

Learning Time Dependent Choice

Zachary Chase* Siddharth Prasad†

September 11, 2018

Abstract

We explore questions dealing with the learnability of models of choice over time. We present a large class of preference models defined by a structural criterion for which we are able to obtain an exponential improvement over previously known learning bounds for more general preference models. This in particular implies that the three most important discounted utility models of intertemporal choice – exponential, hyperbolic, and quasi-hyperbolic discounting – are learnable in the PAC setting with VC dimension that grows logarithmically in the number of time periods. We also examine these models in the framework of active learning. We find that the commonly studied stream-based setting is in general difficult to analyze for preference models, but we provide a redeeming situation in which the learner can indeed improve upon the guarantees provided by PAC learning. In contrast to the stream-based setting, we show that if the learner is given full power over the data he learns from – in the form of learning via membership queries – even very naive algorithms significantly outperform the guarantees provided by higher level active learning algorithms.

1 Introduction

We study the learnability of economic models of choice over time. Our setting is that of an analyst who first observes an agent’s choices between plans that specify payoffs over time, and then attempts to learn the preference parameters guiding the choices. While such parameters are stylized – in reality subjects are not likely to perform standardized computations according to private parameters before making decisions – experiments have shown that they often provide accurate descriptions of how an agent behaves. By observing enough choice data, one can hope to learn the economic parameters that most closely describe the agent’s preferences. Thus, learning theory provides an especially meaningful lens with which to view the theory of choice – it allows us to answer questions regarding the volume of data required to faithfully predict future decisions made by an observed agent. The overarching goal of this paper is to identify structural criteria that yield strong learnability results for preferences over time under different restrictions placed on the learner/analyst. The criteria we present captures a large class of preference models that give the agent significant freedom in weighting decisions against time delays. In particular, it encompasses the most popular models of time dependent choice used by economists.

*California Institute of Technology, zchase@caltech.edu

†California Institute of Technology, sprasad@caltech.edu

The main economic application of our results is in understanding the learnability of models of intertemporal choice. Intertemporal choice is what governs an agent’s decisions over several time periods. The most important models of intertemporal choice are discounted utility models, in which agents evaluate plans by discounting actions as they are delayed – in analogy to how markets value the loss or gain of money over time. The first axiomatic treatment of discounting was by Koopmans in 1960 [18], in which he demonstrates that simple postulates for preferences over an infinite time horizon yield “impatience.” The three most commonly studied discounting models are *exponential*, *hyperbolic*, and *quasi-hyperbolic*, and all have been studied by both economists and computer scientists (though less so by the latter) as well as researchers from various other fields. The importance of discounted utility in economics cannot be overstated – it is the canonical framework used by economists to study choice over time.

Problems of learning economic parameters have received recent attention from computer scientists; see, e.g., [1, 2, 3, 16, 22]. Inspired by a general theme of demanding computational robustness from economic models (Echenique, Golovin, and Wierman provide a nice discussion of this topic in [11]), the tools of learning theory provide relevant and exciting perspectives from which to view economic models that have been around for several decades. In contrast to the usual goal of truthfully extracting the agent’s parameters adopted by classical mechanism design, the learning problem aims to efficiently extract a truthful agent’s parameters in the restricted message space of binary classification. Our paper contributes to the line of work that specifically studies models of choice using the perspectives of learning theory. This confluence of decision theory and learning theory was initiated by Basu and Echenique [2], who consider the learning problem for models of choice under uncertainty. Our investigation in this paper is motivated by models of how agents make choices over time. We provide learnability results that are fine tuned to structural requirements on such models.

We now summarize our main contributions at a high level. Section 3 contains a more detailed exposition of our results.

Summary of results and techniques

Our situation is one of an analyst trying to learn the parameters governing an agent’s preferences over time. The two main learning themes we consider are (1) when the analyst has no control over the data he sees and (2) when the analyst has some control over the data he sees. The first theme is aptly captured by *probably approximately correct (PAC) learning*. To analyze the second theme, we investigate two models of *active learning*: stream-based selective sampling and membership queries.

In the first part, we study the PAC model, where the analyst is presented with pairs of alternatives and a label for each pair indicating the agent’s preference between the alternatives. The data points are drawn according to some unknown distribution, and the analyst has no control over the data he is presented with. Our main result here is a structural criterion on preference models that allows for a drastic improvement over the PAC learning complexity bounds achieved in [2]. We stipulate that the agent weights time-delayed payoffs according to polynomials, which allows for considerable freedom in how payoffs are weighted. Under this requirement, we show that such classes of preference models admit an exponential improvement in sample complexity bounds over the more general preference models considered in [2]. This is achieved via a computation of the VC dimension (which quantifies the complexity of

PAC learning). A simple application of our result shows that each of the discounted utility models are learnable, with sample complexity that grows logarithmically in the number of time periods T over which decisions are being made. The computation of the VC dimension is due to a natural connection between pairs of choices and the signs of polynomials that arise from the choices.

In the second part, we consider active learning models, where the analyst is given a certain amount of control over the data that he uses to learn. The two active learning models we study are *stream-based selective sampling* and learning via *membership queries*. In the former, the analyst is given some control over what data he learns from: as in the PAC setting he is presented with points drawn from an unknown distribution, but now the analyst chooses whether or not to see the label representing the agent’s choice for each point. In the latter, the analyst has complete control over the data he learns from: the analyst can at any time request the label for any point. The former model seems to have been commonly adopted in order to study the very general problem of concept learning, when there is no extra information about the structure of the concepts. We find that the *disagreement methods* used to study the stream-based setting are in general difficult to analyze in the context of preference models – requiring quantitative information about the underlying distribution from which points are drawn. However, we provide a redeeming situation (by examining a particular distribution) where we obtain an improvement over the PAC guarantees. Membership queries, on the other hand, allow us to heavily exploit the structure of the preference models we consider. We present a naive membership query-based algorithm that significantly outperforms the guarantees provided in the stream-based setting. Learning via membership queries, we conclude, seems to be the appropriate model to actively learn economic parameters. It allows the analyst to make use of the preference relations’ structure, and also precisely captures the situation in which the analyst and agent are participating in a real time experiment.

Related work

Discounted utility models of intertemporal choice have been studied extensively not only by economists, but also by researchers from various other fields. We first briefly survey some of the relevant work pertaining to the exponential, quasi-hyperbolic, and hyperbolic discounting models and then survey existing work in the more general topic of learning economic parameters.

In the exponential discounting model, the agent evaluates his utilities based on a discount factor $\delta \in (0, 1)$, where a delay of t time periods incurs an exponential discount in utility by δ^t . Climate change policies are traditionally evaluated according to an exponential discounting model – for example, the *Stern review* on the economics of climate change deals with issues of how to choose an appropriate discount rate in evaluating such policies [21]. Chambers and Echenique [8] present results related to the problem of aggregating discount rates proposed by a group of experts facing disagreement. While it is the most commonly used discounting model due to its simplicity, the exponential discounting model has been criticized due to its inability to match empirical data recording actual human behavior. Quasi-hyperbolic and hyperbolic discounting aim to mend such issues. The quasi-hyperbolic discounting model is parametrized by $\beta, \delta \in (0, 1)$, where a delay of t time periods incurs a discount in utility by $\beta\delta^t$, and was first introduced by Phelps and Pollack [20] to study preferences over generations. They proposed that the constant β discount factor represents how much a given generation t

is affected by the utilities of other people relative to their own – and remark that $\beta = 1$ represents “perfect altruism,” while $\beta < 1$ represents “imperfect altruism.” Kleinberg and Oren [17] study agents with quasi-hyperbolic discounting and propose a graph-theoretic model to investigate phenomena such as procrastination and abandonment of long-range tasks. Hyperbolic discounting aims to capture the notion that people are more impatient in making short term decisions (today vs. tomorrow) than long term decisions (365 days from today vs. 366 days from today)¹, and is modeled via a discount of $(1 + t\alpha)^{-1}$ at time t . Researchers in fields such as psychology and neuroscience [4, 15] have adopted the hyperbolic discounting model to study, for example, issues of self control and anticipation in humans and animals, and have compared the predictions by the different discounted utility models to neurobiological data obtained via MRI scans. Chabris et al. [7] give an exposition of the discounted utility models of intertemporal choice and survey sociological research that examines empirical data pertaining to how discount rates are affected by factors like age, drug use, gambling, etc.

