Approximation methods in inductive inference

William R. Moser

University of Florida, 358 Little Hall, Gainesville, FL 32611, USA
and Metwave Communications, Redmond, WA 98052, USA

Abstract

In many areas of scientific inquiry, the phenomena under investigation are viewed as functions on the real numbers. Since observational precision is limited, it makes sense to view these phenomena as bounded functions on the rationals. One may translate the basic notions of recursion theory into this framework by first interpreting a partial recursive function as a function on \( \mathbb{Q} \). The standard notions of inductive inference carry over as well, with no change in the theory.

When considering the class of computable functions on \( \mathbb{Q} \), there are a number of natural ways in which to define the distance between two functions. We utilize standard metrics to explore notions of approximate inference – our inference machines will attempt to guess values which converge to the correct answer in these metrics. We show that the new inference notions, \( NV_\infty, EX_\infty, \) and \( BC_\infty \), infer more classes of functions than their standard counterparts, \( NV, EX, \) and \( BC \). Furthermore, we give precise inclusions between the new inference notions and those in the standard inference hierarchy. We also explore weaker notions of approximate inference, leading to inference hierarchies analogous to the \( EX'' \) and \( BC'' \) hierarchies. Oracle inductive inference is also considered, and we give sufficient conditions under which approximate inference from a generic oracle \( G \) is equivalent to approximate inference with only finitely many queries to \( G \).

© 1998 Elsevier Science B.V. All rights reserved.

AMS classification: 03D20, 68T05

Keywords: Inductive inference

1. Introduction

1.1. Basic models of inductive inference

In the experimental sciences, it is typical to make generalizations from a finite set of data, amending such generalizations as needed to account for new data. Often, the data are collected at discrete time intervals and so can be viewed as ordered in a natural way. Also, the "experiment" need never end – we may continue to record a new datum during each time interval and make a new generalization at this time. It is this situation which we take as our starting point. As a further simplification we
view such sequences of data as functions on the natural numbers. In the next two sections, we give an informal account of the basic models of inductive inference, and some common variants. See Popper [19] for a philosophical perspective on notions of inductive inference. A more detailed mathematical treatment, giving the formal definitions and stating the major results of inductive inference, is found in Section 2.

Gold [11] was among the first to formalize an example-based theory of learning. The two basic inductive inference paradigms we informally describe below are essentially due to him. The reader is referred to the excellent survey [18] by Odifreddi for a detailed introduction to the subject of inductive inference from a recursion theoretic perspective. In the simplest inference model, we wish to define a procedure which works as a predictor for a given class of phenomena. Given a sequence of data (a partial record of a phenomenon in the class under consideration), we wish to predict what will occur next. In this type of inference, called next-value inference, we use the sequence of data \( \langle d_0, \ldots, d_{n-1} \rangle \) collected prior to time \( n \) to predict the value of the datum \( d_n \) which we will collect at time \( n \). In view of Church's thesis, we will assume that we must do so algorithmically. That is, our inference procedure must be a recursive function — we may not use coin flips, oracle consultations, appeals to higher beings, etc. in our attempt to predict the next datum.

We must now consider the accuracy requirements of our prediction algorithm. At time \( t = 0 \), we are hardly in a position to make a reasonable guess for \( d_0 \). If the underlying phenomenon is complex, we may not be in a position to make a reasonable guess for \( d_N \) at time \( t = N \), even for large values of \( N \). We will thus allow our prediction algorithm to make some mistakes, but we will require of the algorithm that it is eventually always accurate. In other words, for an algorithm to be considered predictive, it must make only finitely many errors in predicting the data. Henceforth, if a class \( \mathcal{C} \) of phenomena is predicted by an algorithm \( M \), we will say that \( M \) \( NV \)-identifies \( \mathcal{C} \), or that \( \mathcal{C} \in NV \).

Notice that this restriction on the error rate of our prediction algorithm essentially forces us to assume that the phenomena we are attempting to predict are also algorithmic, since our prediction machine eventually outputs the same values as those observed experimentally. Thus, we impose the further restriction on our class of phenomena that they may be viewed as recursive functions; hence, our "world" is the class of recursive functions.

Since we now view all phenomena as "black boxes" implementing algorithms of some sort, it is natural to use the data gathered from a given "black box" to attempt to guess its underlying algorithm. Instead of simply guessing the next output, we wish to explain what is going on inside the box. We still have only the data stream to work with; so, at time \( n \) we will again use the data gathered prior to this time to infer the underlying rule. As with the predictive machine, we allow the explanatory machine to make finitely many mistakes, as long as it eventually settles on an algorithm for each phenomenon in the class it explains. If a class \( \mathcal{C} \) of recursive functions is explained by an algorithm \( M \), we will say that \( M \) \( EX \)-identifies \( \mathcal{C} \), or that \( \mathcal{C} \in EX \).
1.2. Variants of the basic models

In a scientific explanation of the phenomenon underlying observed data, it is usually the case that the explanation is consistent with the data. In view of this, we may require of our explanatory machine that the algorithm output at stage $n$ agrees with the data given up to this time. In addition, the unsolvability of the Halting problem tells us that it is in general impossible to determine whether our output algorithms are total. Thus, it is also natural to allow our explanatory machine to output algorithms which are consistent with the available data at time $n$, but which may not return an answer for times $m > n$ if we attempt to use the algorithm to predict such future values. If we denote the family of phenomena classes which can be explained in this manner by $EX_{cons}$, then $EX_{cons}$ strictly contains $NV$. If, however, we restrict our explanatory machine to output only total algorithms, then the two families are equal.

Many other variations on the basic model have been explored. Blum and Blum [4] require that if an explanatory machine eventually stabilizes on an algorithm, then this algorithm accurately describes the phenomenon. This type of inference is known as identification by reliable explanation, and the family of identified classes as $EX_{rel}$. $EX_{rel}$-identification is more general than $EX_{cons}$-identification, but more restrictive than $EX$-identification. Feldman [9] requires only that the explanatory machine eventually output algorithms which are extensionally identical; the same algorithm need not appear twice, but the algorithms output should all produce identical outputs on identical inputs. This type of inference is known as behaviorally correct identification, and it subsumes $EX$-identification.

One may further relax the requirements for an explanatory machine by accepting explanations which come “close” to describing the observed phenomenon. Specifically, we may allow the algorithm upon which the explanatory machine finally settles to differ finitely from the observed data. Furthermore, we may require a uniform upper bound on the number of errors allowed, or we may be a bit less restrictive, settling for identification with arbitrary finite errors. Case and Smith [6] show that the hierarchies of the induced inference families are proper and that behaviorally correct identification with arbitrary finite errors is powerful enough to identify all the recursive functions.

Many other variants appear in the literature. For example, Royer [23] defines notions of probabilistic inference, and Case et al. [5] define notions of limiting inference which are quite different from the notions of approximate inference explored in the sequel.

1.3. Approximate inference

Not all scientific inquiry requires such strong notions of inference. Often, the phenomenon under investigation is assumed to be a bounded continuous function on the real numbers. After receiving only a finite number of datum, the phenomenon is then represented by a function interpolating the known data. While the interpolant may equal the underlying function at only a few points, it is nevertheless viewed as a reasonable
representation of this function. In the sections that follow, we will formalize notions of inference arising from this point of view.

In Section 3, we introduce the concept of inference of a recursive rational-valued function and define an important subclass of these functions, \textit{RUC}, which cannot be inferred by the standard inference methods, but for which a natural approximation technique exists. This technique leads to new "approximate" inference paradigms which subsume the standard paradigms and allow us to easily infer the class \textit{RUC}. In Section 4, we define hierarchies of approximate inference classes, the "epsilon" inference classes, which enable us to infer more classes than with the approximate inference classes, while retaining the finitistic flavor of the standard inference notions. In Section 5, we consider approximate inference using generic sets as oracles. Finally, we conclude in Section 6 with some further directions for research in the area of approximate inference.

2. Preliminaries

2.1. Notions from recursion theory

Let $I_Q = [0, 1] \cap \mathbb{Q}$. We assume that \{q_0, q_1, \ldots, q_n, \ldots\} is a fixed 1–1 effective enumeration of $I_Q$ such that $q_0 = 0$, and $q_1 = 1$. $[I_Q]_{<\omega}^\omega$ will denote the set of finite sequences of rationals. An arbitrary (partial) recursive function can be viewed as a function on $I_Q$, by interpreting the natural number $m$ as the $m$th rational in the enumeration \{q_n\}_{n \in \omega}. More formally, we could define the rational interpretation $f_Q$ of a (partial) recursive $f$ by $f_Q(q_m) = q_n$ iff $f(m) = n$. In this framework, we could carry out all basic recursion theory, proving the enumeration theorem, the SF and fixed point theorems, and so forth. In the sequel, we will dispense with most of this formalism, preferring to use notation to indicate the types of functions under consideration.

Let $\langle a_0, \ldots, a_n \rangle$ denote the usual coding of a finite sequence of natural numbers (or rationals) by a natural number. For a function $f$ and natural number $n$, $f|_n$ denotes $\langle f(0), \ldots, f(n) \rangle$ (or $\langle f(q_0), \ldots, f(q_n) \rangle$). We denote finite binary sequences by $\sigma, \tau$. In this context, $|\sigma|$ denotes the length of the $\sigma$. If $\sigma$ is an initial segment of $\tau$ (or $B \in 2^\omega$), we denote this by $\sigma < \tau$ ($\sigma \prec B$). We will use the following to denote intervals in $2^\omega: I(\sigma) = \{B \in 2^\omega \mid \sigma \prec B \}$.

$\{\phi_0, \phi_1, \ldots, \phi_n, \ldots\}$ enumerates the partial recursive functions, and $\Phi$ denotes an arbitrary partial recursive function. $\{\Phi_0, \Phi_1, \ldots, \Phi_n, \ldots\}$ enumerates the corresponding functions on $I_Q$, with $\Phi$ denoting an arbitrary partial recursive function with domain and range contained in $I_Q$. $M$ (with adornments) will denote a (standard) recursive Turing machine, thought of as a machine which infers a class of functions. Such machines $M$ will usually be used in the context of the standard inference classes, operating on sequences of natural numbers, and producing natural numbers as outputs. We will also use them in the context of standard inference of functions on $I_Q$. We will often use the
notation "\(\forall x\)" in this context to mean "for all but finitely many \(x\)". We will use the symbol \(G\) for two different purposes. In the context of defining approximate inference machines, \(G\) (with adornments) denotes a recursive Turing machine with inputs from \([I_{\mathbb{Q}}]<^{\omega}\) and outputs in \(I_{\mathbb{Q}}\) or \(\mathbb{N}\), depending on whether its guesses are to be function values, or indices of functions. In the context of oracle inductive inference, \(G\) will be used to denote a generic set.

