M})$. Thus we call $\mathfrak{R}^{\mathrm{s}}(\mathbf{M}; h)$ the asymptotic rate of convergence (RoC) of SCBI.

4. Sample Efficiency

In this section, we present some empirical results comparing the sample efficiency of SCBI and BI.

4.1. Asymptotic RoC Comparison

We first compare the asymptotic rate of convergence ($\mathfrak{R}^{\mathrm{b}}$ for BI and $\mathfrak{R}^{\mathrm{s}}$ for SCBI, see Theorems 3.2 and 3.6). The matrix $\mathbf{M}$ is sampled through m i.i.d. uniform distributions on Δ^{n-1}, one for each column.

For each column-normalized matrix $\mathbf{M}$, we compute two variables to compare BI with SCBI: the probability $\mathfrak{P} := \Pr\left(\frac{1}{m}\sum_{h \in \mathcal{H}} \mathfrak{R}^{\mathrm{s}}(\mathbf{M}; h) \geq \frac{1}{m}\sum_{h \in \mathcal{H}} \mathfrak{R}^{\mathrm{b}}(\mathbf{M}; h)\right)$

and the expected value of averaged difference $\mathfrak{E} := \mathbb{E}\left[\frac{1}{m}\sum_{h\in\mathcal{H}}\mathfrak{R}^{\mathrm{s}}(\mathbf{M};h) - \frac{1}{m}\sum_{h\in\mathcal{H}}\mathfrak{R}^{\mathrm{b}}(\mathbf{M};h)\right]$.

Two-column Cases. Consider the case where $\mathbf{M}$ is of shape $n \times 2$ with the two columns sampled from Δ^{n-1} uniformly and independently, we simulated for $n = 2, 3, \ldots, 50$ with a size-10^{10} Monte Carlo method for each n to calculate $\mathfrak{P}$ and $\mathfrak{E}$. The result is shown in Fig. 2(A)(B).

We can reduce the calculation of $\mathfrak{E}$ to a numerical integral $\mathfrak{E} = \int_{(\Delta^{n-1})^2} \ln\left(\sum_{i=1}^{n}\frac{\mathbf{x}_i}{\mathbf{y}_i}\right)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} - \ln n - \frac{n-1}{n}$. [3]

Since $\mathfrak{P}$ goes too close to 1 as the rank grows, we use $-\ln(1-\mathfrak{P})$ to show the increasing in detail. [4]

More Columns of a Fixed Row Size. To verify the general cases, we simulated $\mathfrak{P}$ and $\mathfrak{E}$ by Monte Carlo on matrices of 10-row and various-column shapes, see Fig. 2(C)(D). We sampled 10^8 different $\mathbf{M}$ of shape $10 \times m$ for each $2 \le m \le 10$. Empirical results show that $\mathfrak{E}$ decreases slowly but $\mathfrak{P}$ still increase logistically as m grows.

Square Matrices. Fig. 2(E)(F) shows the square cases with size from 2 to 50, simulated by size 10^8 Monte Carlo.

The empirical $\mathfrak{P}$ is the mean of N (sample-size) i.i.d. variables valued 0 or 1, thus the standard deviation of a single variable is smaller than 1. By Central Limit Theorem, the standard deviation $\sigma(\mathfrak{P}) < N^{-1/2}$ (precision threshold). So we draw lines $y = N^{-1/2}$ in each log-figure, but only in one figure the line lies in the view area.

In all simulated cases, we observe that $\mathfrak{E} > 0$ and $\mathfrak{P} > 0.5$, indicating that SCBI converges faster than BI in most cases and in average. It is also observed that SCBI behaves even better as matrix size grows, especially when the teacher has more choices on the data to be chosen (i.e., more rows).

4.2. Distribution of Inference Results

The promises of cooperation is that one may infer hypotheses from small amounts of data. Hence, we compare SCBI with BI after small, fixed numbers of rounds.

We sample matrices of shape 20×20 whose columns are distributed evenly in Δ^{19} to demonstrate. Equivalently, they are column-normalizations of the uniformly sampled matrices whose sum of all entries is one.