The study of economic models has witnessed a recent influx of work from computer scientists dealing with questions of robustness under various notions of complexity (learning complexity, computational complexity, communication complexity, etc.). Kalai [16] in 2001 studied the learnability of choice functions, where the observed choices are in the form of a given set of alternatives along with the most preferred alternative from the set. Beigman and Vohra [3], Zadimoghaddam and Roth [22], and Balcan et al. [1] investigate the problem of learning utility functions in the context of an expected utility maximizing agent in a demand environment. Most recently (and most related to our work), Basu and Echenique [2] study the learnability of preference models of choice under uncertainty, in which an agent is uncertain about states of a lottery and is made to choose between acts that encode utilities over each state. Here, the different models of choice under uncertainty arise from different ways of representing the subjective probability held by an agent. They are also the first to study learnability in the decision-theoretic setting where choice is modeled by preference relations rather than by expected utility maximizing behavior in a demand setting. However, it does not appear that the learnability of models of intertemporal choice has been previously studied.

2 Model and Preliminaries

We now formally set up the discounted utility models of intertemporal choice and state the standard definitions from learning theory in the context of preference relations. Much of the following material regarding learning and preference relations is taken from [2] since we require a similar list of definitions and setup. First, we sketch our high level model.

Let X be a Euclidean space equipped with a Borel σ -algebra. A *preference relation* on X is a binary relation $\succsim \subseteq X \times X$ such that \succsim is measurable with respect to the product σ -algebra on $X \times X$. A *model* \mathcal{P} of preference relations is a collection of preference relations.

An agent makes choices from pairs of alternatives $(x^i, y^i)_{i=1}^n$ that are drawn according to some unknown distribution on $X \times X$. The choices are presented as labels $(a_i)_{i=1}^n$ where $a_i = 1$ if the agent chooses x^i and $a_i = 0$ if the agent chooses y^i . A *dataset* is any finite sequence of pairs of plans and their labels $((x^1, y^1), a_1), \dots, ((x^n, y^n), a_n)$. An analyst observes a dataset, and attempts to guess the preference relation governing the agent’s choices. A *learning rule*

¹In particular note that exponential discounting does not capture this issue, i.e. it is *dynamically consistent*, in that preferences do not change according to shifts in time.

is any map σ from datasets to preference relations. The output of the learning rule is the analyst's hypothesis as to what the agent's true preference relation is, having seen some finite dataset.

2.1 Learnability

The two notions of learnability we consider are the PAC model and the active model. We now state the standard definitions of PAC and active learning in the context of preference relations. Most of the following definitions for the PAC setting are taken from [2] since the setup involving preference relations is identical. These definitions of course apply to the more general setting of concept learning (for example, see [6]).

A collection \mathcal{P} of preference relations is (PAC) learnable if there is a learning rule σ such that for every $0 < \varepsilon, \delta < 1$, there is $s(\varepsilon, \delta) \in \mathbb{N}$ such that for every $n \geq s(\varepsilon, \delta)$, $\succsim \in \mathcal{P}$, and $\mu \in \Delta(X \times X)$,

$$\mu^n(\{(x^1, y^1), \dots, (x^n, y^n) : \mu(\succsim^* \Delta \succsim) > \varepsilon\}) < \delta,$$

where

$$\succsim^* = \sigma(\{(x^1, y^1), I_{x^1 \succsim y^1}, \dots, (x^n, y^n), I_{x^n \succsim y^n}\})$$

is the hypothesis preference relation produced by the learning rule². The quantity $s(\varepsilon, \delta)$ is called the *sample complexity* of the learning rule σ .

The complexity of learning is commonly quantified by the Vapnik-Chervonenkis (VC) dimension, which we now define. A set of points $\{(x^1, y^1), \dots, (x^n, y^n)\}$ from $X \times X$ is *shattered* by a model of preferences \mathcal{P} if for every vector of labels $(a_1, \dots, a_n) \in \{0, 1\}^n$, there is a preference relation $\succsim \in \mathcal{P}$ that realizes the labelling, i.e. for $i = 1, \dots, n$ we have that $x^i \succsim y^i$ if and only if $a_i = 1$. In this case, \mathcal{P} is said to *rationalize* the dataset $\{((x^1, y^1), a_1), \dots, ((x^n, y^n), a_n)\}$. The VC dimension of \mathcal{P} , denoted by $VC(\mathcal{P})$, is the largest integer n such that there exist n points that are shattered by \mathcal{P} .

Blumer et al. [6] in 1989 proved that learnability is equivalent to having a finite VC dimension³.

Theorem 2.1. *A model of preferences \mathcal{P} is learnable if and only if $VC(\mathcal{P}) < \infty$.*

The VC dimension (denoted by d for the remainder of this subsection) also plays a role in the sample complexity of learning a model of preferences. In the same paper, Blumer et al. [6] show that any algorithm that outputs a hypothesis consistent with the data seen is a valid learning rule requiring sample complexity

$$s(\varepsilon, \delta) = O\left(\frac{1}{\varepsilon} \left(d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}\right)\right).$$

In 2016, Hanneke [14] showed that these bounds (after a small improvement) are tight: the optimal sample complexity of PAC learning is

$$s(\varepsilon, \delta) = \Theta\left(\frac{1}{\varepsilon} \left(d + \log \frac{1}{\delta}\right)\right).$$

² μ^n denotes the product measure induced by μ on $(X \times X)^n$.

³This result requires \mathcal{P} to satisfy a certain measurability requirement. We note in Section 4 that the models of choice we consider all satisfy said requirement.

The other learning model we consider is the active learning framework, where the analyst has some control over the data from which he learns. In stream-based selective sampling, points drawn according to an unknown distribution are presented to the analyst as before, but without the labels. The analyst can choose whether or not to query the label of a given point, and the complexity of the learning rule is measured by *label complexity*, i.e. the number of labels requested by the analyst. Disagreement based active learning refers to the paradigm in which the learner only requests labels on points that significantly reduce the hypothesis space. The disagreement of a preference model with respect to the underlying distribution is quantified through the *disagreement coefficient* θ , which is defined in Section 5. A finite disagreement coefficient implies (for the underlying distribution) an exponential improvement in label complexity over the sample complexity of PAC learning. For example, the CAL algorithm [9, 10, 13], a simple disagreement based learning algorithm, yields a label complexity of

$$\ell_{CAL}(\varepsilon, \delta) = O\left(\theta \log \frac{1}{\varepsilon} \left(d \log \theta + \log \frac{\log(1/\varepsilon)}{\delta}\right)\right).$$

In the membership queries model, the analyst is allowed to request the label for any point at any time. There appears to be a dearth of literature/results pertaining to the complexity of membership query algorithms for learning when the hypothesis space is infinite. One explanation for this is that improvements to the “passive” disagreement based methods used in the stream-based setting would need specific information about the problem domain: disagreement based methods are designed to work on a very general class of concept learning problems without assuming anything about the learning space. In our case, we have specific details about how the preference relations take shape. Thus, the membership query model turns out to be an interesting and useful perspective to use in the study of learning preference models.

For a more detailed survey of active learning, see [10].

2.2 Discounted utility

We now present the definitions for the discounted utility models of intertemporal choice. An agent chooses between *plans* or vectors in $X = \mathbb{R}^T$ that encode payoffs over T time periods. A preference relation over plans is a binary relation $\succsim \subseteq \mathbb{R}^T \times \mathbb{R}^T$.

The most important model of intertemporal choice is the *discounted utility model*, in which the agent’s payoffs x_t for having chosen a plan $x \in \mathbb{R}^T$ are reduced, or discounted, as t increases from 1 to T . In its most general form, we can characterize the preference relations that follow time discounting as follows:

Definition 2.1 (Discounted utility model). The class of preference relations $\mathcal{P}_{\mathcal{D}}$ that satisfy the *discounted utility model* are those \succsim such that there exists a decreasing map $D : \{1, \dots, T\} \rightarrow (0, 1)$ where

$$x \succsim y \text{ if and only if } \sum_{t=1}^T D(t)x_t \geq \sum_{t=1}^T D(t)y_t.$$

We use the following notation for the preference models arising from the three most commonly studied discounting functions D :

- $\mathcal{P}_{\mathcal{D}}$ denotes the set of preferences that satisfy the discounted utility model.
- $\mathcal{P}_{\mathcal{ED}}$ denotes the set of preferences that satisfy the discounted utility model with *exponential discounting*: $D(t) = \delta^t$ for $\delta \in (0, 1)$.
- $\mathcal{P}_{\mathcal{HD}}$ denotes the set of preferences that satisfy the discounted utility model with *hyperbolic discounting*: $D(t) = \frac{1}{1+t\alpha}$ for $\alpha > 0$.
- $\mathcal{P}_{\mathcal{QHD}}$ denotes the set of preferences that satisfy the discounted utility model with *quasi-hyperbolic discounting*: $D(t) = 1$ if $t = 1$, $D(t) = \beta \cdot \delta^{t-1}$ if $t > 1$ for $\beta, \delta \in (0, 1)$.