\(REC\) will denote the standard (total) recursive functions (i.e. those recursive on \(\mathbb{N}\)). \(QREC\) will denote the class of total recursive functions from \(I_{\mathbb{Q}}\) to \(I_{\mathbb{Q}}\); we will also call these the recursive rational-valued functions. We shall denote the subclass of recursive \(\{0,1\}\)-valued functions by \(QSET\). \(RC\) denotes the recursive rational functions which are continuous on \(I_{\mathbb{Q}}\), and \(RUC\) denotes those which are uniformly continuous on \(I_{\mathbb{Q}}\). Since \(I_{\mathbb{Q}}\) is neither compact nor connected, continuity does not imply uniform continuity, so that \(RUC \subseteq RC\). Note that we may embed \(REC\) into \(RUC\) by mapping \(f\) to \(f_{\mathbb{Q}}\), and then mapping \(f_{\mathbb{Q}}\) to \(\hat{f}_{\mathbb{Q}}\) defined by \(\hat{f}_{\mathbb{Q}}(0) = 0\), and for \(n > 0\), \(\hat{f}_{\mathbb{Q}}(1/n) = f_{\mathbb{Q}}(q_{n-1})/n\), linearly interpolating these values for other rationals.

See [17, 21, 25] for developments of basic recursion theory. All notions from measure theory, real analysis, and topology can be found in [12, 22, 16], respectively.

2.2. Inductive inference

We give a brief summary of the standard inductive inference classes and the theorems relating them. See [6] for a thorough treatment of these notions. The three basic notions of inductive inference are as follows:

**Definition 1.** A class \(\mathcal{C}\) of recursive functions is next-value identifiable (\(\mathcal{C} \in NV\)) if there is an \(M : [\omega]<^{\omega} \rightarrow \omega\) such that for every \(f \in \mathcal{C}\),

\[
\forall n \quad f(n) = M(\langle f(0), \ldots, f(n-1) \rangle).
\]

Thus, an \(NV\)-inference machine \(M\) tries to correctly guess the sequence \(\{f(0), f(1), \ldots\}\), making only finitely many errors. We also say that \(M\) \(NV\)-identifies \(\mathcal{C}\), and that \(\mathcal{C} \in NV\) via \(M\).

**Definition 2.** A class \(\mathcal{C}\) of recursive functions is explanatorily identifiable (\(\mathcal{C} \in EX\)) if there is a \(M : [\omega]<^{\omega} \rightarrow \omega\) so that for every \(f \in \mathcal{C}\), there is an index \(e\) of \(f\) such that

\[
\forall n \quad M(\langle f(0), \ldots, f(n-1) \rangle) = e.
\]

Thus, an \(EX\)-inference machine \(M\) tries to settle on a correct index for the input \(f\), making only finitely many errors. Gold [11] also introduces the notion of explanatory consistency:
**Definition 3.** A class $\mathcal{C}$ of recursive functions is identifiable by consistent explanation ($\mathcal{C} \in EX_{cons}$) if $\mathcal{C} \in EX$ via $M$, and for all $n$,

$$\phi_M((f(0), \ldots, f(n-1)))|_n = f|_n.$$  

Feldman [9] further weakens the notion of $EX$-inference as follows:

**Definition 4.** A class $\mathcal{C}$ of recursive functions is behaviorally correctly identifiable ($\mathcal{C} \in BC$) if there is a $M : [\omega]^{<\omega} \rightarrow \omega$ so that for every $f \in \mathcal{C}$,

$$\forall n \phi_M((f(0), \ldots, f(n-1))) = f.$$  

Thus, a $BC$-inference machine $M$ must, after finitely many stages, output only indices of the input function $f$. However, unlike the $EX$ case, we may output any index of $f$, infinitely often changing our mind about which index to output.

The notions $PEX$ and $PBC$ are similar to $EX$ and $BC$, except that $M$ may only output indices of total functions. Odifreddi [18] credits Barzdin with observing that $NV = PEX$. Gold [11] shows that $NV \nsubseteq EX_{cons}$, Blum and Blum [4] show that $EX_{cons} \nsubseteq EX$, and Barzdin [3] shows that $EX \nsubseteq BC$. Thus, we have the following hierarchy of inference notions.

**Proposition 5.** $NV = PEX \nsubseteq EX_{cons} \nsubseteq EX \nsubseteq BC$.

Blum and Blum [4], and Case and Smith [6] introduce the concept of inference with anomalies:

**Definition 6.** If $f$ is total, we say that $\phi_e$ is an $n$-variant of $f$ ($\phi_e \simeq^n f$) if

$$| \{ x | \phi_e(x) \uparrow \text{ or } \phi_e(x) \downarrow \neq f(x) \} | \leq n.$$  

We say simply that $\phi_e$ is a variant of $f$ ($\phi_e \simeq^* f$) if $\phi_e$ is an $n$-variant of $f$ for some $n$.

**Definition 7.** A class $\mathcal{C}$ of recursive functions is explanatorily identifiable with $n$ errors ($\mathcal{C} \in EX^n$) if there is an $M : [\omega]^{<\omega} \rightarrow \omega$ so that for every $f \in \mathcal{C}$, there is an index $e$ such that $\phi_e \simeq^n f$, and

$$\forall n M((f(0), \ldots, f(n-1))) = e.$$  

**Definition 8.** A class $\mathcal{C}$ of recursive functions is explanatorily identifiable with finitely many errors ($\mathcal{C} \in EX^*$) if there is an $M : [\omega]^{<\omega} \rightarrow \omega$ so that for every $f \in \mathcal{C}$, there is an index $e$ such that $\phi_e \simeq^* f$, and

$$\forall n M((f(0), \ldots, f(n-1))) = e.$$  

Note that $EX^0 = EX$. Case and Smith [6] show the following:
Proposition 9. For all $n$, 
- $EX^n \subseteq EX^*$ (in fact, $\cup_n EX^n \subseteq EX^*$), and 
- $EX^n \subseteq EX^{n+1}$.

Case and Smith [6] also show that $EX^* \subseteq BC$. They also extend the concept of inference with anomalies to $BC - BC^n$ and $BC^*$ are defined similarly to their $EX$-counterparts. The previous proposition also holds for $BC^n$ and $BC^*$. Odifreddi [18] credits Harrington with showing that $REC \in BC^*$.

We thus have the following hierarchy of inference classes:

$$NV = PEX \subseteq EX_{cons} \subseteq EX \subseteq EX^1 \subseteq EX^2 \subset \ldots \subset \bigcup_n EX^n \subseteq EX^* \subseteq BC \subseteq BC^1 \subseteq BC^2 \subset \ldots \subset \bigcup_n BC^n \subseteq BC^*.$$ 

In the sequel, we shall see that the notions of approximate inference yield a hierarchy which is no longer linear.

2.3. Inference with oracles

Fortnow et al. [10], and Kummer and Stephan [14] examine the concept of oracle inductive inference, in which the inference machine $M$ is allowed queries to an arbitrary set $A \subseteq \mathbb{N}$.

Definition 10. We say that an oracle Turing machine (OTM) $M^{(i)}$ is categorical if for every $A \subseteq \mathbb{N}$, $M^A$ is total (in the sequel, we shall usually identify such $A$ with its characteristic function).

We give the definitions of $NV[A], NV[A^*]$ here. The definitions for $EX[A], EX[A^*], BC[A],$ and $BC[A^*]$ are obtained from the definitions of $EX$ and $BC$ similarly.

Definition 11. Let $A \subseteq \mathbb{N}$. A class $C$ of recursive functions is next-value identifiable from $A$ ($C \in NV[A]$) if there is a categorical OTM $M^{(i)}$ so that for every $f \in C$,

$$\forall n \ f(n) = M^A((f(0), \ldots, f(n-1))).$$

Definition 12. Let $A \subseteq \mathbb{N}$. A class $C$ of recursive functions is next-value identifiable from $A$ with finitely many queries ($C \in NV[A^*]$) if there is a categorical OTM $M^{(i)}$ so that $C \in NV[A]$ via $M^A$, and for each $f \in C$, there is an $n$ so that $f$ is inferred by $M^A^{(i)(0 \ldots n-1)}$.

For oracle inference machines, the fundamental task is to determine which oracles yield no increase in inference power and which ones allow us to infer all of $REC$. 
Definition 13. We say that $A$ is $NV$-trivial if $NV[A] = NV$, and $NV$-omniscient if $REC \in NV[A]$. We define these notions for other inference criteria similarly.

$K$ denotes the Halting problem, $K = \{(x, y) \mid \phi_x(y) \downarrow\}$. It is well-known that $K$ is $EX$-omniscient. As another example, consider

$$TOT = \{e \mid \forall y \phi_e(y) \downarrow\}.$$ 

It is easy to see that $REC \in EX[TOT^*]$.

Definition 14. $G \in 2^\omega$ is generic if for each $\Sigma_1$ $W \subseteq \{0, 1\}^*$, either
- $(\exists \sigma < G) (\forall \tau \geq \sigma) \tau \notin W$, or
- $(\exists \sigma < G) \sigma \in W$.

Following Fortnow et al. [10], we use the notation $\mathcal{G}(A)$ to mean that either $A$ is recursive, or $A \leq_T K$ and $A \equiv_T G$ for some generic $G$.

It has recently been shown that $EX[A] = EX \iff \mathcal{G}(A)$ and $BC[A] = BC \iff \mathcal{G}(A)$. The forward implication in both cases is proved in Fortnow et al. [10]. Slaman and Solovay [24] were the first to prove the reverse implication for the $EX$ case. Kummer and Stephan [14] give an easier proof of the reverse implication for the $EX$ case, and also show the reverse implication for the $BC$ case.

3. Approximate inference of recursive functions

3.1. Notions of approximate inference

We are now in a position to define notions of inference, $NV_\infty$, $EX_\infty$, and $BC_\infty$, which are in a sense “continuous” analogues of the standard notions $NV$, $EX$, and $BC$. The idea for these new notions comes from the following situation. Suppose that $f : [0, 1] \to [0, 1]$ is continuous, and we are allowed to ask “What is $f(x)$?” for (only) countably many $x$. We may then ask, in this manner, for $f(0), f(1), f(1/2), f(1/3), f(2/3)$, and so on. After we ask for the $n$th value, we can then form an approximation $f_n$ to the graph of $f$, perhaps by splines, or by a polynomial interpolation. Since we know that $f$ is continuous, we know that the sequence of continuous functions $\{f_n\}$ converges uniformly to $f$. Thus, we have a procedure whereby we can build reasonable approximations to $f$ in stages, with the knowledge that, “in the limit”, we recover $f$ itself. With this procedure in mind, we define the approximate inference classes.