Assume that the correct hypothesis to teach is $h \in \mathcal{P}(\mathcal{H})$ We first simulate a 5-round inference behavior, exploring all possible ways that the teacher may teach, then calculate the expectation and standard deviation of $\theta(h)$. With 300 matrices sampled in the above way, Fig. 3 shows this comparison between BI and SCBI.

Similarly, we extend the number of rounds to 30 by Monte Carlo since an exact calculation on exhausting all possible teaching paths becomes impossible. With sampling 500 matrices independently, we simulate a teacher teaches 2000 times to round 30 for each matrix, and the statistics are also shown in Fig. 3. From Fig. 3, we observe that SCBI have better expectation and less variance in the short run.

In conclusion, experiments indicate that SCBI is both more efficient asymptotically, and in the short run.

5. Stability

In this section, we study the robustness of SCBI by setting the initial conditions of teacher and learner different. This could happen when agents do not have full access to their partner's exact state.

Theory. In this section, we no longer have assumption (iii). Let $\mathbf{T}$ and $\mathbf{L}$ be matrices of teacher and learner (not necessarily have (iii)). Let θ_0^T and θ_0^L be elements in $\mathcal{P}(\mathcal{H})$ representing the prior on hypotheses that the teacher and learner use in the estimation of inference (teacher) and in the actual inference (learner), i.e., $\mu_0^T = \delta_{\theta_0^T}$ and $\mu_0^L = \delta_{\theta_0^L}$. During the inference, let μ_k^T and μ_k^L be the distribution of posteriors of the teacher and the learner after round k, and denote the corresponding random variables by θ_k^T and θ_k^L, for all positive k and ∞, where ∞ represents the limit in probability.

Let D be a random variable on $\mathcal{D}$, we define an operator $\Psi_{\mathrm{D}}^{\mathbf{L}} : \mathcal{P}(\mathcal{P}(\mathcal{H})) \longrightarrow \mathcal{P}(\mathcal{P}(\mathcal{H}))$ similar to the Ψ in Section 3. Let $T_d(\theta) = \mathbf{L}_{(d,\text{-})}^{\langle n\theta\rangle}$, then $\mathrm{d}(\Psi_{\mathrm{D}}^{\mathbf{L}}(\mu))(\theta) := \sum\limits_{d\in\mathcal{D}} \mathrm{P}(\mathrm{D}=d)\mathrm{d}(T_{d*}\mu)(\theta)$.

Proposition 5.1. *Given a sequence of identical independent $\mathcal{D}$-valued random variables $(\mathrm{D}_i)_{i\ge1}$ following the uniform distribution. Let $\mu_0 \in \mathcal{P}(\mathcal{P}(\mathcal{H}))$ be a prior distribution on $\mathcal{P}(\mathcal{H})$, and $\mu_{k+1} = \Psi_{\mathrm{D}_{k+1}}^{\mathbf{L}}(\mu_k)$, then μ_k converges, in probability, to $\sum_{i\in\mathcal{H}} a_i\delta_i$ where $a_i = \mathbb{E}_{\mu_0}[\theta(i)]$.*

Remark 5. This proposition helps accelerate the simulation, that one may terminate the teaching process when θ_k^T is sufficiently close to δ_h, since Prop. 5.1 guarantees that the expectation of the learner's posterior on the true hypothesis h at that time is close enough to the eventual probability of getting δ_h, i.e. $\mathbb{E}\theta_\infty^L(h) \approx \mathbb{E}\theta_k^L(h)$.

Definition 5.2. We call $\mathbb{E}\theta_\infty^L(h) := \lim\limits_{k\to\infty}\mathbb{E}_{\mu_k}(\theta(h))$ the **successful rate** of the inference given $\mathbf{T}$, $\mathbf{L}$, θ_0^T and θ_0^L. By the setup in Section 2, the failure probability, $1 - \mathbb{E}\theta_\infty^L(h)$, is $\frac{1}{2}||\mathbb{E}\theta_\infty^L - \delta_h||_1$, half of the 1-distance on $\mathcal{P}(\mathcal{H})$.