For a more thorough exposition on the various discounted utility models of intertemporal choice, see [7].

3 Main Results

In this section we provide a formal discussion and interpretation of our results, which is split into two themes: the first dealing with an analyst who has no control over the learning data, the second dealing with an analyst who has some control over the learning data.

A powerless analyst

The first part of our paper investigates the situation of an analyst trying to learn the preference relation by which an agent makes choices, but has no control over what choices he gets to observe, and is agnostic to the process by which they are drawn. We thus adopt the PAC learning model.

The agent chooses between plans that encode payoffs over T periods of time and evaluates the total payoff of a plan vector $x \in \mathbb{R}^T$ according to private weights w_1, \dots, w_T that he multiplicatively applies to each state: $\text{payoff}(x) = \sum_{t=1}^T w_t x_t$. This defines a model of preference relations, which we denote by $\mathcal{P}_{\mathcal{W}}$, where for any $\succsim \in \mathcal{P}_{\mathcal{W}}$, there exists a vector of weights $w = (w_1, \dots, w_T) \in \mathbb{R}^T$ such that

$$x \succsim y \text{ if and only if } w \cdot x \geq w \cdot y.$$

In [2], it is shown that $T - 1 \leq VC(\mathcal{P}_{\mathcal{W}}) \leq T + 1$. In the context of choice over time, however, this model is extremely general and does not capture any of the intuitive notions of how an agent values payoffs when they are delayed⁴. For example, the discounted utility models of intertemporal choice require the weights to be of a particular functional form. Moreover, when there is no structure to the discount function we cannot improve the bounds on $\mathcal{P}_{\mathcal{W}}$:

Proposition 3.1. $T - 1 \leq VC(\mathcal{P}_{\mathcal{D}}) \leq T + 1$.

This leads us to the motivating question of the first part of the paper: what structural conditions can we impose on the weights w_1, \dots, w_T such that this bound can be improved?

We investigate the situation where the agent computes his weights by evaluating polynomials at a private parameter δ . Specifically, let Q_1, \dots, Q_T be polynomials of degree at

⁴In [2] the complete control over weights is used to model choice under uncertainty, which calls for such generality since the agent's beliefs/weights are given by an element of the probability simplex on \mathbb{R}^T .

most d , and suppose the agent evaluates total payoff of a plan vector $x \in \mathbb{R}^T$ by $\text{payoff}(x) = \sum_{t=1}^T Q_t(\delta)x_t$. Consequently, let $\mathcal{P}_{\mathcal{PW}}$ be the model of preference relations parametrized by δ such that

$$x \succsim y \text{ if and only if } \sum_{t=1}^T Q_t(\delta)x_t \geq \sum_{t=1}^T Q_t(\delta)y_t.$$

This class of preference models allows us to approximate preference relations where the weights are given by any real valued functions – we choose Q_1, \dots, Q_T to be the appropriate Taylor polynomials. Moreover, existing models of intertemporal choice fit this characterization – for example $\mathcal{P}_{\mathcal{ED}}$ and $\mathcal{P}_{\mathcal{HD}}$.

We additionally consider a slightly larger class of preference models where the agent has a private parameter β (in addition to δ) that in evaluating total payoff of a plan vector $x \in \mathbb{R}^T$ allows the agent to modify the constant term $\sum_{t=1}^T Q_t(0)x_t$ of the polynomial $\sum_{t=1}^T Q_t(\delta)x_t$. This model aims to more generally capture the effects of the β parameter in quasi-hyperbolic discounting. For polynomials Q_1, \dots, Q_T of degree at most d , let $\mathcal{P}_{\mathcal{BPW}}$ be the model of preference relations parametrized by β and δ such that $x \succsim y$ if and only if

$$\left(\frac{1}{\beta} - 1\right) \sum_{t=1}^T Q_t(0)x_t + \sum_{t=1}^T Q_t(\delta)x_t \geq \left(\frac{1}{\beta} - 1\right) \sum_{t=1}^T Q_t(0)y_t + \sum_{t=1}^T Q_t(\delta)y_t.^5$$

Our main results show that with this additional structure on the preference model, we can achieve an exponential improvement in the bounds for the VC dimension of $\mathcal{P}_{\mathcal{W}}$ obtained in [2]⁶.

Theorem 3.2. *For every $\varepsilon > 0$, there exists a d_ε such that for every $d \geq d_\varepsilon$ we have $VC(\mathcal{P}_{\mathcal{PW}}), VC(\mathcal{P}_{\mathcal{BPW}}) \leq (1 + \varepsilon) \log d$ for any T and any T polynomials Q_1, \dots, Q_T of degree at most d .*

Note that when Q_1, \dots, Q_T have degree at most polynomial in T , we obtain an exponential improvement over the linear growth of $VC(\mathcal{P}_{\mathcal{W}})$. We show that in this case, we get a tight (asymptotic) bound of $\log T$:

Theorem 3.3. *Let Q_1, \dots, Q_T be polynomials in δ of degree at most $T - 1$ that span the space of polynomials in δ of degree at most $T - 1$. Then $VC(\mathcal{P}_{\mathcal{PW}}), VC(\mathcal{P}_{\mathcal{BPW}}) \geq \log(T - 1)$.*

An interesting feature of Theorems 3.2 and 3.3 is that for fixed Q_1, \dots, Q_T with degrees at most $T - 1$, $VC(\mathcal{P}_{\mathcal{PW}})$ and $VC(\mathcal{P}_{\mathcal{BPW}})$ satisfy the same asymptotic bounds, so giving the agent an extra parameter that allows control over the constant term of the polynomial $\sum_{t=1}^T Q_t(\delta)x_t$ does not introduce a significant amount of richness to the model.

Applying Theorems 3.2 and 3.3 to the discounted utility models, we have:

Corollary 3.4. $VC(\mathcal{P}_{\mathcal{ED}}), VC(\mathcal{P}_{\mathcal{HD}}), VC(\mathcal{P}_{\mathcal{QHD}}) \sim \log(T - 1)$

⁵We use a factor of $(1/\beta - 1)$ since it yields a clean description of quasi-hyperbolic discounting.

⁶It is important to note that the classes $\mathcal{P}_{\mathcal{PW}}$ and $\mathcal{P}_{\mathcal{BPW}}$ are defined for a given Q_1, \dots, Q_T . That is, the analyst knows Q_1, \dots, Q_T , and is trying to learn the parameters β and δ . If the Q_1, \dots, Q_T are private information only available to the agent, we are in no better shape than in the case of $\mathcal{P}_{\mathcal{W}}$.

Thus, $\mathcal{P}_{\mathcal{D}}$, $\mathcal{P}_{\mathcal{E}\mathcal{D}}$, $\mathcal{P}_{\mathcal{H}\mathcal{D}}$, and $\mathcal{P}_{\mathcal{Q}\mathcal{H}\mathcal{D}}$ are all learnable. $\mathcal{P}_{\mathcal{D}}$ requires a minimum sample size that grows linearly with T , while $\mathcal{P}_{\mathcal{E}\mathcal{D}}$, $\mathcal{P}_{\mathcal{H}\mathcal{D}}$, and $\mathcal{P}_{\mathcal{Q}\mathcal{H}\mathcal{D}}$ require a minimum sample size that grows logarithmically in T .

The main technique in proving Theorems 3.2 and 3.3 is interpreting the shattering criteria as a statement about the sign combinations achieved by a collection of polynomials. The upper bound on the VC dimension follows from an upper bound on the number of sign combinations a collection of polynomials can achieve. In demonstrating the lower bound on the VC dimension, we construct a set of points that is shattered by finding polynomials achieving all possible sign combinations – chosen according to a Hamiltonian path in the $\log(T - 1)$ -dimensional hypercube.

A powerful analyst

Our other results concern the active learning framework, which broadly deals with situations in which the analyst has some control over the choices he observes and learns from. The two models we consider are *stream-based selective sampling* and learning via *membership queries*. A large body of active learning research is devoted to the stream-based model, specifically focusing on disagreement based algorithms – a class of learning algorithms that instructs the analyst only to request labels on points he sees that reduce the hypothesis space significantly. In the most general setting of concept learning, this is a useful framework since the error guarantees can be described using the same setup as the PAC model. Moreover without additional information about the problem domain, it is unclear how to devise efficient algorithms that are more specific in instructing the analyst on what questions to ask.