Definition 15. A class $\mathcal{C}$ of recursive rational functions is next-value approximable ($\mathcal{C} \in NV_\infty$) if there is a $G : [I_\mathcal{Q}]^{<\omega} \to I_\mathcal{Q}$ such that for every $f \in \mathcal{C}$

$$\lim_{n \to \infty} |f(q_n) - G((f(q_0), f(q_1), \ldots, f(q_{n-1})))| = 0.$$
This definition compares to standard $NV$-inference, as applied to rational functions, where $G$ infers $f$ only if
\[ f(q_n) = G((f(q_0), f(q_1), \ldots, f(q_{n-1}))) \]
for all but finitely many $n$.

**Definition 16.** A class $\mathcal{C}$ of recursive rational functions is *explanatorily approximable* ($\mathcal{C} \in \text{EX}_\infty$) if there is a recursive $G: [I_Q]^{<\omega} \to \mathbb{N}$ so that for every $f \in \mathcal{C}$, there is an index $e$ such that
- $\Phi_e$ is total,
- for all but finitely many $n$, $G((f(q_0), \ldots, f(q_n))) = e$, and
- $\lim_{n \to \infty} |f(q_n) - \Phi_e(q_n)| = 0$.

Thus, an $\text{EX}_\infty$-inference machine behaves like an $\text{EX}$-inference machine, except that the index it settles on must only approximate the input $f$ in the $NV_\infty$ sense. The definition for $\mathcal{C} \in \text{PEX}_\infty$ is similar, but $G$ must output only indices for total functions.

**Definition 17.** A class $\mathcal{C}$ of recursive rational functions is *behaviorally correctly approximable* ($\mathcal{C} \in \text{BC}_\infty$) if there is a recursive $G: [I_Q]^{<\omega} \to \mathbb{N}$ such that for every $f \in \mathcal{C}$,
\[ \lim_{n \to \infty} \| f - \Phi_{G((f(q_0), f(q_1), \ldots, f(q_n)))} \|_\infty = 0, \]
where $\| \Phi \|_\infty = \sup_{x \in I_Q} |\Phi(x)|$ if $\Phi$ is total, and $\| \Phi \|_\infty = 1$ if $\Phi$ is partial. The definition for $\mathcal{C} \in \text{PBC}_\infty$ is similar, but $G$ must output only indices for total functions.

This definition compares to standard $BC$-inference, as applied to rational functions, where $G$ infers $f$ only if
\[ f = \Phi_{G((f(q_0), f(q_1), \ldots, f(q_{n-1})))} \]
for all but finitely many $n$. We recall that the more restrictive notion of $EX$-inference requires further that there is a fixed $e$ such that $G((f(q_0), f(q_1), \ldots, f(q_{n-1}))) = e$ for all but finitely many $n$. The less restrictive notion of $BC^*$-inference allows $G$ to infer $f$ if, for all but finitely many $n$,
\[ f(x) = \Phi_{G((f(q_0), f(q_1), \ldots, f(q_{n-1})))}(x) \]
except on a finite set. We note again that the class $\mathcal{QREC}$ is in $BC^*$, but that neither $\mathcal{QREC}$ nor $\mathcal{QSET}$ are in $BC$.

An obvious way to weaken the above definition is to require only that a $BC_\infty$ machine $G$ output guesses $\{f_n\}$ so that the $f_n$'s converge pointwise to the input function $f$. The modified inference notion is then equivalent to $BC^*$, for if $f \in \mathcal{QREC}$, we approximate $f$ at stage $n$ by $f_n$, defined by
\[ f_n(q_k) = \begin{cases} f(q_k), & k < n, \\ 0 & \text{otherwise}. \end{cases} \]
Note that by definition of \( NV \) (resp. \( EX \), \( BC \)), we have \( NV \subseteq NV_{\infty} \) (resp. \( EX \subseteq EX_{\infty} \), \( BC \subseteq BC_{\infty} \)).

3.2. Relationships among the inference classes

The results of this section determine the lattice of inclusions among the inference classes \( NV \), \( EX \), \( BC \), \( NV_{\infty} \), \( EX_{\infty} \), \( PE_{\infty} \), \( BC_{\infty} \), and \( PBC_{\infty} \). As was previously mentioned, the standard inference notions \( NV \), \( EX \), and \( BC \) can easily be translated into inference notions about the recursive rational functions. We shall avoid this formalism, taking for example the assertion "\( NV \subseteq NV_{\infty} \)" to mean that the analogue of \( NV \) (in the framework of functions on \( I_{Q} \)) is contained in \( NV_{\infty} \).

It is easy to see that restricting approximate-inference machines to output only indices of total functions results in strictly less powerful inference notions:

**Theorem 18.** \( PEX_{\infty} \subseteq EX_{\infty} \) and \( PEX_{\infty} \subseteq PBC_{\infty} \).

**Proof.** \( PEX_{\infty} \subseteq EX_{\infty} \) is trivial. To see that \( PEX_{\infty} \subseteq PBC_{\infty} \), let \( C \in PEX_{\infty} \) via \( G \), and define a \( PBC_{\infty} \)-inference machine \( \hat{G} \) as follows: at stage \( n \), \( \hat{G} \) outputs an index for \( \phi_{G}(f(q_0),f(q_1),...,f(q_{n-1})) \) patched with the values \( \{(q_0,f(q_0)),..., (q_{n-1},f(q_{n-1}))\} \). Then

\[
\lim_{n \to \infty} \| f - \phi_{\hat{G}}(f(q_0),f(q_1),...,f(q_{n-1})) \|_{\infty} = 0,
\]

so that \( f \) is \( PBC_{\infty} \)-inferred. \( \square \)

This proof yields the following:

**Corollary 19.** \( EX_{\infty} \subseteq BC_{\infty} \).

**Theorem 20.** \( PBC_{\infty} \subseteq BC_{\infty} \) and \( PBC_{\infty} \subseteq NV_{\infty} \).

**Proof.** \( PBC_{\infty} \subseteq BC_{\infty} \) is trivial. To see that \( PBC_{\infty} \subseteq NV_{\infty} \), let \( C \in PBC_{\infty} \) via \( G \), and let \( f \in C \). Since \( G \) must output an index of a total function for any input sequence, \( |f(q_n) - \phi_{G}(f(q_0),f(q_1),...,f(q_{n-1})))| \) is defined for any \( n \). Furthermore,

\[
\lim_{n \to \infty} |f(q_n) - \phi_{G}(f(q_0),f(q_1),...,f(q_{n-1})))| = 0,
\]

so that \( f \) is \( NV_{\infty} \)-inferred. Note that this proof does not show that \( BC_{\infty} \subseteq NV_{\infty} \) -- \( BC_{\infty} \)-machines \( G \) are not required to output indices of total functions. \( \square \)

Interpolation of known function values is a natural inference procedure in the setting of recursive rational functions, and so we have

**Theorem 21.** \( RUC \in PBC_{\infty} \).
**Proof.** Let $f \in RUC$. At stage $n$, having received $(f(q_0), f(q_1), \ldots, f(q_{n-1}))$, simply output (an index of) the linear interpolation $f_n$ of these $n$ values. The functions $\{f_n\}$ are uniformly continuous on $I_Q$, and converge pointwise to $f$, therefore $f_n \to f$ uniformly. \(\square\)

By the above three theorems, we have

**Corollary 22.** $RUC \in BC_{\infty}$, $RUC \in NV_{\infty}$.

However, the interpolation method cannot be used to $EX_{\infty}$-infer $RUC$, since it requires changing indices infinitely often. In fact, there is no algorithm which will $EX_{\infty}$-infer $RUC$. The idea of the proof is essentially the same as that in the proof that $REC \notin EX$.

**Theorem 23.** $RUC \notin EX_{\infty}$.

**Proof.** Suppose $G$ $EX_{\infty}$-infers $RUC$. We construct an $f \in RUC$ on which $G$ changes its mind infinitely often, for a contradiction. Let $\{a_i\}$ be an increasing, computable sequence of rationals with limit 1, and set $f(0) = f(1) = 0$. We use $G$ to define $f$ on all of $I_Q$ as follows. At stage $n$, if $q_n < a_n$, do nothing. Otherwise, $q_n \in [a_k, a_{k+1})$ for some $k \geq n$. Extend $f$ in two ways: $f_1$ will be $f$ extended by 0, and $f_2$ is $f$ extended with a "hat" of height $1/k$ in $[a_k, a_{k+1})$. Both $f_1$ and $f_2$ are in $RUC$, and differ by more than $1/2k$ on a positive interval. Thus, there is a stage greater than $n$ at which $G$ outputs different indices on $f_1$ and $f_2$. Extend $f$ by whichever of $f_1$, $f_2$ forces $G$ to output a different index than $G$ output on $f$ at stage $n-1$. \(\square\)

As a corollary to Theorems 19 and 23, we have

**Corollary 24.** $EX_{\infty} \subseteq BC_{\infty}$.

Thus, we see that $RUC$ plays two crucial roles in approximate inference. It is foremost a natural class of functions which motivates the idea of approximate inference. It also separates inference notions in a natural way – it is not a construct devised solely to prove a theorem. In addition, although $PEX \neq PBC$, we have shown

**Corollary 25.** $PEX_{\infty} \subseteq PBC_{\infty}$.

The interpolation procedure used above also does not suffice to infer an arbitrary $RC$ function. We show that $RC$ is not $BC_{\infty}$- or $NV_{\infty}$-inferable. We will need a way for an inference machine $G$ to use its own guesses as inputs:

**Definition 26.** For any $G, \sigma = (a_0, \ldots, a_{m-1})$, we define $T_{G, \sigma}$ on $I_Q$ by recursion as follows:

$$T_{G, \sigma}(q_n) = \begin{cases} G((a_0, \ldots, a_{n-1})) & \text{if } n < m, \\ G((a_0, \ldots, a_{m-1}, Rnd(T_{G, \sigma}(q_m)), \ldots, Rnd(T_{G, \sigma}(q_{n-1})))) & \text{otherwise}, \end{cases}$$
where
\[ Rnd(x) = \begin{cases} 
0 & \text{if } x < 1/2, \\
1 & \text{otherwise.} 
\end{cases} \]

**Theorem 27.** $RC \notin NV_\infty$. 

**Proof.** By way of contradiction, we suppose that $RC \in NV_\infty$ via $G$ and show that we can construct a $\hat{G}$ to $NV$-infer $QSET$. Let $0 < r < 1$ be any computable irrational, and let $\{r_i\}$ be any increasing computable sequence of computable irrationals in $[0, 1]$ with limit $r$. For $A \in QREC$, define $\hat{A}$ by
\[ \hat{A}(q) = \begin{cases} 
1 & \text{if } q \in (r_i, r_{i+1}) \text{ and } A(q_i) > 1/2, \\
0 & \text{otherwise.} 
\end{cases} \]

Note that since the $\{r_i\}$ are irrational, $\hat{A}$ is continuous on $IQ$ (i.e. $\hat{A} \in RC$), although not uniformly so. Construct $\hat{G}$ as follows. On input $\langle f(q_0), \ldots, f(q_{n-1}) \rangle$ (from $f \in QREC$), output $T_{\hat{G}, i}(q_0, \ldots, q_{n-1})$ for the largest $m$ such that we can compute $\hat{f}|_m$ from $f|_n$. If $A \in QSET$, then $A \in RC$, so by hypothesis, $G$ outputs guesses within $1/3$, say, of $A(q_n)$ for $n$ greater than some $N$. Let $M$ denote the least natural number greater than $N$ such that $\hat{A}|_N$ can be computed from $A|_M$. Then if $A \in QSET$, for stages $n > M$, $\hat{G}$ outputs $A(q_n)$.