Simulations with Perturbation on Priors. We simulated the square cases of rank 3 and 4. We sample 5 matrices ($\mathbf{M}_1$ to $\mathbf{M}_5$) of size 3×3, whose columns distribute uniformly on $\mathcal{P}(\{d_1, d_2, d_3\}) = \Delta^2$, and 5 priors ($\theta_1$ to θ_5) in $\mathcal{P}(\mathcal{H})$,

[3] Details can be found in Supplementary Material.

[4] We guess an empirical formula $-\ln(1-\mathfrak{P}) \approx \frac{1}{2}\ln(x(x+1)/(x-1.5)) + 0.1x - 0.3$, see Supplementary Material.

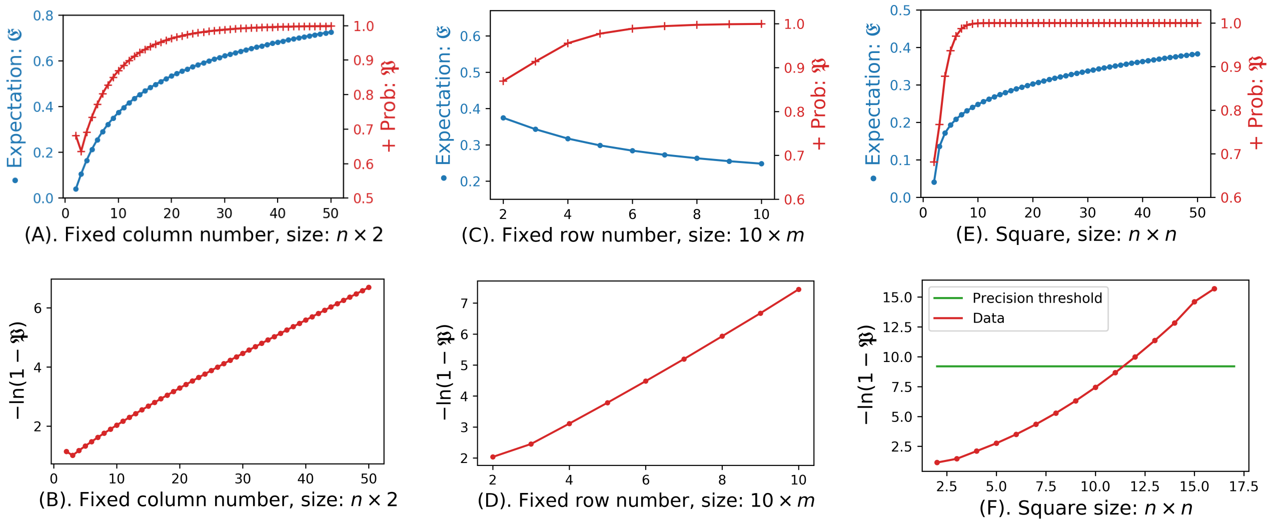

Figure 2. Comparison of RoC between BI and SCBI. (A), (C), (E): the comparison on $\mathfrak{P}$ in blue and on $\mathfrak{E}$ in red. (B), (D), (F): plotting $-\ln(1-\mathfrak{P})$. (A), (B): two-column case, number of rows from 2 to 50. Monte Carlo of 10^{10} samples for each point on figure. (C), (D): 10-row case, number of columns from 2 to 10. Monte Carlo of size 10^8. (E), (F): square case, number of rows from 2 to 50. Monte Carlo of size 10^8. The horizontal line in (F) is the theoretical threshold of precision by central limit theorem. For $n > 17$, MC provides $\mathfrak{P} = 1$ ($\mathfrak{R}^{\mathrm{s}} > \mathfrak{R}^{\mathrm{b}}$ for all samples). From the figures, except in (C) where $\mathfrak{E}$ decays slowly when column number grows, the two values $\mathfrak{E}$ and $\mathfrak{P}$ increases as size grows in all the other cases. Moreover, $\mathfrak{P}$ grows to 1 logistically in all situations.