We find that the stream-based model is in general difficult to analyze for the preference relations we work with. This difficulty seems to arise from the apparent need to quantify disagreement in order to explicitly write down learning guarantees. Though in most general situations it is unclear how to quantify disagreement for our preference relations, we present a redeeming situation for which we are able to provide a precise analysis of the learning guarantees for $\mathcal{P}_{\mathcal{E}\mathcal{D}}$. Here, the analyst can learn $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ with an exponential improvement in label complexity over the guarantees provided by the PAC model. This is achieved via a computation of the disagreement coefficient (defined in Section 5) of $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ for a specific distribution.

Theorem 3.5. *There exists a distribution μ on $\mathbb{R}^T \times \mathbb{R}^T$ for which the disagreement coefficient of $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ is $\theta = 2$. Thus, for this distribution,*

$$\ell_{CAL}(\varepsilon) = \tilde{O}\left(\log T \log \frac{1}{\varepsilon}\right),$$

where the \tilde{O} notation suppresses terms that are logarithmic in $\log T$ and $\log 1/\varepsilon$.

The measure μ we construct is induced by the product Lebesgue measure on $(0, 1)^{T-1}$, and allows us to precisely translate statements about disagreement into statements about the roots of polynomials arising from a given choice. Once we have defined μ , the calculation of θ follows from basic probability arguments.

Now, in our case the analyst has structural information regarding the preference relation of the agent he is questioning. We find that allowing the analyst full control over the membership queries he makes yields a learning algorithm that, despite its simplicity, takes advantage of

this extra structure and yields a significant improvement in complexity over the stream-based setting. Additionally, the membership queries model naturally describes an experimental environment in which the analyst is able to ask the agent questions in real time.

We show that when the preference model satisfies some relatively benign structural requirements, even very naive algorithms outperform the guarantees provided by CAL in the stream-based setting. The example algorithm we give, relying on a simple binary search, has a query complexity of $O(\log 1/\varepsilon)$, which gets rid of the $\log T$ dependence in Theorem 3.5.

The class of preference models is defined as follows: let $g_1, \dots, g_T : \mathbb{R} \rightarrow \mathbb{R}$ be a collection of functions satisfying the properties listed in Section 5.3 and consider the model of preference relations \mathcal{P} parametrized by δ where $x \succsim y$ if and only if $\sum_{t=1}^T g_t(\delta)x_t \geq \sum_{t=1}^T g_t(\delta)y_t$ ⁷. We have

Proposition 3.6. *There exists an algorithm that takes as input $\varepsilon > 0$ and using $O(\log 1/\varepsilon)$ membership queries outputs δ^h such that $|\delta - \delta^h| \leq \varepsilon$, where δ parametrizes the target preference relation in \mathcal{P} .*

The remainder of the paper is devoted to proving the results discussed in this section.

4 PAC Learning

In this section we prove Theorems 3.2 and 3.3. We first note a preliminary upper bound due to Basu and Echenique [2]. Let $\mathcal{P}_{\mathcal{I}}$ be the set of preference relations that satisfy the following axioms:

Order: For all x, y either $x \succsim y$ or $y \succsim x$ (completeness). For all x, y, z , if $x \succsim y$ and $y \succsim z$, then $x \succsim z$ (transitivity).

Independence: For all x, y, z and for any $\lambda \in (0, 1)$, $x \succsim y$ if and only if $\lambda x + (1 - \lambda)z \succsim \lambda y + (1 - \lambda)z$.

The class $\mathcal{P}_{\mathcal{I}}$ satisfies the property that for any $\succsim \in \mathcal{P}_{\mathcal{I}}$, there are finitely many vectors q_1, \dots, q_K , with $K \leq T$, such that $x \succsim y$ if and only if $(q_k \cdot x)_{k=1}^K \geq_L (q_k \cdot y)_{k=1}^K$, where \geq_L denotes the lexicographic order [5]. Then, $\mathcal{P}_{\mathcal{PW}}, \mathcal{P}_{\mathcal{BPW}} \subset \mathcal{P}_{\mathcal{W}} \subset \mathcal{P}_{\mathcal{I}}$, since the aforementioned characterization is satisfied with $K = 1$ and $q_1 = (w_1, \dots, w_T)$.

This has two main consequences. First, $\mathcal{P}_{\mathcal{PW}}$ and $\mathcal{P}_{\mathcal{BPW}}$ (and thus all the discounted utility models) satisfy the measurability requirement discussed in Lemma 4 of [2] for the equivalence result of Theorem 2.1 to hold. Second, the VC dimensions of $\mathcal{P}_{\mathcal{PW}}$ and $\mathcal{P}_{\mathcal{BPW}}$ are all bounded above by $T + 1$ (and in particular $T - 1 \leq VC(\mathcal{P}_{\mathcal{W}}) \leq T + 1$). This follows due to Theorem 3.1 of [2], in which an argument similar to that required to compute the VC dimension of the class of half-spaces is used to show that $VC(\mathcal{P}_{\mathcal{I}}) = T + 1$. In all cases excluding the most general model of discounted utility, we are able to bring this down to $\log(T - 1)$ (which we then show is tight by demonstrating the corresponding lower bound).

We begin by demonstrating that even in the discounted utility setting, without any structure we cannot do better than the learning bounds obtained for $\mathcal{P}_{\mathcal{I}}$.

Proposition 3.1. $T - 1 \leq VC(\mathcal{P}_{\mathcal{D}}) \leq T + 1$.

⁷As before, the g_1, \dots, g_T are known to the analyst.

Proof. That $VC(\mathcal{P}_{\mathcal{D}}) \leq T + 1$ follows from Theorem 3.1 of [2], since $\mathcal{P}_{\mathcal{D}} \subset \mathcal{P}_{\mathcal{I}}$.

Here is a simple construction that shows $VC(\mathcal{P}_{\mathcal{D}}) \geq T - 1$. Fix an $\varepsilon > 0$. Let e_1, \dots, e_T be the standard unit vectors in \mathbb{R}^T , and consider the set of points $\{(x^1, y^1), \dots, (x^{T-1}, y^{T-1})\}$, where $x^i = (1 - \varepsilon)e_i$ and $y^i = e_{i+1}$.

This set is shattered by $\mathcal{P}_{\mathcal{D}}$: for any $(a_i)_{i=1}^{T-1}$, choose $D(1)$ arbitrarily from $(0, 1)$, and if $D(i)$ has been defined, inductively define $D(i+1)$ such that $D(i+1) \leq D(i)(1 - \varepsilon)$ if $a_i = 1$ and $D(i) > D(i+1) > (1 - \varepsilon)D(i)$ if $a_i = 0$. \square

We now prove Theorems 3.2 and 3.3, which are restated below for convenience.

Theorem 3.2. *For every $\varepsilon > 0$, there exists a d_ε such that for every $d \geq d_\varepsilon$ we have $VC(\mathcal{P}_{\mathcal{PW}}), VC(\mathcal{P}_{\mathcal{BPW}}) \leq (1 + \varepsilon) \log d$ for any T and any T polynomials Q_1, \dots, Q_T of degree at most d .*

Proof. It suffices to establish the bound for $\mathcal{P}_{\mathcal{BPW}}$.

Let (z^1, \dots, z^n) be a set of points in $\mathbb{R}^T \times \mathbb{R}^T$, $z^i = (x^i, y^i)$. For each $z^i = (x^i, y^i)$, define the plan $f^i := x^i - y^i$. Then, note that (z^1, \dots, z^n) is shattered by $\mathcal{P}_{\mathcal{BPW}}$ if and only if $((f^1, 0), \dots, (f^n, 0))$ is shattered by $\mathcal{P}_{\mathcal{BPW}}$. Hence, we may (and do) restrict attention to datasets of the form $((f^1, 0), \dots, (f^n, 0))$.