The notions $BC_\infty$, $EX_\infty$, and $NV_\infty$ are the $IQ$-domain generalizations of $BC$, $EX$, and $NV$, as we show in the next few theorems. Note that $NV = PEX$ and that $PEX \subset PEX_\infty$ by definition, whence $NV \subset PEX_\infty$. For completeness, however, we give the following direct proof:

**Theorem 29.** $NV \subset PEX_\infty$. 

Proof. As above, we suppose that $RC \in BC_\infty$ via $G$, and show that we can construct a $\hat{G}$ to $BC$-infer $QSET$. Let $r$, $\{r_i\}$ be as above, and for $f$, define $\hat{f}$ as previously. Now let $f \in QREC$. We construct $\hat{G}$ as follows. On input $\langle f(q_0), \ldots, f(q_{n-1}) \rangle$, output $\phi(G(\langle \hat{f}(q_0), \ldots, \hat{f}(q_{n-1}) \rangle))$, for the largest $m$ such that we can compute $\hat{f}|_m$ from $f|_n$. Thus, if $A \in QSET$, then $A \in RC$, so by hypothesis, $G$ outputs indices of functions that are everywhere within $1/3$, say, of $\hat{A}$ for $n$ greater than some $N$. Let $M$ denote the least natural number greater than $N$ such that $\hat{A}|_N$ can be computed from $A|_M$. But then, $\hat{G}$ outputs indices of $A$ for $n > M$. 

The notions $BC_\infty$, $EX_\infty$, and $NV_\infty$ are the $IQ$-domain generalizations of $BC$, $EX$, and $NV$, as we show in the next few theorems. Note that $NV = PEX$ and that $PEX \subset PEX_\infty$ by definition, whence $NV \subset PEX_\infty$. For completeness, however, we give the following direct proof:
Proof. Suppose $\mathcal{C} \in NV$ via $G$, and $f \in \mathcal{C}$. We define a $\text{PEX}_\infty$-inference machine $H$ for $\mathcal{C}$ as follows. At stage $n$, output an index for the function $h$ such that $h(q_i) = f(q_i)$ for each $i < n$ and $h(q_m) = G(h(q_0), h(q_1), \ldots, h(q_{m-1}))$ for each $m \geq n$. Since $G$ makes only finitely many incorrect guesses, the sequence of indices output is eventually constant (and is in fact an index for $f$). □

Theorem 30. $BC \subsetneq BC_{\infty}$.

Proof. Containment is clear. To show that it is strict, we show that $RUC \notin BC$. We suppose that $RUC \in BC$, and then show that $\otimes \text{REC} \in BC$, a contradiction. Let $RUC \in BC$ via $G$; we will define a recursive machine $H$ which $BC$-infers every recursive function on rationals. Let $f$ be a fixed recursive function. First recall that for any rational function $f$, we may define an associated continuous rational function $\hat{f} \in RUC$ by defining $\hat{f}(0) = 0$ and $\hat{f}(1/i) = f(q_{i-1})/i$ for $i > 0$ and defining $\hat{f}$ for any other rational point by linear interpolation. Then the values of $\hat{f}(q_i)$ may be computed for $q_i \geq 1/n$ from $f(q_0), f(q_1), \ldots, f(q_{n-1})$. Let $s(n)$ be the least $s$ such that $q_{i+1} < 1/n$, so that $\lim_{n \to \infty} s(n) = \infty$. Now let $E((f(q_0), f(q_1), \ldots, f(q_{n-1}))) = G((\hat{f}(q_0), \hat{f}(q_1), \ldots, \hat{f}(q_{s(n)})))$. Since $\hat{f} \in RUC$, we have by assumption that $e_n = E((f(q_0), f(q_1), \ldots, f(q_{n-1})))$ is an index of $\hat{f}$ for all sufficiently large $n$. It is now straightforward to compute from $e_n$ an index $H((f(q_0), f(q_1), \ldots, f(q_{n-1})))$ for the function which recovers $f$ from $\hat{f}$. That is, if $e_n$ is an index for the function $\hat{f}_n$, then $H((f(q_0), f(q_1), \ldots, f(q_{n-1})))$ is an index for the function $f_n$ such that $f_n(q_{i-1}) = \hat{f}_n(1/i)$. Thus, $H$ infers the arbitrary recursive function $f$. □

In the previous literature (see [18, 1, 6]), it has been shown that $NV \equiv EX \subsetneq BC$. The next few theorems show that the inductive inference hierarchy is no longer linear when we consider the approximate inference classes, since $NV_{\infty}$ and $BC_{\infty}$ are incomparable.

Theorem 31. $EX \not\subset NV_{\infty}$.

Proof. Blum and Blum [4] observe that the class $\{f : \phi_{1/f(0)} = f\}$ is $EX$-inferable. It is easy to see that $\mathcal{C} = \{f : 1/f(0) \in \mathbb{N} \text{ and } \Phi_{f/\mathbb{N}} = \mathbb{N}\}$ is a member of (the analogue of) $EX$, in the context of recursive rational functions. We show that $\mathcal{C} \notin NV_{\infty}$. For a contradiction, suppose that $\mathcal{C} \in NV_{\infty}$ via $G$. By the $S_n$ Theorem, we may define $\Phi_{\mathcal{C}}$ as follows. Let $\Phi_{\mathcal{C}}(0) = 1/e$. If $G((\Phi_{\mathcal{C}}(0))) < 1/2$, let $\Phi_{\mathcal{C}}(1) = 1$, otherwise let $\Phi_{\mathcal{C}}(1) = 0$. Continue this process by recursion, so that, for each $n$, $\Phi_{\mathcal{C}}(n + 1) = 1$, if $G((\Phi_{\mathcal{C}}(q_0), \Phi_{\mathcal{C}}(q_1), \ldots, \Phi_{\mathcal{C}}(q_n))) < 1/2$, and $\Phi_{\mathcal{C}}(n + 1) = 0$, otherwise. Observe that $\Phi_{\mathcal{C}}$ is a total recursive function for all $e$. By the Fixed-Point Theorem, there is an index $e$ s.t. $\Phi_{\mathcal{C}}(e) = \Phi_e$. Then $\Phi_e \in \mathcal{C}$, but

$$|\Phi_e(q_n) - G((\Phi_e(q_0), \Phi_e(q_1), \ldots, \Phi_e(q_{n-1})))| \geq 1/2$$

for all $n$, a contradiction. □
Corollary 32. \( \text{EX} \nsubseteq PBC_{\infty}, \ QREC \nsubseteq \text{NV}_{\infty}, \ PBC_{\infty} \subseteq BC_{\infty}, \ \text{EX}_{\infty} \nsubseteq \text{NV}_{\infty}, \) and \( BC_{\infty} \nsubseteq \text{NV}_{\infty}. \)

Proof. \( \text{EX} \nsubseteq PBC_{\infty}, \) since \( PBC_{\infty} \subseteq \text{NV}_{\infty}, \) \( QREC \nsubseteq \text{NV}_{\infty}, \) since otherwise \( \text{EX} \subseteq \text{NV}_{\infty}. \) \( PBC_{\infty} \subseteq BC_{\infty}, \) since if \( PBC_{\infty} = BC_{\infty}, \) then \( BC_{\infty} \subseteq \text{NV}_{\infty} \) by Theorem 20, so that \( \text{EX} \subseteq \text{NV}_{\infty}, \) contradicting Theorem 31. Finally, since \( \text{EX} \subseteq \text{EX}_{\infty} \) and \( \text{EX} \subseteq BC \subseteq BC_{\infty}, \) the last two assertions are true. \( \square \)

Proposition 33. \( BC \nsubseteq \text{EX}_{\infty}. \)

Proof. The proof of \( \text{EX}^1 - \text{EX} \neq \emptyset \) in [4] also shows that \( \text{EX}^1 - \text{EX}_{\infty} \neq \emptyset. \) \( \square \)

The next theorem, along with the previous corollary, shows that \( \text{NV}_{\infty} \) and \( BC_{\infty} \) are incomparable.