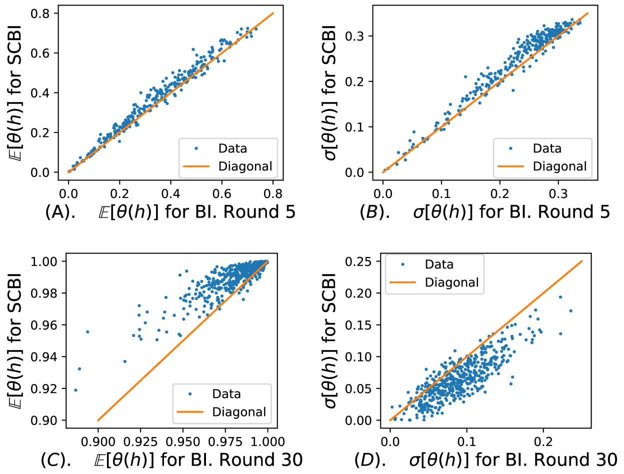

Figure 3. Comparison between BI and SCBI on 20×20 matrices: Top: 300 points (matrices) of round 5 accurate value. Bottom: 500 points of round 30 using Monte Carlo of size 2000. Left: comparison on expectations of learner's posterior on h. Right: comparison on the standard deviations. Orange line is the diagonal.

used as θ_0^T. Similarly, we sample 3 matrices ($\mathbf{M}_1'$, $\mathbf{M}_2'$ and $\mathbf{M}_3'$) of size 4×4, and 3 priors (θ_1', θ_2', θ_3') from Δ^3 in the same way as above. In both cases, we assume $h = 1 \in \mathcal{H}$ to be the true hypothesis to teach.

Our simulation is based on Monte Carlo method of 10^4 teaching sequences (for each single point plotted) then use Proposition 5.1 to calculate the successful rate of inference. For 3×3 matrices, we perturb θ_0^L in two ways: (1) take θ_0^L around θ_0^T distributed evenly on concentric circles, thus 630 points for each θ_0^T are taken. In this area, there are 84 points lying on 6 given directions (60° apart, see Supplementary Material for figures). (2) sample θ_0^L evenly in the whole simplex $\mathcal{P}(\mathcal{H}) = \Delta^2$ (300 points for each θ_0^T). For 4×4 matrices, we perturb θ_0^L in two ways: (1) along 15 randomly chosen directions in Δ^3 evenly take 21 points on each direction, and (2) sample 300 points evenly in Δ^3. Then we have the following figure samples (for figures demonstrating the entire simulation, please see Supplementary Material). From the figures we see: 1. left pictures indicate that the learner's expected posterior on h is roughly linear to perturbations along a line. 2. right pictures indicate that the learner's expected posterior on h is closely bounded by a multiple of the learner's prior on true h. Thus we have the following conjecture:

Conjecture 5.3. *Given $\mathbf{L} = \mathbf{T} = \mathbf{M}$ and θ_0^T, let h be the true hypothesis to teach. For any $\epsilon > 0$, let θ_0^L be learner's prior with a distance to θ_0^T less than ϵ. Then the successful rate for sufficiently many rounds is greater than $1 - C\epsilon$, where $C = \frac{1}{\theta_0^T(h)}$.*

Simulations with Perturbation on Matrices. We now investigate robustness of SCBI to differences between agents' matrices. Let $\mathbf{T}$ and $\mathbf{L}$ be stochastic, and let $\mathbf{L}$ be perturbed from $\mathbf{T}$. The simulations are performed on the matrices $\mathbf{M}_1$ to $\mathbf{M}_5$ mentioned above with a fixed common prior θ_1.