We have that $((f^1, 0), \dots, (f^n, 0))$ is shattered by $\mathcal{P}_{\mathcal{BPW}}$ if and only if for all vectors $(a_1, \dots, a_n) \in \{0, 1\}^n$, there exists a δ and β (which determines the preference relation) such that

$$(Q_T(\delta)f_T^i + \dots + Q_1(\delta)f_1^i) + \left(\frac{1}{\beta} - 1\right) (Q_T(0)f_T^i + \dots + Q_1(0)f_1^i) \geq 0 \text{ whenever } a_i = 1,$$

and

$$(Q_T(\delta)f_T^i + \dots + Q_1(\delta)f_1^i) + \left(\frac{1}{\beta} - 1\right) (Q_T(0)f_T^i + \dots + Q_1(0)f_1^i) < 0 \text{ whenever } a_i = 0.$$

We first show that for all $\varepsilon > 0$, for sufficiently large d we have $VC(\mathcal{P}_{\mathcal{BPW}}) \leq (1 + \varepsilon) \log d$. Note that if the n points $((f^1, 0), \dots, (f^n, 0))$ can be shattered, there are polynomials P_1, \dots, P_n in δ (where P_i is the polynomial $Q_1(\delta)f_1^i + \dots + Q_T(\delta)f_T^i$), each of degree at most d , such that for every labeling $(a_1, \dots, a_n) \in \{0, 1\}^n$, there exists a δ and β such that

$$(\text{sgn}(P_1(\delta) + (1/\beta - 1)P_1(0)), \dots, \text{sgn}(P_n(\delta) + (1/\beta - 1)P_n(0))) = (a_1, \dots, a_n)^8.$$

First, for any n polynomials P_1, \dots, P_n of degree at most d , we give an upper bound on the number of possible values $(\text{sgn}(P_1(\delta)), \dots, \text{sgn}(P_n(\delta)))$ can realize. Each polynomial has at most d real roots, so together P_1, \dots, P_n have at most nd distinct real roots. Since sign changes can only occur at the roots, there are at most $nd + 1$ possible values of $\{0, 1\}^n$ that $(\text{sgn}(P_1(\delta)), \dots, \text{sgn}(P_n(\delta)))$ can realize.

Now, for a fixed δ , varying β shifts the collection of polynomials

$$P_1(\delta) + (1/\beta - 1)P_1(0), \dots, P_n(\delta) + (1/\beta - 1)P_n(0)$$

⁸For notational convenience, let $\text{sgn}(x)$ be 1 if $x \geq 0$ and 0 otherwise.

vertically, which in the worst case induces sign changes in all entries. We thus get at most an additional n new sign combinations for every sign combination realized by $(\text{sgn}(P_1(\delta)), \dots, \text{sgn}(P_n(\delta)))$. Hence, there are at most $nd + 1 + n(nd + 1) = (n^2 + n)d + n + 1$ possible values of $\{0, 1\}^n$ that

$$(\text{sgn}(P_1(\delta) + (1/\beta - 1)P_1(0)), \dots, \text{sgn}(P_n(\delta) + (1/\beta - 1)P_n(0)))$$

can realize.

In order for all 2^n elements of $\{0, 1\}^n$ to be realized, it must be that

$$(n^2 + n)d + n + 1 \geq 2^n.$$

If $n > (1 + \varepsilon) \log d$, then for large enough d this inequality does not hold, and so any set of n points cannot be shattered. Thus, for all $\varepsilon > 0$, $n \leq (1 + \varepsilon) \log d$ for large enough d , i.e. $VC(\mathcal{P}) \leq (1 + \varepsilon) \log d$. \square

We now establish the corresponding lower bound when the polynomials Q_1, \dots, Q_T span the space of polynomials of degree at most $T - 1$.

Theorem 3.3. *Let Q_1, \dots, Q_T be polynomials in δ of degree at most $T - 1$ that span the space of polynomials in δ of degree at most $T - 1$. Then $VC(\mathcal{P}_{\mathcal{PW}}), VC(\mathcal{P}_{\mathcal{BPW}}) \geq \log(T - 1)$.*

Proof. It suffices to establish the bound for $\mathcal{P}_{\mathcal{PW}}$.

Consider the graph on $\{0, 1\}^n$ where two vertices are connected by an edge if they differ in exactly one location. Fix a Hamiltonian path v_1, v_2, \dots, v_{2^n} in this graph (the existence of which is well known). Let $b_{1,2}, \dots, b_{2^n-1,2^n}$ be the sequence where $b_{i,i+1}$ is the index of the location at which v_i and v_{i+1} differ. Note that if $n = \log(T - 1)$, the graph has $T - 1$ vertices, so each index in $\{1, \dots, n\}$ can appear in the sequence $(b_{i,i+1})$ at most $T - 1$ times.

Now, let $r_1 < r_2 < \dots < r_{2^n}$ be any points in $(0, 1)$. Define n polynomials P_1, \dots, P_n by $P_k(\delta) = \prod_{b_{i,i+1}=k} (\delta - r_i)$, so the roots of P_k are precisely the r_i 's that correspond to a flip in the entry at the k th position of a vertex in the path. Then, $(\text{sgn}(P_1(\delta)), \dots, \text{sgn}(P_n(\delta)))$ realizes every element of $\{0, 1\}^n$.

Since Q_1, \dots, Q_T span the space of polynomials of degree at most $T - 1$, for each P_i we can find f_1^i, \dots, f_T^i such that

$$P_i(\delta) = Q_1(\delta)f_1^i + \dots + Q_T(\delta)f_T^i,$$

which gives us a collection of $\log(T - 1)$ points that is shattered. Hence $\log(T - 1) \leq VC(\mathcal{P})$. \square

It is readily seen that $\mathcal{P}_{\mathcal{ED}}$ and $\mathcal{P}_{\mathcal{QHD}}$ satisfy the conditions of Theorems 3.2 and 3.3⁹. We give a quick argument verifying that $\mathcal{P}_{\mathcal{HD}}$ does as well: For $1 \leq t \leq T$, let $Q_t(\alpha) = \prod_{\ell \in \{1, \dots, T\} \setminus \{t\}} (1 + \ell\alpha)$ (these are the polynomials obtained by clearing denominators of the hyperbolic discount factors). We argue that $\{Q_t(\alpha)\}_{t=1}^T$ are linearly independent over the vector space of polynomials in α of degree at most $T - 1$. Indeed, if $Q_1(\alpha)f_1 + \dots + Q_T(\alpha)f_T = 0$, then we must have that $f_1 = \dots = f_T = 0$ since at $\alpha = \frac{-1}{t}$ we get $Q_t(\alpha)f_t = 0$. Hence, $\mathcal{P}_{\mathcal{HD}}$ is simply $\mathcal{P}_{\mathcal{PW}}$ with $Q_t(\alpha) = \prod_{\ell \in \{1, \dots, T\} \setminus \{t\}} (1 + \ell\alpha)$.

⁹ $\mathcal{P}_{\mathcal{ED}}$ is given by $\mathcal{P}_{\mathcal{PW}}$ with $Q_t(\delta) = \delta^{t-1}$, and $\mathcal{P}_{\mathcal{QHD}}$ is given by $\mathcal{P}_{\mathcal{BPW}}$ with $Q_t(\delta) = \delta^{t-1}$.

Remark. These bounds also hold in the scenario where the agent can report indifference in the data. More precisely, the condition for $x \succsim y$ is now a strict inequality, and we have three possible labels for the pair (x, y) : $+1$ indicates that $x \succ y$, -1 indicates that $y \succ x$, and 0 indicates that $x \sim y$. Then, using a Hamiltonian path on $\{-1, 0, 1\}^n$, we can construct polynomials P_1, \dots, P_n such that $(\text{sgn}(P_1(\delta)), \dots, \text{sgn}(P_n(\delta)))$ realizes all elements of $\{-1, 0, 1\}^n$ as δ ranges from 0 to 1 (where sgn is the true sign function).

4.1 A Remark on Efficient Learnability

While PAC learnability is a positive result, it does not take into account the computational complexity of computing a hypothesis. Blumer et. al. [6] show that any learning rule that outputs a hypothesis consistent with the data seen yields with high probability a hypothesis that has very low error. However, if the problem of outputting a consistent hypothesis is computationally intractable, PAC learnability on its own is perhaps unsatisfying. In this section we note that the discounted utility models of intertemporal choice are *efficiently* learnable. This is due to an algorithm of Grigor'ev and Vorobjov [12] for solving a system of polynomial inequalities.

For notational convenience, it will be useful to write $\mathcal{P} = \{\mathcal{P}^T\}_{T \geq 1}$, for each of the models above, where \mathcal{P}^T is the collection of preference relations for a given T . Moreover, suppose acts are chosen from $[-1, 1]^T$ instead of \mathbb{R}^T . It is clear that this does not change any of the analysis above.

Polynomial learnability, as defined by Blumer et. al. [6], stipulates that the learning rule be computable in $\text{poly}(1/\varepsilon, 1/\delta, T)$ -time (where ε and δ denote the error threshold and confidence threshold respectively). Polynomial learnability is equivalent to the task of outputting a hypothesis consistent with the given data set in polynomial time [6].

Definition 4.1. A *randomized polynomial hypothesis finder (r-poly hy-fi)* for \mathcal{P} is a randomized polynomial time algorithm that takes as input a sample of a preference relation in \mathcal{P} , and for some $\gamma > 0$, with probability at least γ produces a hypothesis that is consistent with the sample.