Theorem 34. \( \text{NV}_{\infty} \nsubseteq BC_{\infty}. \)

Proof. Let \( \mathcal{C} = \{ f \in QREC: \lim_{n \to \infty} f(q_{2n}) = 1/2, \ f(q_{2n}) < 1/2 \Rightarrow f(q_{2n+1}) = 0, \) and \( f(q_{2n}) \geq 1/2 \Rightarrow f(q_{2n+1}) = 1 \}. \mathcal{C} \in \text{NV}_{\infty} \) by the following algorithm: at stages \( 2n, \) simply output \( 1/2, \) and at stages \( 2n + 1, \) use the \( f(q_{2n}) \) to predict whether the next value will be \( 0 \) or \( 1. \) We show that if \( \mathcal{C} \in BC_{\infty} \) (via \( G, \) say), then \( QSET \in BC \), a contradiction. We define a machine \( H \) which \( BC \)-infers \( QSET \) as follows. For a recursive set of rationals \( A, \) define \( f_A(q_{2k}) = 1/2 + 1/k \) if \( q_k \in A, \) \( f_A(q_{2k}) = 1/2 - 1/k \) if \( q_k \notin A, \) and \( f_A(q_{2k+1}) = \chi_A(k). \) It is clear that \( f_A \in \mathcal{C} \) for any \( A, \) so that by hypothesis \( f_A \) will be \( BC_{\infty} \)-inferred by \( G. \)

Now, if \( A \in QREC, \) on input \( \langle \chi_A(q_0), \chi_A(q_1), \ldots, \chi_A(q_{n-1}) \rangle, \) we may compute \( \langle f_A(q_0), f_A(q_1), \ldots, f_A(q_{2n-1}) \rangle \) as defined above and then compute
\[
e_n = G((f_A(q_0), f_A(q_1), \ldots, f_A(q_{2n-1}))).
\]

Since \( f_A \in \mathcal{C}, \) there is an \( N \) such that for all \( n > N, \) we have \( \| f_A - \Phi_{e_n} \| < 1/2. \) Thus for \( n > N \) and any \( q, f_A(q) = 1 \) if and only if \( \Phi_{e_n}(q) > 1/2. \) Now use the \( S_n^m \) Theorem to compute from the program \( e_n \) a program \( H((\chi_A(q_0), \chi_A(q_1), \ldots, \chi_A(q_{n-1}))) = a_n \) so that
\[
\Phi_{e_n}(q_k) = \begin{cases} 1 & \text{if } \Phi_{e_n}(q_{2k+1}) > 1/2, \\ 0 & \text{otherwise.} \end{cases}
\]

Since \( \chi_A(q_k) = f_A(q_{2k+1}), \) it follows that \( H \) \( BC \)-identifies the arbitrary recursive set \( A. \) \( \square \)

Corollary 35. \( PBC_{\infty} \subseteq \text{NV}_{\infty}, QREC \nsubseteq BC_{\infty}. \)

We have shown that the relationships among the approximate inference notions are not always analogous to the relationships among the notions of standard inference. In the sequel, we explore the reasons for this.
3.3. The extended inference hierarchy

Fig. 1 illustrates the inclusions derived in the previous section among the various inference notions. When we add the notions $NV_\infty$ and $BC_\infty$ to the picture, the inference hierarchy is no longer linear, and as mentioned earlier, analogues of some theorems of "standard" inference no longer hold. We offer a heuristic argument why this is to be expected.

In the "standard" inference setting, it is easily shown that $NV \subseteq BC$, but that $NV = PBC$. The proof that $PBC \subseteq NV$ carries over to approximate inference. The reverse inclusion does not. Consider the usual proof that $NV \subseteq PBC$. Given $\mathcal{C} \in NV$ (via $M$, say), and $f \in \mathcal{C}$, at stage $n$ we output a program $M_n$ which on input $k < n$ computes $M(f(q_0), f(q_1), \ldots, f(q_k))$, on input $k = n$ computes $M(\langle f(0), f(1), \ldots, f(n - 1) \rangle, M(\langle f(0), \ldots, f(n - 1) \rangle))$, and so on by recursion. Since $M$ is completely accurate in its guesses from some stage $N$ onward, the program $M_n$ computes $f$ if $n > N$, and so $\mathcal{C} \in PBC$.

Suppose we try to translate this proof to the new inference setting, as follows: given $\mathcal{C} \in NV_\infty$ via $G$, and $f \in \mathcal{C}$, use $G_n$ (defined analogously to $M_n$) as a guess for $f$ at stage $n$. Unfortunately, this does not work. Roughly speaking, if at stage $n$, $G$'s guess for $f(q_n)$ differs from $f(q_n)$ by $\varepsilon > 0$, however small, then we expect...
that this error will not only propagate, but worsen as we ask \( G_n \) to compute on inputs \( q_k \) for large \( k \), so that for \( k \gg n \), \( |G_n(q_k) - f(q_k)| \gg 0 \). In particular, we set \( G_n(q_{n+1}) = G(f(q_0), \ldots, f(q_{n-1}), G_n(q_n)) \), but since \( G_n(q_n) \neq f(q_n) \), we do not know that \( |G_n(q_{n+1}) - f(q_{n+1})| < \varepsilon \). In fact, even assuming a strong continuity property for \( G \), such as

\[
\|x - y\| < \varepsilon \Rightarrow \|G(x) - G(y)\| < \varepsilon
\]

we may only conclude that \( |G_n(q_{n+1}) - f(q_{n+1})| < 2\varepsilon \).

Of course, in the proof that \( NV \subset PBC \), we might expect \( M_n \)'s error to propagate in the same way. The difference is that \( M \) is accurate from some finite stage onward, so that from this stage on, there is no error to propagate. In contrast, we cannot be sure that \( G \)'s guesses are ever completely accurate, only that they get "better" as time goes on. Thus, every \( G_n \) magnifies the error \( |G(f(q_0), f(q_1), \ldots, f(q_{n-1})) - f(n)| \) when input \( q_k, k \gg n \), so that we cannot expect that \( f \) is the uniform limit of the sequence \( \{G_n\} \).

Recall that all of the standard inference notions have the following "finitistic" component which is lacking in the notions of approximate inference given thus far. Using \( NV \) as an example, if \( f \) is inferred via \( M \), then by some (finite) stage \( N \), \( M \) will ever after predict correctly the next value to be input from \( f \) (although \( M \) does not "know" when this stage \( N \) occurs). Thus, unlike the approximate inference methods defined up to now, there is a criterion for accuracy which is met at some finite stage. The next section gives methods for approximate inference which also have this feature.

4. Weaker notions of approximate inference

4.1. The "epsilon" inference criteria

Recall the motivation for approximate inference from the previous section. \( f : [0, 1] \rightarrow [0, 1] \) is continuous, and we are allowed to ask for a sequence of values \( f(x) \), say \( f(0), f(1), f(1/2), f(1/3), f(2/3) \), and so on. As before, after we ask for the \( n \)th value, we may construct an approximation \( f_n \) to the graph of \( f \), by some form of continuous interpolation. At some point in this procedure, we may be satisfied that we are "close enough" for our particular purposes, and so, no longer wish to continue to build the approximations \( \{f_n\} \), settling instead on, say, \( g = f_N \) for some fixed \( N \) in all later computations. If we have in mind that we wish to be within, say, \( \varepsilon = 0.00001 \) of \( f(x) \) for all \( x \), we can keep asking for new values of \( f(x) \), checking these to make sure that \( |g(x) - f(x)| < \varepsilon \). If we eventually happen upon an \( x \) for which this does not hold, we may then update our interpolant \( g \). However, since the \( f_n \)'s converge uniformly to \( f \), we will only have to make finitely many such updates. Thus, analogous to the standard inference notions, this procedure allows us to meet our inference criterion by some finite stage, although we cannot in general determine when this stage occurs.
We now modify the notions of approximate inference accordingly to formalize this idea. The $BC_\infty$ inference criterion is essentially one of uniform convergence. Thus, a natural weakening of this criterion is to require convergence of the guesses only to within $\varepsilon$ for a fixed $\varepsilon > 0$. We may similarly weaken other approximate inference criteria. In the sequel, we take $\varepsilon$ to be rational, for purposes of computability.

**Definition 36.** Let $\varepsilon > 0$. A class $\mathcal{C}$ of recursive rational functions is next-value approximable to within $\varepsilon$ ($\mathcal{C} \in NV_\varepsilon$) if there is a $G : [I_Q]^{<\omega} \rightarrow I_Q$ such that for every $f \in \mathcal{C}$,

$$\limsup_n |f(q_n) - G(\langle f(q_0), \ldots, f(q_{n-1}) \rangle)| < \varepsilon.$$

**Definition 37.** Let $\varepsilon > 0$. A class $\mathcal{C}$ of recursive rational functions is explanatorily approximable to within $\varepsilon$ ($\mathcal{C} \in EX_\varepsilon$) if there is a $G : [I_Q]^{<\omega} \rightarrow \mathbb{N}$ such that for every $f \in \mathcal{C}$, there is an index $e$ such that

- $\Phi_e$ is total,
- for all but finitely many $n$, $G(\langle f(q_0), \ldots, f(q_{n-1}) \rangle) = e$, and

$$\limsup_n |f(q_n) - \Phi_e(q_n)| < \varepsilon.$$

The definition for $\mathcal{C} \in PEX_\varepsilon$ is similar, but $G$ must output only indices for total functions.

**Definition 38.** A class $\mathcal{C}$ of recursive rational functions is behaviorally correctly approximable to within $\varepsilon$ ($\mathcal{C} \in BC_\varepsilon$) if there is a recursive $G : [I_Q]^{<\omega} \rightarrow \mathbb{N}$ such that for every $f \in \mathcal{C}$,

$$\limsup_n \|f - \Phi_G(\langle f(q_0), \ldots, f(q_{n-1}) \rangle)\|_{\infty} < \varepsilon.$$

So, $G$ $BC_\varepsilon$-infers $\mathcal{C}$ if for each $f \in \mathcal{C}$, $\exists N \forall n > N \ G(\langle f(q_0), f(q_1), \ldots, f(q_n) \rangle)$ is an index of a recursive function whose values are strictly within $\varepsilon$ of $f$’s values. The definition for $\mathcal{C} \in PBC_\varepsilon$ is similar, but $G$ must output only indices for total functions.

Each of the above criteria yields a hierarchy of inference classes parameterized by $\varepsilon > 0$. Clearly, for $\varepsilon > 1/2$, the class $\mathcal{I}_\varepsilon$ contains $QREC$ for every inference notion $\mathcal{I} \in \{NV, EX, PEX, BC, PBC\}$, so the hierarchies collapse above $\varepsilon = 1/2$. We will show that the hierarchies do not collapse below $\varepsilon = 1/2$, and are in fact strictly monotone. In the sequel, $\delta, \varepsilon$ range over rational numbers in $[0, 1]$. Let $\delta f$ be the function obtained from $f$ by pointwise multiplication by $\delta$, and for a class of functions $\mathcal{C}$, let $\varepsilon \mathcal{C}$ denote the class $\{\delta f \mid f \in \mathcal{C}, 0 \leq \delta \leq \varepsilon\}$. It is easy to see that for any $\varepsilon > 0$, if $0 < \delta < 2\varepsilon$, then $\delta QREC$ is an element of each $\mathcal{I}_\varepsilon$ for $\mathcal{I} \in \{NV, EX, PEX, BC, PBC\}$.

We begin by showing that these new inference notions are strictly weaker than the notions of approximate inference, in the sense that the new notions infer more classes of functions.
Theorem 39. For any $\varepsilon > 0$, $NV_\infty \subseteq NV_\varepsilon$.