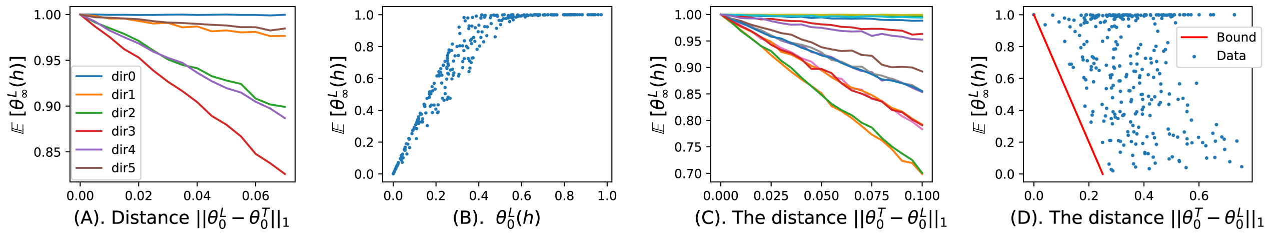

Figure 4. From left to right: (A). Rank 3, $\mathbf{M}_3$ and θ_1, θ_0^L is perturbed along six directions. (B). Rank 3, $\mathbf{M}_3$ and θ_1, sample θ_0^L uniformly in Δ^2. (C). Rank 4, $\mathbf{M}_1'$ and θ_1', along 15 different directions. (D). Rank 4, $\mathbf{M}_1'$ and θ_1', sample θ_0^L uniformly in Δ^3.

Let all matrices mentioned be column-normalized (this does not affect SCBI since cross-ratios and marginal conditions determines the Sinkhorn scaling results), we call the column determined by the true hypothesis h (the first column in our simulation) the target column ("tr. h" on Fig. 5), the column which $\mathfrak{R}^{\mathrm{s}}$ uses (argmin column) the relevant column ("rel. h") and the other column the irrelevant column ("irr. h"). Let $\mathbf{T}$ be given, and let $\mathbf{L}$ be obtained from $\mathbf{T}$ by perturbing along the relevant / irrelevant column.

Without loss of generality, we assume that only one column of the learner's matrix $\mathbf{L}$ is perturbed at a time as other perturbations may be treated as compositions of such.

For each $\mathbf{T}$ and each column h', we apply 330 perturbations on concentric circles around $\mathbf{T}$ (the disc), 90 perturbations preserving the normalized-KL divergence $(\mathrm{KL}(\mathbf{e}/n, \mathscr{N}_{\mathrm{vec}}(\mathbf{L}_{(\_,h')}/\mathbf{L}_{(\_,1)}), 1))$ used in $\mathfrak{R}^{\mathrm{s}}$) from the target column and 50 linear interpolations with target column. Each point in Fig. 5 is estimated using a size-10^4 Monte Carlo method using Proposition 5.1. From the graphs, we can see that the successful rate varies continuously on perturbations, slow on one direction (the yellow strip crossing the center) and rapid on the perpendicular direction (color changed to blue rapidly).

6. Grid World: an Application

Consider a 3×5 grid world with two possible terminal goals, A and B, and a starting position S as shown below. Let the reward at the terminal position h_t be R. Assuming no step costs, the value of a grid that distanced k from h_t is then $R \times \gamma^k$ (in the RL-sense), where $\gamma < 1$ is the discount factor.

A						B
			⇑			
		⇐	S	⇒		

Suppose the structure of the world is accessible to both agents whereas the true location of the goal h_t is only known to a teacher. The teacher performs a sequence of actions to teach h_t to a learner. At each round, there are three available actions, *left*, *up* and *right*. After observing the teacher's actions, the learner updates their belief on h_t accordingly.

We now compare BI and SCBI agents' behaviours under this grid world. In terms of previous notations, the hypothesis set $\mathcal{H} = \{A, B\}$, the data set $\mathcal{D} = \{\textit{left}, \textit{up}, \textit{right}\}$. Let the learner's prior over $\mathcal{H}$ be $\theta_0 = (0.5, 0.5)$ and the true hypothesis h_t be A, then at each blue grid, agents' (unnormalized) initial matrix $\mathbf{M} = \begin{array}{c} \\ \text{left} \\ \text{up} \\ \text{right} \end{array} \!\!\begin{array}{c} \begin{array}{cc} A & \quad\quad B \end{array} \\ \begin{pmatrix} \gamma^{(k-1)} & \gamma^{(k+1)} \\ \gamma^{(k-1)} & \gamma^{(k-1)} \\ \gamma^{(k+1)} & \gamma^{(k-1)} \end{pmatrix} \end{array}$.