Theorem 4.1. \mathcal{P} is properly polynomially learnable if and only if there is an r-poly hy-fi for \mathcal{P} and $VC(\mathcal{P}^T)$ grows only polynomially in T .

We have just shown that $VC(\mathcal{P}^T) \sim \log(T - 1)$. Using techniques involving algebraic geometry, Grigor'ev and Vorobjov [12] present an algorithm to solve a system of N polynomial inequalities with a $\text{poly}(N, T)$ runtime. This serves as our r-poly hy-fi, and thus we obtain:

Theorem 4.2. $\mathcal{P}_{\mathcal{E}\mathcal{D}}$, $\mathcal{P}_{\mathcal{H}\mathcal{D}}$, and $\mathcal{P}_{\mathcal{Q}\mathcal{H}\mathcal{D}}$ are all properly polynomially learnable.

5 Active Learning

In this section we study two models of active learning: *stream-based selective sampling* and learning via *membership queries*. We first define a distribution for which the disagreement coefficient of $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ is 2, showing that disagreement methods (specifically the CAL algorithm [9, 10, 13]) in the stream-based model can yield an exponential improvement over the sample complexity of PAC learning (thus proving Theorem 3.5).

We then consider learning via membership queries and show that in this setting even very naive algorithms outperform the disagreement methods in the stream-based model (that is, the analyst needs to ask fewer questions to the agent in order to learn his preference than the number of label requests he would need to make using disagreement methods).

5.1 Preliminaries

For notational convenience, \succsim_δ will refer to the preference relation in $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ with discounting factor δ .

Let μ be a distribution on $\mathbb{R}^T \times \mathbb{R}^T$. μ induces a metric on $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ by $d(\succsim_\delta, \succsim_\gamma) = \mu(\succsim_\delta \Delta \succsim_\gamma)$, and thus we can define the closed ball of radius R centered at \succsim_δ by

$$B(\succsim_\delta, R) = \{\succsim_\gamma : d(\succsim_\delta, \succsim_\gamma) \leq R\}.$$

For $V \subseteq \mathcal{P}_{\mathcal{E}\mathcal{U}}$, the disagreement region of V , $\text{Dis}(V)$ is defined by

$$\text{Dis}(V) = \{(x, y) \in \mathbb{R}^T \times \mathbb{R}^T : \exists \succsim_\delta, \succsim_\gamma \in V \text{ s.t. } (x, y) \in \succsim_\delta \Delta \succsim_\gamma\} = \bigcup_{\succsim_\delta, \succsim_\gamma \in V} (\succsim_\delta \Delta \succsim_\gamma)$$

Intuitively, $\text{Dis}(V)$ is the collection of points (x, y) such that we can find two hypothesis relations in the current version space that rank x and y differently.

If \succsim_δ is the target preference relation, the disagreement coefficient of \succsim_δ with respect to μ is the quantity

$$\theta = \sup_{R>0} \frac{\mu(\text{Dis}(B(\succsim_\delta, R)))}{R}$$

5.2 Disagreement based active learning

In this subsection, we define a distribution μ on $\mathbb{R}^T \times \mathbb{R}^T$ and show that the disagreement coefficient of $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ with respect to μ is 2.

Choosing a measure

The main challenge here is that θ depends on the underlying distribution over $\mathbb{R}^T \times \mathbb{R}^T$. Since preferences are polynomial inequalities, the disagreement coefficient seems to lend itself to a characterization involving polynomials and their roots, which is the motivation for our choice of distribution. For a general distribution μ over $\mathbb{R}^T \times \mathbb{R}^T$, it is not clear how to compute the disagreement coefficient.

We show that $\theta = 2$ for a suitably chosen distribution on $\mathbb{R}^T \times \mathbb{R}^T$, which is induced by the Lebesgue measure on $(0, 1)^{T-1}$. This allows us to work with a measure on sets of roots of polynomials that arise from the definition of the preference relations.

Let μ^{**} be a measure on $(0, 1)^{T-1}$. We interchangeably represent elements of \mathbb{R}^T as polynomials P of degree at most $T - 1$ or as $T - 1$ -tuples of coefficients. Let \sim be the equivalence relation on \mathbb{R}^T defined by $P \sim Q \iff P = cQ$ for some constant c , and let \mathbb{R}^T / \sim be the resulting quotient space. Let $g : (0, 1)^{T-1} \rightarrow \mathbb{R}^T / \sim$ be the map taking a tuple of roots to the equivalence class of the polynomials with those roots, and let $h : \mathbb{R}^T \times \mathbb{R}^T \rightarrow \mathbb{R}^T / \sim$ be the map $h(x, y) = [x - y]$.

We define the following measures μ^* and μ on $g((0, 1)^{T-1}) \subset \mathbb{R}^T / \sim$ and $h^{-1}(g((0, 1)^{T-1})) \subset \mathbb{R}^T \times \mathbb{R}^T$ respectively.

- Define μ^* on all sets $S \subset g((0,1)^{T-1})$ such that

$$\{(r_1(P), \dots, r_{T-1}(P)) \in (0,1)^{T-1} : P \in S\}$$

(where $r_1(P), \dots, r_{T-1}(P)$ denote the roots of P) is μ^{**} -measurable, for which we set

$$\mu^*(S) = \mu^{**}(\{(r_1(P), \dots, r_{T-1}(P)) \in (0,1)^{T-1} : P \in S\}).$$

- Define μ on all sets $S \subset h^{-1}(g((0,1)^{T-1}))$ such that

$$\{[z] \in g((0,1)^{T-1}) : \exists(x, y) \in S \text{ s.t. } z \sim x - y\}$$

is μ^* -measurable, for which we set

$$\mu(S) = \mu^*(\{[z] \in g((0,1)^{T-1}) : \exists(x, y) \in S \text{ s.t. } z \sim x - y\}).$$

Intuitively, μ^* is defined only on those polynomials that have all their roots in $(0,1)$. When $T = 2$, this is a desirable property since the analyst is only presented with polynomials that have some disagreement in $(0,1)$. He is not presented with meaningless polynomials that are, for example, always positive on $(0,1)$ (the analyst has nothing to learn from such polynomials since such a polynomial will be preferred to 0 for all $\delta \in (0,1)$). For $T \geq 3$, this is a more restrictive property since the analyst is only presented with polynomials that have all $T - 1$ roots in $(0,1)$.

Let μ^{**} be the product Lebesgue measure on $(0,1)^{T-1}$. Choosing μ^{**} in this fashion allows us to neatly characterize $B(\succsim_\delta, R)$.

Let X_1, \dots, X_{T-1} be uniform i.i.d. random variables on $(0,1)$, and let $Y_{\delta,\gamma}$ be the random variable $Y_{\delta,\gamma} = |\{i : X_i \text{ is between } \delta \text{ and } \gamma\}|$. Let $E_{\delta,\gamma}^{odd}$ denote the event that $Y_{\delta,\gamma}$ is odd, let $E_{\delta,\gamma}^k$ denote the event $Y_{\delta,\gamma} = k$, and let $E_{\delta,\gamma}^{\geq k}$ denote the event $Y_{\delta,\gamma} \geq k$.

Lemma 5.1. $\succsim_\gamma \in B(\succsim_\delta, R)$ if and only if $\mathbb{P}[E_{\delta,\gamma}^{odd}] \leq R$.

Proof. Given $(x, y) \in \mathbb{R}^T \times \mathbb{R}^T$, let $P_{x-y}(X) = \sum_{t=1}^T X^{t-1} \cdot (x_t - y_t)$. Then,

$$\begin{aligned} \succsim_\gamma \in B(\succsim_\delta, R) &\iff \mu(\{(x, y) \in h^{-1}(g((0,1)^{T-1})) : \text{sgn}(P_{x-y}(\delta)) \neq \text{sgn}(P_{x-y}(\gamma))\}) \leq R \\ &\iff \mu^*(\{[P] \in g((0,1)^{T-1}) : \text{sgn}(P(\delta)) \neq \text{sgn}(P(\gamma))\}) \leq R \\ &\iff \mu^{**}(\{(r_1, \dots, r_{T-1}) \in (0,1)^{T-1} : \text{sgn}(\prod(\delta - r_i)) \neq \text{sgn}(\prod(\gamma - r_i))\}) \leq R \end{aligned}$$

But $\text{sgn}(\prod(\delta - r_i)) \neq \text{sgn}(\prod(\gamma - r_i))$ occurs exactly when an odd number of roots lie between γ and δ (modulo a set of measure 0 since the probability that we have a root of multiplicity greater than 1 is 0). \square

While it seems difficult to write down a general characterization of μ , Propositions 5.2 and 5.3 give some basic observations regarding the σ -algebras on which μ^* and μ are defined. We defer their proofs (along with a description of μ^* in the case $T = 2$) to the appendix:

Proposition 5.2. *The σ -algebra on $g((0,1)^{T-1})$ is the Borel σ -algebra.*

The σ -algebra induced on $h^{-1}(g((0,1)^{T-1}))$ does not appear to yield a clean characterization, but we can show the weaker statement that μ is a Borel measure, i.e. it is defined on all open sets of $h^{-1}(g((0,1)^{T-1}))$.