Proof. Note that $\varepsilon QREC \subseteq NV_\varepsilon$. We show that $\varepsilon QREC \notin NV_\infty$: for any $G$ (which potentially $NV_\infty$-infers $\varepsilon QREC$), we find an $f$ in $\varepsilon QREC$ not inferred by $G$. Let $\Psi$ be defined by

$$\Psi(q_n) = \begin{cases} 
\varepsilon & \text{if } G(\langle \Psi(q_0), \Psi(q_1), \ldots, \Psi(q_{n-1}) \rangle) < \varepsilon/2, \\
0 & \text{otherwise.}
\end{cases}$$

Since $\Psi$ is recursive, and $\|\Psi\|_\infty \leq \varepsilon$, $\Psi$ is an element of $\varepsilon QREC$. But

$$\limsup_n |\Psi(q_n) - G(\langle \Psi(q_0), \Psi(q_1), \ldots, \Psi(q_{n-1}) \rangle)| \leq \varepsilon/2,$$

so $\varepsilon QREC$ is not $NV_\infty$-inferred by $G$. \qed

Theorem 40. For $\varepsilon > 0$, $EX_\infty \subseteq EX_\varepsilon$.

Proof. Note that $\varepsilon QREC \subseteq EX_\varepsilon$. We show that $\varepsilon QREC \notin EX_\infty$: Suppose that $\varepsilon QREC \in EX_\infty$ Then $\varepsilon QSET \in EX_\infty$ (via $G$, say), since $QSET \subseteq QREC$. But then if $f \in QSET$, feed $\varepsilon f$ to $G$. Let $e_n$ be the index output by $G$ at stage $n$, i.e. $e_n = G(\langle \varepsilon f(q_0), \ldots, \varepsilon f(q_{n-1}) \rangle)$. Then, since $G EX_\infty$-infers $\varepsilon f$, there is an index $e$ and a stage $N$ so that for each $n > N$, $e_n = e$, and

$$\lim_k |\Phi_e(q_k) - \varepsilon f(q_k)| < \varepsilon/2.$$

We use this to $EX_\infty$-infer $QSET$ as follows. For each $n$ we output an index of the function $\Psi_n$ defined by

$$\Psi_n(x) = \begin{cases} 
1 & \text{if } \Phi_{e_n}(x) \downarrow > \varepsilon/2, \\
0 & \text{if } \Phi_{e_n}(x) \downarrow \leq \varepsilon/2, \\
\text{undefined} & \text{otherwise.}
\end{cases}$$

Since $e_n = e$ after stage $N$, with $|\Phi_e(q_k) - \varepsilon f(q_k)| < \varepsilon/2$ for sufficiently large $k$, then because $f \in QSET$, we have $|\Psi_n(q_k) - f(q_k)| = 0$ for such $k$. But this procedure is uniform in $f$, so it defines an $EX_\infty$-inference machine for $QSET$, a contradiction. Hence $\varepsilon QREC \notin EX_\varepsilon - EX_\infty$. \qed

Theorem 41. For $\varepsilon > 0$, $BC_\infty \subseteq BC_\varepsilon$.

Proof. Note that $\varepsilon QREC \subseteq BC_\varepsilon$. We show that $\varepsilon QREC \notin BC_\infty$: Suppose that $\varepsilon QREC \in BC_\infty$. Then $\varepsilon QSET \in BC_\infty$ (via $G$, say), since $QSET \subseteq QREC$. But then if $f \in QSET$, feed $\varepsilon f$ to $G$. Let $e_n$ be the index output by $G$ at stage $n$, i.e. $e_n = G(\langle \varepsilon f(q_0), \ldots, \varepsilon f(q_{n-1}) \rangle)$, then since $G BC_\infty$-infers $\varepsilon f$, we have

$$\lim_n \|\Phi_{e_n} - \varepsilon f\|_\infty = 0,$$
so there is a stage \( N \) so that for each \( n > N \),
\[
\| \Phi_{e_n} - \varepsilon f \|_\infty < \varepsilon/2.
\]
We use this to \( BC_\infty \)-infer \( Q\text{SET} \) as follows. For each \( n \) we output an index of the function \( \Psi_n \) defined by
\[
\Psi_n(x) = \begin{cases} 
1 & \text{if } \Phi_{e_n}(x) > \varepsilon/2, \\
0 & \text{if } \Phi_{e_n}(x) \leq \varepsilon/2, \\
\text{undefined} & \text{otherwise.}
\end{cases}
\]
Since the \( \| \Phi_{e_n} - \varepsilon f \|_\infty < \varepsilon/2 \) after stage \( N \), if \( f \in Q\text{SET} \), then \( \Psi_n = f \) after stage \( N \). But this procedure is uniform in \( f \), so it defines a \( BC_\infty \)-inference machine for \( Q\text{SET} \), a contradiction. Hence \( \varepsilon Q\text{REC} \in BC_\varepsilon - BC_\infty \).  

The preceding two proofs, mutadis mutandis, yield analogous results for \( P\text{EX}_\infty \) and \( P\text{BC}_\infty \):

**Corollary 42.** For \( \varepsilon > 0 \), \( P\text{EX}_\infty \not\subseteq P\text{EX}_\varepsilon \).

**Corollary 43.** For \( \varepsilon > 0 \), \( P\text{BC}_\infty \not\subseteq P\text{BC}_\varepsilon \).

Above, we exhibited, for each \( \varepsilon \), a class \( \mathcal{C}_\varepsilon \) such that \( \mathcal{C}_\varepsilon \in EX_\varepsilon - EX_\infty \). We can further show that there is a class \( \mathcal{C} \) with \( \mathcal{C} \in \bigcap_\varepsilon EX_\varepsilon - EX_\infty \), and that \( \mathcal{C} \in EX_\varepsilon \) uniformly in \( \varepsilon \). This class \( \mathcal{C} \) is just \( R\text{UC} \).

**Theorem 44.** \( R\text{UC} \in EX_\varepsilon \) for all \( \varepsilon \).

**Proof.** The machine \( G_\varepsilon \) which accomplishes this is a simple variant of the linear interpolation procedure used previously. Fix \( f \in Q\text{REC} \), let \( L \) be, initially, the zero function, and let \( L_n \) denote the linear interpolation of \( (f(q_0), \ldots, f(q_{n-1})) \). We construct \( G_\varepsilon \) as follows: at stage \( n \), if \( L(q_{n-1}) \) is not within \( \varepsilon \) of \( f(q_{n-1}) \), set \( L = L_n \). Output \( L \).

Now, if \( f \in R\text{UC} \), eventually, we will reach a stage after which all of the linear interpolants \( L_n \) are everywhere within \( \varepsilon \) of \( f \). \( G_\varepsilon \) will stabilize on the first such interpolant.  

In the case of \( NV_\infty \) it is straightforward to show that there is a \( \mathcal{C} \) such that \( \mathcal{C} \in \bigcap_\varepsilon NV_\varepsilon - NV_\infty \), although not necessarily uniformly. For each \( n > 0 \) and \( x \), let \( s_n(x) = (1/2^{n+1})x + 1/2^n \). Then for fixed \( n \), for any \( f \in Q\text{REC} \), and all \( x \in I_Q \), we may define \( S_{f,n}(x) = s_n(f(x)) \). Now, enumerate \( Q\text{SET} \) arbitrarily as \( \{f_1, f_2, \ldots\} \). In the next two theorems, let \( \mathcal{C} = \{S_{f,i}\} \), and let \( v(n) \) be an effective enumeration of \( \mathbb{N} \), in which each natural number appears infinitely often.

**Theorem 45.** \( NV_\infty \not\subseteq \bigcap_\varepsilon NV_\varepsilon \).
Proof. We show that $\mathcal{C} \in NV_{1/k}$ for any $k > 1$, but that $\mathcal{C} \notin NV_\infty$. Fix $k > 1$, and let $f_i = S_{f,i}$. Then $\mathcal{C} \in NV_{1/k}$ via $G_k$, defined as follows (input $f \in \mathcal{QREC}$):

Stage $n$:

Step 1: Find $l > 0$ such that $f(q_{n-1}) \in [1/2^l, 1/2^{l-1})$ (if no such $l$ exists -- i.e. $f(q_n)$ is 0 or 1 -- output 1, and goto stage $n + 1$).

Step 2: If $l < k$, output $f_i(q_n)$, otherwise, output $1/2^l$.

End of Construction.

Clearly, if $f \in \mathcal{C}$, then $G_k$ $NV_{1/k}$-infers $f$. Note that the construction of $G_k$ is not uniform in $k$, since no enumeration of $\mathcal{QREC}$ is effective. Intuitively, this is why $\mathcal{C}$ is not $NV_\infty$-inferable. We suppose $\mathcal{C}$ is $NV_\infty$-inferable, via $G$, say and show that we can then $NV$-infer $\mathcal{QSET}$, for a contradiction. Construct $M$ as follows (input $f \in \mathcal{QREC}$, and let $e = 0$):

Stage $n$:

Step 1: Compute $v_n = G(S_{f,v(e)}|n)$.

Step 2: If $v_n$ is closer to $1/2^v(e)$ than to $1/2^{v(e)+1}$ output 1, else output 0.

Step 3: If

$$| G(S_{f,v(e)}|n-1) - S_{f,v(e)}(q_{n-1}) | > \frac{1}{2^{v(e)+1}},$$

increment $e$.

End of Construction.

Now if $A \in \mathcal{QSET}$, then by the construction, since $G$ infers $\mathcal{C}$, $M$ infers $A$. Given $f = f_i \in \mathcal{QSET}$, let $\hat{f} = f_i = S_{f,i}$. Then $\hat{f} \in \mathcal{C}$, and $\hat{f}(n) = \frac{1}{2^i}$ or $\frac{1}{2^i} + \frac{1}{2^{i+1}}$ for each $n$. Let $N$ be large enough so that

$$| G((\hat{f}(q_0), \ldots, \hat{f}(q_{n-1}))) - \hat{f}(q_n) | < \frac{1}{2^{i+2}}$$

for all $n > N$. There are two cases to consider.

Case 1: There is a stage $m \geq N$ where $v(e) = i$. Then for every $n \geq m$

$$G(S_{f,v(e)}|n-1) = G((\hat{f}(q_0), \ldots, \hat{f}(q_{n-2}))),$$

and $S_{f,v(e)}(q_{n-1}) = \hat{f}(q_{n-1})$, so that the difference computed in Step 3 is less than $1/2^{i+1}$, which means that $v(e)$ remains equal to $i$ thereafter. Thus, for all $n \geq m$,

$$| v_n - \hat{f}(q_n) | < 1/2^{i+2},$$

so that by Step 2 of the construction, since $f \in \mathcal{QSET}$, $M((f(q_0), \ldots, f(q_{n-1}))) = f(q_n)$ for all $n \geq m$.