Assume both BI teacher and SCBI teacher start with grid S. Based on $\mathbf{M}$, the BI teacher would choose equally between *left* and *up*, whereas the SCBI teacher is more likely to choose *left* as the teacher's likelihood matrix $\mathbf{T} = \begin{pmatrix} 2/(3+3\gamma^2) & 2\gamma^2/(3+3\gamma^2) \\ 1/3 & 1/3 \\ 2\gamma^2/(3+3\gamma^2) & 2/(3+3\gamma^2) \end{pmatrix}$, obtained from Sinkhorn scaling on $\mathbf{M}$, assigns higher probability for *left*. Hence, comparing to the BI teacher who only aims for the final goal, the SCBI teacher tends to cooperate with the learner by selecting less ambiguous moves towards the goal. This point is aligned with the core idea of many existing models of cooperation in cognitive development (Jara-Ettinger et al., 2016; Bridgers et al., in press), pragmatic reasoning (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013) and robotics (Ho et al., 2016; Fisac et al., 2017).

Moreover, even under the same teaching data, the SCBI learner is more likely to infer h_t than the BI learner. For instance, given the teacher's trajectory $\{\textit{left}, \textit{up}\}$, the left plot in Fig. 6 shows the SCBI and BI learners' posteriors on the true hypothesis h_t. Hence, comparing to the BI learner who reads the teacher's action literally, the SCBI learner interprets teacher's data corporately by updating belief sequentially after each round.

Regarding the stability, consider the case where the learner's discount factor is either greater or less (with equal probability) than the teacher's by 0.1. The right plot in Fig. 6 illustrates the expected difference between the learner's posterior on h_t after observing a teacher's trajectory of length 2 and the teacher's estimation of it.

As discussed in Sec 4.1, showing in Fig. 2, as the board gets wider and the number of possible goals gets more (i.e. the number of hypotheses increases), the gap between

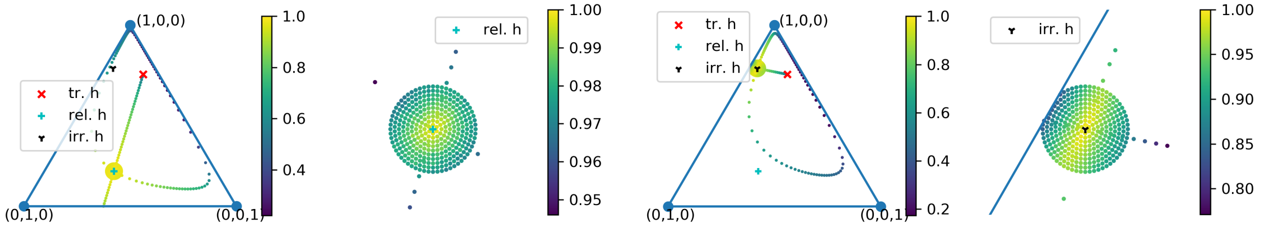

Figure 5. The perturbations on $\mathbf{M}_3$ along a column, and their zoomed-in version (with different color scale). The crosses shows the position of three normalized columns of $\mathbf{T} = \mathbf{M}_3$, the location of the dots represent the perturbed column of $\mathbf{T}$ (unperturbed columns are represented by crosses on figures which are not the center of disc) and whereas their colors depict the successful rate of inference. Left two figures are perturbations on the irrelevant column. Right two figures are perturbations on the relevant column.