Proposition 5.3. *μ is a Borel measure on $h^{-1}(g((0,1)^{T-1}))$.*

Computing θ

We now show $\theta = 2$ for the distribution μ as chosen above. We use the notation $d = d_{\delta,\gamma} := |\delta - \gamma|$ to denote the distance between δ and γ . Let \succsim_δ be the target preference relation.

First, note that since $Y_{\delta,\gamma}$ is distributed according to $\text{Bin}(T-1, d)$, $\mathbb{P}[E_{\delta,\gamma}^{\text{odd}}] = \frac{1-(1-2d)^{T-1}}{2}$ ¹⁰. We use this fact to derive an explicit description of the preference relations \succsim_γ contained in the ball $B(\succsim_\delta, R)$ in terms of d . We break the analysis up into a few cases.

First, when $R > \frac{1}{2}$, we have $\sup_{R > \frac{1}{2}} \frac{\mu(\text{Dis}(B(\succsim_\delta, R)))}{R} = 2$. Let $R \leq \frac{1}{2}$. By Lemma 5.1,

$$\succsim_\gamma \in B(\succsim_\delta, R) \iff \mathbb{P}[E_{\delta,\gamma}^{\text{odd}}] = \frac{1 - (1 - 2d)^{T-1}}{2} \leq R \quad (1)$$

Suppose $d \leq \frac{1}{2}$. Then $1 - 2d$ and $1 - 2R$ are both non-negative, so rearranging Equation (1) yields

$$d \leq \frac{1 - (1 - 2R)^{1/(T-1)}}{2}. \quad (2)$$

Suppose $d > \frac{1}{2}$, so $1 - 2d < 0$. Rearranging Equation (1), we get $1 - 2R \leq (1 - 2d)^{T-1}$. If $T - 1$ is odd, $(1 - 2d)^{T-1}$ is negative, so $1 - 2R \leq (1 - 2d)^{T-1}$ does not hold. Thus, when $T - 1$ is odd the ball consists of \succsim_γ such that d satisfies condition (2). If $T - 1$ is even, $(1 - 2d)^{T-1}$ is positive, so we get

$$d \geq \frac{1 + (1 - 2R)^{1/(T-1)}}{2}. \quad (3)$$

Thus, when $T - 1$ is even the ball consists of \succsim_γ such that d satisfies conditions (2) or (3).

Now, the disagreement region of $B(\succsim_\delta, R)$ consists of all points (x, y) such that the polynomial P_{x-y} (as defined in Lemma 5.1) has a root γ such that $\succsim_\gamma \in B(\succsim_\delta, R)$ (since we can find two hypotheses that disagree on (x, y) by taking a point slightly below γ and a point slightly above γ such that P_{x-y} has no sign changes in between). Hence, with $R_1 = \frac{1-(1-2R)^{1/(T-1)}}{2}$ and $R_2 = \frac{1+(1-2R)^{1/(T-1)}}{2}$, we have that

$$\mu(\text{Dis}(B(\succsim_\delta, R))) = \begin{cases} \mathbb{P}[E_{\delta-R_1, \delta+R_1}^{\geq 1}] & \text{if } T-1 \text{ is odd} \\ \mathbb{P}[E_{\delta-R_1, \delta+R_1}^{\geq 1} \cup E_{0, \delta-R_2}^{\geq 1} \cup E_{\delta+R_2, 1}^{\geq 1}] & \text{if } T-1 \text{ is even} \end{cases}$$

We have

$$\mathbb{P}[E_{\delta-R_1, \delta+R_1}^{\geq 1}] = 1 - (1 - 2R_1)^{T-1} = 2R,$$

and

$$\mathbb{P}[E_{\delta-R_1, \delta+R_1}^{\geq 1} \cup E_{0, \delta-R_2}^{\geq 1} \cup E_{\delta+R_2, 1}^{\geq 1}] = 1 - (2(R_2 - R_1))^{T-1} = 1 - 2^{T-1}(1 - 2R).$$

Therefore, when $T - 1$ is odd $\sup_{0 < R \leq 1/2} \frac{\mu(\text{Dis}(B(\succsim_\delta, R)))}{R} = 2$ and when $T - 1$ is even

$$\sup_{0 < R \leq 1/2} \frac{\mu(\text{Dis}(B(\succsim_\delta, R)))}{R} = \sup_{0 < R \leq 1/2} \frac{1 - 2^{T-1}(1 - 2R)}{R} = 2,$$

¹⁰This is due to the general fact that if X is a random variable distributed according to $\text{Bin}(n, p)$, the probability that X is odd is $\frac{1-(1-2p)^n}{2}$.

which is achieved at $R = 1/2$ since $\frac{1-2^{T-1}(1-2R)}{R}$ is increasing on $0 < R \leq \frac{1}{2}$.

Finally, $\theta = \sup_{R>0} \frac{\mu(\text{Dis}(B(\succ_{\delta}, R)))}{R} = 2$.

We have thus established Theorem 3.5:

Theorem 3.5. *There exists a distribution μ on $\mathbb{R}^T \times \mathbb{R}^T$ for which the disagreement coefficient of $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ is $\theta = 2$. Thus, for this distribution,*

$$\ell_{CAL}(\varepsilon) = \tilde{O} \left(\log T \log \frac{1}{\varepsilon} \right),$$

where the \tilde{O} notation suppresses terms that are logarithmic in $\log T$ and $\log 1/\varepsilon$.

5.3 Learning via membership queries

In this subsection, we present a simple membership queries algorithm that outperforms the guarantees provided by the disagreement based CAL algorithm. For simplicity, we restrict attention to preference models that are parametrized by a single parameter (e.g. $\mathcal{P}_{\mathcal{E}\mathcal{D}}$ and $\mathcal{P}_{\mathcal{H}\mathcal{D}}$).

Let $g_1, \dots, g_T : \mathbb{R} \rightarrow \mathbb{R}$ be a collection of functions such that there exist $1 \leq t_1, t_2 \leq T$ satisfying

1. $M := \sup_{\delta} \frac{g_{t_1}(\delta)}{g_{t_2}(\delta)}$ is finite, and
2. The map $\delta \mapsto \frac{g_{t_1}(\delta)}{g_{t_2}(\delta)}$ satisfies an inverse Lipschitz condition with constant C :

$$|\delta - \delta'| \leq C \left| \frac{g_{t_1}(\delta)}{g_{t_2}(\delta)} - \frac{g_{t_1}(\delta')}{g_{t_2}(\delta')} \right|.$$

Consider the model of preference relations \mathcal{P} parametrized by δ such that

$$x \succsim y \text{ if and only if } \sum_{t=1}^T g_t(\delta)x_t \geq \sum_{t=1}^T g_t(\delta)y_t.$$

Proposition 3.6. *There exists an algorithm that takes as input $\varepsilon > 0$ and using $O(\log 1/\varepsilon)$ membership queries outputs δ^h such that $|\delta - \delta^h| \leq \varepsilon$, where δ parametrizes the target preference relation in \mathcal{P} .*

Proof. Fix a $\rho > 0$ and an η -cover of $[0, M\rho]$, where $0 < \eta \leq \frac{\rho\varepsilon}{C}$. Let b_ρ be the quantity such that the agent is indifferent between receiving a payoff of ρ at time t_1 or receiving a payoff of b_ρ at time t_2 , i.e. b_ρ solves

$$g_{t_2}(\delta)b_\rho = g_{t_1}(\delta)\rho.$$

By running a binary search over the η -cover of $[\rho, M\rho]$, the analyst can find an approximation b_ρ^h to the indifference point for which $|b_\rho - b_\rho^h| \leq \eta$ (the binary search is performed on the parameter b_ρ^h by requesting labels for pairs of the form $(\rho e_{t_1}, b_\rho^h e_{t_2})$). The analyst then outputs the δ^h that solves $g_{t_2}(\delta^h)b_\rho^h = g_{t_1}(\delta^h)\rho$.