Case 2: There is no stage $m \geq N$ where $v(e) = i$. Since $v(e) = i$ for infinitely many $e$, this means that $e$ remains fixed from some stage $m$ onwards. But this implies, by Step 3, that $G$ is successfully inferring $S_{f,v(e)}(q_n)$ for all $n \geq m$. It follows, as in the
previous case, that $M$ infers $f$. Note that in this case, $G$ is inferring $S_{f,v}(e)$ even though $S_{f,v}(e)$ may not be in $\mathcal{C}$. \(\square\)

We remark that the above proof technique is independent of the enumeration of $QSEC$ chosen, and therefore does not work to show that $BC_\infty \subseteq \cap \varepsilon BC_\varepsilon$. To see this, enumerate $QSET$ as follows: let $f_i = \phi_i$ if $i$ is the index of a total recursive function which is not identically zero, otherwise let $f_i = \lambda x.0$. Then the class $\mathcal{C} = \{S_{f_i,1}\}$ is EX-inferable (and therefore $BC_\infty$-inferable) by the machine which at stage $n$ outputs

- an index of $\lambda x.0$ if input the constant sequence $(1/2^{e+1}, \ldots, 1/2^{e+1})$ of length $n$,
- the index $e$ if input any other sequence $(a_0, \ldots, a_{n-1})$ such that $a_i \in [1/2^e, (1/2^e) + 1/2^{e-1})$ for all $i$,
- 0, otherwise.

We do not know of a proof for $BC_\infty \subseteq \cap BC_\varepsilon$, but we conjecture that the statement is true.

4.2. Monotonicity of the “epsilon” Hierarchies

By adapting the proofs of the previous section, it is easy to show that the “epsilon” hierarchies do not collapse for $\varepsilon \leq 1/2$:

**Theorem 46.** For any $0 < \varepsilon < 1/2$, $2\varepsilon QREC \notin NV_\varepsilon$.

**Proof.** We show that for any $G$ (which potentially $NV_\varepsilon$-infers $2\varepsilon QREC$), we can find an $f$ in $2\varepsilon QREC$ not inferred by $G$. Let $\Psi$ be defined by

$$\Psi(q_n) = \begin{cases} 2\varepsilon & \text{if } G(\langle \Psi(q_0), \Psi(q_1), \ldots, \Psi(q_{n-1}) \rangle) < \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\Psi$ is recursive, and $||\Psi||_\infty \leq 2\varepsilon$, $\Psi$ is an element of $2\varepsilon QREC$. But

$$\limsup_n \Psi(q_n) - G(\langle \Psi(q_0), \Psi(q_1), \ldots, \Psi(q_{n-1}) \rangle) \geq \varepsilon,$$

so $2\varepsilon QREC$ is not $NV_\varepsilon$-inferred by $G$. \(\square\)

Thus, the $\{NV_\varepsilon\}_{0 < \varepsilon \leq 1/2}$ hierarchy does not collapse:

**Corollary 47.** If $0 < \varepsilon_1 < \varepsilon_2 \leq 1/2$, then $NV_{\varepsilon_1} \subsetneq NV_{\varepsilon_2}$.

**Proof.** Any $NV_{\varepsilon_1}$-inference machine is by definition an $NV_{\varepsilon_2}$-inference machine, so $NV_{\varepsilon_1} \subseteq NV_{\varepsilon_2}$. This containment is strict, however, by the preceding theorem, since $2\varepsilon_1 QREC \in NV_{\varepsilon_2}$, via $G$ which simply outputs $\varepsilon_1$ on any input. \(\square\)

Also, since $NV_\varepsilon$ contains both $PEX_\varepsilon$ and $PBC_\varepsilon$, we have:

**Corollary 48.** For $\varepsilon \leq 1/2$, $2\varepsilon QREC \notin PBC_\varepsilon$ and $2\varepsilon QREC \notin PEX_\varepsilon$. 
Corollary 49. If $0 < \varepsilon_1 < \varepsilon_2 \leqslant 1/2$, then $PBC_{\varepsilon_1} \subseteq PBC_{\varepsilon_2}$, and $PEX_{\varepsilon_1} \subseteq PEX_{\varepsilon_2}$.

Theorem 50. For $0 < \varepsilon \leqslant 1/2$, $2\varepsilon QREC \notin EX_\varepsilon$.

Proof. We suppose $2\varepsilon QREC \in EX_\varepsilon$, for a contradiction. Then $2\varepsilon QSET \in EX_\varepsilon$ (via $G$, say), since $QSET \subseteq QREC$. But then if $f \in QSET$, feed $2\varepsilon f$ to $G$. Let $e_n$ be the index output by $G$ at stage $n$, i.e. $e_n = G((2\varepsilon f(q_0), \ldots, 2\varepsilon f(q_{n-1})))$. Then, since $G$ $EX_\varepsilon$-infers $2\varepsilon f$, there is an index $e$ and a stage $N$ so that for each $n > N$, $e_n = e$, and

$$\lim_{k} |\Phi_e(q_k) - 2\varepsilon f(q_k)| < \varepsilon.$$

We use this to $EX$-infer $QSET$ as follows. For each $n$ we output an index of the function $\Psi_n$ defined by

$$\Psi_n(x) = \begin{cases} 
1 & \text{if } \Phi_{e_n}(x) \downarrow > \varepsilon, \\
0 & \text{if } \Phi_{e_n}(x) \downarrow \leqslant \varepsilon, \\
\text{undefined} & \text{otherwise.}
\end{cases}$$

Since $e_n = e$ after stage $N$, and $|\Phi_e(q_k) - 2\varepsilon f(q_k)| < \varepsilon$ for sufficiently large $k$, then $|\Psi_n(q_k) - f(q_k)| = 0$ for such $k$. But this procedure is uniform in $f$, so it defines an $EX$-inference machine for $QSET$, a contradiction. Hence $2\varepsilon QREC \notin EX_\varepsilon$. \(\square\)

Corollary 51. If $0 < \varepsilon_1 < \varepsilon_2 \leqslant 1/2$, then $EX_{\varepsilon_1} \subseteq EX_{\varepsilon_2}$.

Proof. Any $EX_{\varepsilon_1}$-inference machine is by definition an $EX_{\varepsilon_2}$-inference machine, so $EX_{\varepsilon_1} \subseteq EX_{\varepsilon_2}$. This containment is strict, however, by the preceding corollary, since $2\varepsilon_1 QREC \in EX_{\varepsilon_2}$, via $G$ which simply outputs an index for the constant function $f(q) = \varepsilon_1$ on any input. \(\square\)

In the following, if $|g - f| < \varepsilon$, we will call $g$ an $\varepsilon$-variant of $f$.

Theorem 52. For $0 < \varepsilon \leqslant 1/2$, $2\varepsilon QREC \notin BC_\varepsilon$.

Proof. We suppose $2\varepsilon QREC \in BC_\varepsilon$, for a contradiction. Then $2\varepsilon QSET \in BC_\varepsilon$ (via $G$, say), since $QSET \subseteq QREC$. But then if $f \in QSET$, feed $2\varepsilon f$ to $G$. Let $e_n$ be the index output by $G$ at stage $n$, i.e. $e_n = G((2\varepsilon f(q_0), \ldots, 2\varepsilon f(q_{n-1})))$. Then, since $G$ $BC_\varepsilon$-infers $2\varepsilon f$, we have

$$\lim_{n} ||\Phi_{e_n} - 2\varepsilon f||_\infty = 0,$$

so there is a stage $N$ so that for each $n > N$,

$$||\Phi_{e_n} - 2\varepsilon f||_\infty < \varepsilon.$$

Thus, for each stage $n > N$, $G$ produces an index of an $\varepsilon$-variant of $f$. We use this act to $BC$-infer $QSET$ as follows. For each $n$ we output an index of the function $\Psi_n$,
defined by

\[ \Psi_n(x) = \begin{cases} 
1 & \text{if } \Phi_{e_n}(x) \downarrow > \varepsilon, \\
0, & \text{if } \Phi_{e_n}(x) \downarrow \leq \varepsilon, \\
\text{undefined} & \text{otherwise.} 
\end{cases} \]

Since the \( \| \Phi_{e_n} - 2\varepsilon f \|_{\infty} < \varepsilon \) after stage \( N \), then \( \Psi_n = f \) after stage \( N \). But this procedure is uniform in \( f \), so it defines a \( BC \)-inference machine for \( QSE \), a contradiction. Hence \( 2\varepsilon \mathbb{Q}REC \notin BC_{\varepsilon} \).

**Corollary 53.** If \( 0 < \varepsilon_1 < \varepsilon_2 \leq 1/2 \), then \( BC_{\varepsilon_1} \subseteq BC_{\varepsilon_2} \).

**Proof.** Any \( BC_{\varepsilon_1} \)-inference machine is by definition an \( BC_{\varepsilon_2} \)-inference machine, so \( BC_{\varepsilon_1} \subseteq BC_{\varepsilon_2} \). This containment is strict, however, by the preceding corollary, since \( 2\varepsilon \mathbb{Q}REC \in BC_{\varepsilon_p} \), via \( G \) which simply outputs an index for the constant function \( f(q) = \varepsilon_1 \) on any input. \( \Box \)

### 4.3. Relationships among the hierarchies

Note that Theorems 31 and 34 are easily modified to yield the following:

**Corollary 54.** (to 31). For any \( 0 < \varepsilon \leq 1/2 \), \( EX \nsubseteq NV_{\varepsilon} \).

**Corollary 55.** (to 34). For any \( 0 < \varepsilon \leq 1/2 \), \( NV_{\infty} \nsubseteq BC_{\varepsilon} \).

Most of the assertions below are just corollaries to the above, and to the theorems of the previous section. For \( A, B \), families of subsets of \( REC \), we use the notation \( A \perp B \) to mean \( A \not\subseteq B \) and \( B \not\subseteq A \).