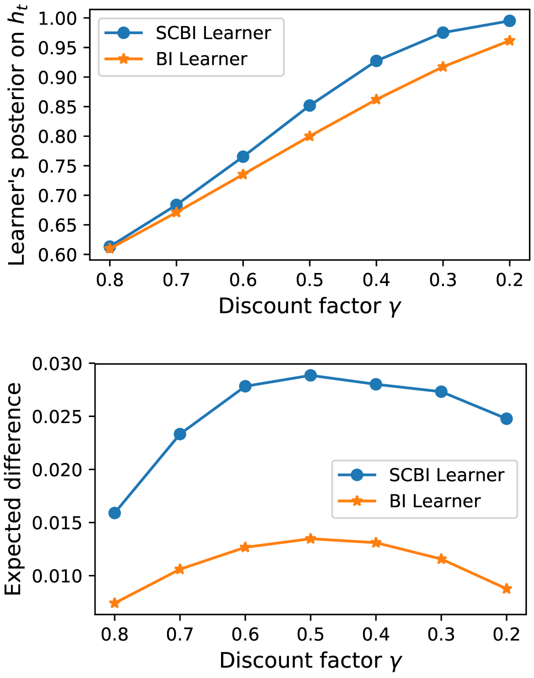

Figure 6. The top plot demonstrates that both BI and SCBI converge to the true hypothesis with SCBI having higher sample efficiency. The bottom plot shows that both BI and SCBI agents are robust to perturbations with SCBI relatively less stable.

posteriors of SCBI and BI learners will increase whereas the expected difference between agents for the same magnitude of perturbation will decrease. Thus, this example illustrates the consistency, sample efficiency, and stability of SCBI versus BI.

7. Related Work

Literatures on Bayesian teaching (Eaves & Shafto, 2016; Eaves Jr et al., 2016), Rational Speech act theory (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013), and machine teaching (Zhu, 2015; 2013) consider the problem of selecting examples that improve a learner's chances of inferring a concept. These literatures differ in that they consider the single step, rather than sequential problem, that they do not formalize learners who reason about the teacher's selection process, and that they models without a mathematical analysis.

The literature on pedagogical reasoning in human learning (Shafto & Goodman, 2008; Shafto et al., 2012; 2014) and cooperative inference (Yang et al., 2018; Wang et al., 2019a;b) in machine learning formalize full recursive reasoning from the perspectives of both the teacher and the learner. These only consider the problem of a single interaction between the teacher and learner.

The literature on curriculum learning considers sequential interactions with a learner by a teacher in which the teacher presents data in an ordered sequence (Bengio et al., 2009), and traces back to various literatures on human and animal learning (Skinner, 1958; Elman, 1993). Curriculum learning involves one of a number of methods for optimizing the sequence of data presented to the learner, most commonly starting with easier / simpler examples first and gradually moving toward more complex or less typical examples. Curriculum learning considers only problems where the teacher optimizes the sequence of examples, where the learner does not reason about the teaching.

8. Conclusions

Cooperation is central to learning in humans and machines. We set out to provide a mathematical foundation for sequential cooperative Bayesian inference (SCBI). We presented new analytic results demonstrating the consistency and asymptotic rate of convergence of SCBI. Empirically, we demonstrated the sample efficiency and stability to perturbations as compared to Bayesian inference, and illustrated with a simple reinforcement learning problem. We therefore provide strong evidence that SCBI satisfies basic desiderata. Future work will aim to provide mathematical proofs of the empirically observed efficiency and stability.

Acknowledgements

This material is based on research sponsored by the Air Force Research Laboratory and DARPA under agreement number FA8750-17-2-0146 and the Army Research Office and DARPA under agreement HR00112020039. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This work was also supported by DoD grant 72531RTREP and NSF MRI 1828528 to PS.

References

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48. ACM, 2009.

Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., et al. An overview of robust bayesian analysis. *Test*, 3(1):5–124, 1994.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., and Schulz, L. The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, 120(3):322–330, 2011.

Bridgers, S., Jara-Ettinger, J., and Gweon, H. Young children consider the expected utility of others' learning to decide what to teachn. *Nature Human Behaviour*, in press.

Csibra, G. and Gergely, G. Natural pedagogy. *Trends in cognitive sciences*, 13(4):148–153, 2009.