We have

$$|\delta - \delta^h| \leq C \left| \frac{g_{t_1}(\delta)}{g_{t_2}(\delta)} - \frac{g_{t_1}(\delta^h)}{g_{t_2}(\delta^h)} \right| = C \left| \frac{b_\rho}{\rho} - \frac{b_\rho^h}{\rho} \right| \leq \frac{C\eta}{\rho} \leq \varepsilon,$$

as desired.

Since $M := \sup_{\delta} \frac{g_{t_1}(\delta)}{g_{t_2}(\delta)}$ is finite, $b_{\rho} \leq M\rho$, so the binary search over the η -cover of $[0, M\rho]$ terminates. \square

Remark. Outputting a hypothesis parameter δ^h that is ε -close to δ is a reasonable measurement for the error of learning via membership queries since there is no underlying distribution providing points to the analyst. However, note that for a distribution on $\mathbb{R}^T \times \mathbb{R}^T$, a hypothesis close to the target parameter implies the set of misclassified points is assigned a small measure, due to continuity of measure.

The main feature of this algorithm is that its query complexity has no dependence on the number of time periods T . Both $\mathcal{P}_{\mathcal{ED}}$ and $\mathcal{P}_{\mathcal{HD}}$ fit the conditions of Proposition 3.6, and thus we obtain a large improvement over the guarantees provided by disagreement methods in the stream-based model. Such methods assume no extra knowledge about the problem domain and are written to fit a wide class of learning problems. When we are learning economic parameters, membership queries allow us to take advantage of the extra structure present in preference models.

Acknowledgements

We would like to thank Federico Echenique and Adam Wierman for several helpful comments and suggestions.

References

- [1] Balcan, M. F., Daniely, A., Mehta, R., Uner, R., & Vazirani, V. V. (2014, December). Learning economic parameters from revealed preferences. In International Conference on Web and Internet Economics (pp. 338-353). Springer, Cham.
- [2] Basu, P., & Echenique, F. (2018, February). Learnability and Models of Decision Making under Uncertainty. In Proceedings of the 2018 ACM Conference on Economics and Computation (pp. 53-53). ACM.
- [3] Beigman, E., & Vohra, R. (2006, June). Learning from revealed preference. In Proceedings of the 7th ACM Conference on Electronic Commerce (pp. 36-42). ACM.
- [4] Berns, G. S., Laibson, D., & Loewenstein, G. (2007). Intertemporal choice – toward an integrative framework. Trends in cognitive sciences, 11(11), 482-488.
- [5] Blume, L., Brandenburger, A., & Dekel, E. (1991). Lexicographic probabilities and choice under uncertainty. Econometrica: Journal of the Econometric Society, 61-79.
- [6] Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1989). Learnability and the Vapnik-Chervonenkis dimension. Journal of the ACM (JACM), 36(4), 929-965.
- [7] Chabris, C. F., Laibson, D. I., & Schuldt, J. P. (2010). Intertemporal choice. In Behavioural and Experimental Economics (pp. 168-177). Palgrave Macmillan, London.
- [8] Chambers, C. P., & Echenique, F. (2018). On multiple discount rates. Econometrica, 86(4), 1325-1346.

- [9] Cohn, D., Atlas, L., & Ladner, R. (1994). Improving generalization with active learning. *Machine learning*, 15(2), 201-221.
- [10] Dasgupta, S. (2011). Two faces of active learning. *Theoretical computer science*, 412(19), 1767-1781.
- [11] Echenique, F., Golovin, D., & Wierman, A. (2011, June). A revealed preference approach to computational complexity in economics. In *Proceedings of the 12th ACM conference on Electronic commerce* (pp. 101-110). ACM.
- [12] Grigoriev, D., & Vorobjov, N. (1988). Solving systems of polynomial inequalities in subexponential time. *J. Symb. Comput.*, 5(1/2), 37-64.
- [13] Hanneke, S. (2009). Theoretical foundations of active learning (No. CMU-ML-09-106). CARNEGIE-MELLON UNIV PITTSBURGH PA MACHINE LEARNING DEPT.
- [14] Hanneke, S. (2016). The optimal sample complexity of PAC learning. *The Journal of Machine Learning Research*, 17(1), 1319-1333.
- [15] Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature neuroscience*, 10(12), 1625.
- [16] Kalai, G. (2003). Learnability and rationality of choice. *Journal of Economic theory*, 113(1), 104-117.
- [17] Kleinberg, J., & Oren, S. (2014, June). Time-inconsistent planning: a computational problem in behavioral economics. In *Proceedings of the fifteenth ACM conference on Economics and computation* (pp. 547-564). ACM.
- [18] Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, 287-309.
- [19] Kreps, D. M. (1988). *Notes on the theory of choice* (Underground Classics in Economics). Westview Press Incorporated.
- [20] Phelps, E. S., & Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, 35(2), 185-199.
- [21] Stern, N., Peters, S., Bakhshi, V., Bowen, A., Cameron, C., Catovsky, S., ... & Edmonson, N. (2006). *Stern Review: The economics of climate change* (Vol. 30, p. 2006). London: HM treasury.
- [22] Zadimoghaddam, M., & Roth, A. (2012, December). Efficiently learning from revealed preference. In *International Workshop on Internet and Network Economics* (pp. 114-127). Springer, Berlin, Heidelberg.

A Properties of μ and μ^*

Proof of Proposition 5.2. Let \mathcal{B} denote the Borel σ -algebra on $(0, 1)^{T-1}$.

Let $X = \mathbb{R}^T / \sim$ endowed with the quotient topology, and $g^* : [0, 1]^{T-1} \rightarrow X$ be the map $g^*(r_1, \dots, r_{T-1}) = [\prod(x - r_i)]$. Explicitly, the terms of $g^*(r_1, \dots, r_{T-1})$ are given by symmetric sums:

$$g^*(r_1, \dots, r_{T-1}) = \left(c, -c \sum_i r_i, c \sum_{i,j} r_i r_j, \dots, (-1)^{T-1} c r_1 \cdots r_{T-1} \right),$$

where c is the appropriate constant for the representative of the equivalence class. Each symmetric sum is a continuous function of $T-1$ variables, so g^* is continuous. Moreover, note that g^* is injective. Then, with $Y = g^*([0, 1]^{T-1})$, we have that $g^* : [0, 1]^{T-1} \rightarrow Y$ is a continuous bijection from a compact set into a Hausdorff space. Hence, g^* is a homeomorphism. Then g , which is the restriction of g^* to $(0, 1)^{T-1}$ is a homeomorphism onto $Z := g((0, 1)^{T-1})$. Thus, the σ -algebra $g(\mathcal{B})$ that we obtain on Z is the Borel σ -algebra. \square

When $T = 2$, we can give an explicit description of μ^* . Identify \mathbb{R}^2 / \sim with the unit circle. Then, a degree 1 polynomial P is identified with the point $(\cos \theta, \sin \theta)$, where $P(x) = (\cos \theta)x + \sin \theta$. $Z := g((0, 1)^{T-1})$ consists of the boundary of the unit circle for which the argument θ satisfies $-\tan \theta \in (0, 1)$. This is satisfied precisely for $\theta \in (3\pi/4, \pi) \cup (7\pi/4, 2\pi)$. Hence, if U is a basic open subset of $\{(\cos \theta, \sin \theta) : \theta \in (3\pi/4, \pi) \cup (7\pi/4, 2\pi)\}$, we can write $U = \{(\cos \theta, \sin \theta) : \theta_1 < \theta < \theta_2\}$ with θ_1, θ_2 both in the same segment of the unit circle and

$$\mu^*(U) = \mu^{**}(\{-\tan \theta : \theta_1 < \theta < \theta_2\}) = |\tan \theta_1 - \tan \theta_2|.$$

Proof of Proposition 5.3. Let $h : \mathbb{R}^T \times \mathbb{R}^T \rightarrow \mathbb{R}^T / \sim$ be the map $h(x, y) = [x - y]$. Let $V \subseteq h^{-1}(g((0, 1)^{T-1}))$ be open. We show that

$$h(V) = \{[z] : \exists(x, y) \in V(z = x - y)\}$$

is open. Indeed, let $z = x - y$ for $(x, y) \in V$ and choose ε small enough such that the square with vertices $\{(x + \varepsilon, y + \varepsilon), (x + \varepsilon, y - \varepsilon), (x - \varepsilon, y + \varepsilon), (x - \varepsilon, y - \varepsilon)\}$ is contained in V . Then, for any $\lambda \leq \varepsilon$, $[z + \lambda] = [(x + \lambda) - y]$ with $(x + \lambda, y) \in V$ and $[z - \lambda] = [x - (y + \lambda)]$ with $(x, y + \lambda) \in V$, so in particular the open ball with radius λ centered at $[z]$ is contained in $h(V)$. \square