Doliwa, T., Fan, G., Simon, H. U., and Zilles, S. Recursive teaching dimension, VC-dimension and sample compression. *Journal of Machine Learning Research*, 15(1): 3107–3131, 2014.

Doob, J. L. Application of the theory of martingales. *Le calcul des probabilites et ses applications*, pp. 23–27, 1949.

Eaves, B. S. and Shafto, P. Parameterizing developmental changes in epistemic trust. *Psychonomic Bulletin & Review*, pp. 1–30, 2016.

Eaves Jr, B. S., Feldman, N. H., Griffiths, T. L., and Shafto, P. Infant-directed speech is consistent with teaching. *Psychological review*, 123(6):758, 2016.

Elman, J. L. Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1):71–99, 1993.

Fisac, J. F., Gates, M. A., Hamrick, J. B., Liu, C., Hadfield-Menell, D., Palaniappan, M., Malik, D., Sastry, S. S., Griffiths, T. L., and Dragan, A. D. Pragmatic-pedagogic value alignment. *arXiv preprint arXiv:1707.06354*, 2017.

Frank, M. C. and Goodman, N. D. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012.

Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452, 2015.

Goodman, N. D. and Stuhlmüller, A. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1):173–184, 2013.

Hadfield-Menell, D., Russell, S. J., Abbeel, P., and Dragan, A. Cooperative inverse reinforcement learning. In *Advances in neural information processing systems*, pp. 3909–3917, 2016.

Hershkowitz, D., Rothblum, U. G., and Schneider, H. Classifications of nonnegative matrices using diagonal equivalence. *SIAM journal on Matrix Analysis and Applications*, 9(4):455–460, 1988.

Ho, M. K., Littman, M., MacGlashan, J., Cushman, F., and Austerweil, J. L. Showing versus doing: Teaching by demonstration. In *Advances in Neural Information Processing Systems*, pp. 3027–3035, 2016.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., and Tenenbaum, J. B. The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences*, 20(8):589–604, 2016.

Kadane, J. B., Chuang, D. T., et al. Stable decision problems. *The Annals of Statistics*, 6(5):1095–1110, 1978.

Miescke, K.-J. and Liese, F. *Statistical Decision Theory: Estimation, Testing, and Selection*. Springer, 2008. doi: https://doi.org/10.1007/978-0-387-73194-0.

Murphy, K. P. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Schneider, M. H. Matrix scaling, entropy minimization, and conjugate duality. i. existence conditions. *Linear Algebra and its Applications*, 114:785–813, 1989.

Shafto, P. and Goodman, N. Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In *Proceedings of the 30th annual conference of the Cognitive Science Society*, pp. 1632–1637. Cognitive Science Society Austin, TX, 2008.

Shafto, P., Goodman, N. D., and Frank, M. C. Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, 7(4):341–351, 2012.

Shafto, P., Goodman, N. D., and Griffiths, T. L. A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71:55–89, 2014.

Skinner, B. F. Teaching machines. *Science*, 128(3330): 969–977, 1958.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285, 2011.

Tomasello, M. *The cultural origins of human cognition*. Harvard University Press, Cambridge, MA, 1999.

Wang, P., Paranamana, P., and Shafto, P. Generalizing the theory of cooperative inference. *AIStats*, 2019a.

Wang, P., Wang, J., Paranamana, P., and Shafto, P. A mathematical theory of cooperative communication, 2019b.

Yang, S. C., Yu, Y., Givchi, A., Wang, P., Vong, W. K., and Shafto, P. Optimal cooperative inference. In *AISTATS*, volume 84 of *Proceedings of Machine Learning Research*, pp. 376–385. PMLR, 2018.

Zhu, X. Machine teaching for bayesian learners in the exponential family. In *Advances in Neural Information Processing Systems*, pp. 1905–1913, 2013.

Zhu, X. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *AAAI*, pp. 4083–4087, 2015.

Zilles, S., Lange, S., Holte, R., and Zinkevich, M. Teaching dimensions based on cooperative learning. In *COLT*, pp. 135–146, 2008.