**Corollary 56.** For all \( 0 < \delta, \varepsilon \leq 1/2 \), \( NV_{\varepsilon} \perp BC_{\delta} \).

**Proof.** \( NV_{\varepsilon} \nsubseteq BC_{\delta} \): \( NV_{\infty} \subset NV_{\varepsilon} \), but by Corollary 55, \( NV_{\infty} \nsubseteq BC_{\delta} \). \( BC_{\delta} \nsubseteq NV_{\varepsilon} \); \( EX \subset BC \subset BC_{\delta} \), but by Corollary 54, \( EX \nsubseteq NV_{\varepsilon} \). \( \Box \)

**Corollary 57.** For all \( 0 < \delta, \varepsilon \leq 1/2 \), \( NV_{\varepsilon} \perp EX_{\delta} \).

**Proof.** \( NV_{\varepsilon} \nsubseteq EX_{\delta} \): \( NV_{\infty} \subset NV_{\varepsilon} \), but by Corollary 55, since \( EX_{\delta} \subset BC_{\delta} \), \( NV_{\infty} \nsubseteq EX_{\delta} \). \( EX_{\delta} \nsubseteq NV_{\varepsilon} \); \( EX \subset EX_{\delta} \), but by Corollary 54, \( EX \nsubseteq NV_{\varepsilon} \). \( \Box \)

**Corollary 58.** If \( 0 < \varepsilon \leq 1/2 \), \( PBC_{\varepsilon} \nsubseteq BC_{\varepsilon} \).

**Proof.** Containment is immediate. It must be strict, since otherwise \( EX \subset NV_{\varepsilon} \), a contradiction. \( \Box \)
Corollary 59. If $0 < \varepsilon \leq 1/2$, $PBC_\varepsilon \subseteq NV_\varepsilon$.

**Proof.** Containment is immediate. It must be strict, since otherwise $NV_\infty \subset BC_\varepsilon$, a contradiction. □

Unlike the case of $PEX_\infty$ and $PBC_\infty$ in the previous section, in the framework of weak approximate inference, we can easily modify the proof of $PEX = PBC$ to yield its analogue:

Lemma 60. If $0 < \varepsilon \leq 1/2$, $PEX_\varepsilon = PBC_\varepsilon$.

**Proof.** If $\xi \in PEX_\varepsilon$ via $G$, on input $f$, at stage $n$ we patch the index $G((f(q_0),..., f(q_n)))$ with the known values $(f(q_0),..., f(q_n))$. If $f \notin \xi$, then this procedure serves to $BC_\varepsilon$-infer $f$. In the other direction, if $\xi \in PBC_\varepsilon$ via $G$, on input $f$, at stage $n$ simply output the least index $e_i = G((f(q_0),..., f(q_i)))$ (for $i < n$) for which $|\Phi_i(q_x) - f(q_x)| < \varepsilon$ for each $x < n$ (if no such $i$ exists, just output $e_n$). □

Corollary 61. If $0 < \varepsilon \leq 1/2$, $PEX_\varepsilon \subseteq EX_\varepsilon$.

**Proof.** Containment is strict: otherwise, since $PEX_\varepsilon = PBC_\varepsilon$, we would have $EX \subseteq NV_\varepsilon$, a contradiction. □

Previously, it was noted that the proof of $EX^1 - EX \neq \emptyset$ in Blum and Blum [4] also shows that $EX^1 - EX_\infty \neq \emptyset$. In fact, it is easily modified to yield $EX^1 - EX_\varepsilon \neq \emptyset$ for $0 < \varepsilon \leq 1/2$. Thus, since $EX^1 \subseteq BC$, the following proposition holds:

Proposition 62. If $0 < \varepsilon \leq 1/2$, $EX_\varepsilon \subseteq BC_\varepsilon$.

Fig. 2 illustrates the inclusions derived in this section.
5. Inference from oracles

5.1. Inference from generic oracles

We now turn to approximate inference using oracles (see Section 2 for notation and definitions). We would like to obtain analogues for approximate inference of the results in Fortnow et al. [10] characterizing oracle triviality for \( \text{EX} \) and \( \text{BC} \), namely that

\[
\text{EX}[A] = \text{EX} \iff \mathcal{G}(A) \quad \text{and} \quad \text{BC}[A] = \text{BC} \iff \mathcal{G}(A),
\]

where \( \mathcal{G}(A) \) is the condition that either \( A \) is recursive, or \( A \leq_T K \) and \( A \equiv_T G \) for some generic \( G \).

As noted previously, the crux of the argument is to show that for generic \( G \), \( \text{EX}[G^*] = \text{EX}[G] \) and that \( \text{BC}[G^*] = \text{BC}[G] \). It appears however, that similar relationships do not hold in the cases of \( \text{NI}/\), \( \text{NI} \), and \( \text{BC} \). In the sequel, we introduce the concept of a modulus of inference, and show that for classes \( \mathcal{C} \) which can be inferred by machines \( M^G \) with recursive inference moduli, only finitely many queries are needed. This, in turn, leads to new notions of approximate inference.

We recall the following definition of genericity provided by Fortnow et al. [10]:

**Definition 63.** \( G \in 2^\omega \) is generic if for each \( \Sigma_1 \) \( W \subseteq \{0,1\}^* \), either

- \( \exists \sigma \prec G \) \( (\forall \tau \geq \sigma) \tau \notin W \), or
- \( \exists \sigma \prec G \) \( \sigma \in W \).  

In the sequel, we will use the following formulation (see [13] for other characterizations).

**Lemma 64.** \( G \) is generic if and only if for each \( \Pi^0_1 \) class \( \mathcal{P} \subseteq 2^\omega \), either

- \( \exists \sigma \prec G \) \( I(\sigma) \subseteq \mathcal{P} \), or
- \( G \notin \mathcal{P} \).

**Proof.** (\( \Rightarrow \)) If \( \mathcal{P} \) is a \( \Pi^0_1 \) class, then \( \mathcal{P} = [T] \) for some recursive tree \( T \) (note that \( \overline{T} \) is \( \Sigma_1 \)). If \( G \notin \mathcal{P} \) there is nothing to prove. Suppose that \( G \in \mathcal{P} \). Then \( \forall \sigma \prec G \), \( \sigma \notin \overline{T} \). Thus, by genericity of \( G \), there is a \( \sigma \prec G \) so that for each \( \tau \geq \sigma, \tau \notin \overline{T} \), whence \( I(\sigma) \subseteq \mathcal{P} \).

(\( \Leftarrow \)) Suppose \( W \) is \( \Sigma_1 \), and \( \forall \sigma \prec G, \sigma \notin W \). Now \( \overline{W} \in \Pi^0_1 \), so \( [\overline{W}] \) is a \( \Pi^0_1 \) class, and \( G \in [\overline{W}] \). Thus, there is a \( \sigma \prec G \) with \( I(\sigma) \subseteq [\overline{W}] \), so for all \( \tau \geq \sigma, \tau \notin W \). \( \square \)

It was shown [10] that for generic \( G \), \( \text{EX}[G] = \text{EX}[G^*] \), and \( \text{BC}[G] = \text{BC}[G^*] \). \( \text{EX}_\infty[G] = \text{EX}_\infty[G^*] \) is essentially a corollary of the first result. However, \( \text{BC}_\infty[G] = \text{BC}_\infty[G^*] \) does not seem to follow from any simple modification of the proof \( \text{BC}[G] = \text{BC}[G^*] \). We suspect that in fact, \( \text{BC}_\infty[G^*] \subseteq \text{BC}_\infty[G] \), and also that \( \text{NI}_\infty[G^*] \subseteq \text{NI}_\infty[G] \). However, we can show that, at least for \( \text{NI}_k \) and \( \text{BC}_k \)-inference from generic oracles, no more power is obtained from an infinite number than from
an arbitrary finite number of queries. The basic idea is to compute from all oracles in an interval \( I(o) \), where \( o \prec G \), as long as the various computations at any given stage are all "close" to each other. If not all the computations are close, then we ask for a longer initial segment \( \sigma \prec \tau \prec G \), and start computing from oracles in \( I(\tau) \). Since \( G \) is generic, we will only have to return to \( G \) finitely many times to obtain these initial segments.

**Definition 65.** For \( \tau \in 2^{<\omega} \), let \( S_\tau \) denote the partial recursive function with domain \( |\tau| \), defined by \( S_\tau(x) = (\tau)_x \). We say that \( \tau \) is \((M, \langle x_0, \ldots, x_n \rangle)\)-minimal if \( M^{S_\tau}(\langle x_0, \ldots, x_n \rangle) \downarrow \), but for all proper initial segments \( \sigma \) of \( \tau \), \( M^{S_\sigma}(\langle x_0, \ldots, x_n \rangle) \uparrow \).

**Lemma 66.** For any categorical \( M(1) \), compact \( D \subseteq 2^\omega \), and sequence \( \langle a_0, \ldots, a_n \rangle \), the set

\[ \{ M^B(\langle a_0, \ldots, a_n \rangle) \mid B \in D \} \]

is finite.

**Proof.** For any \( B \in D \), since \( M^B(\langle a_0, \ldots, a_n \rangle) \downarrow \), there is a \( \sigma_B \prec B \) so that \( M \) uses only \( \sigma_B \) in its computation. Thus, for all \( A \in I(\sigma_B) \), \( M^B(\langle a_0, \ldots, a_n \rangle) = M^A(\langle a_0, \ldots, a_n \rangle) \).

Now, \( \{ I(\sigma_B) \mid B \in D \} \) is an open cover of \( D \), so it contains a finite subcover \( I(\sigma_{B_1}), \ldots, I(\sigma_{B_k}) \). Thus,

\[ \{ M^B(\langle a_0, \ldots, a_n \rangle) \mid B \in D \} = \{ M^{B_1}(\langle a_0, \ldots, a_n \rangle), \ldots, M^{B_k}(\langle a_0, \ldots, a_n \rangle) \} \]

is finite. \( \square \)

Note that if the relation \( B \in D \) is computable, then

\[ \{ M^B(\langle a_0, \ldots, a_n \rangle) \mid B \in D \} \]

is uniformly computable in \( \langle a_0, \ldots, a_n \rangle \).

**Theorem 67.** If \( G \) is generic, then \( NV_{1/k}[G^*] = NV_{1/k}[G] \).

**Proof.** Let \( C \in NV_{1/k}[G] \) via \( M^G \). We construct \( \hat{M} \) to \( NV_{1/k}[G^*]\)-infer \( C \). Let \( f \in QRREC \), and initialize \( \sigma_0 \) to \( \emptyset \). Then \( \hat{M} \) works as follows:

Stage \( n \):

1. **Step 1:** Output \( a_n = M^{\sigma_n 0^*}(\langle f(q_0), \ldots, f(q_{n-1}) \rangle) \).

2. **Step 2:** For each \( \tau \succ \sigma_n \) such that \( \tau \) is \((M, \langle f(q_0), \ldots, f(q_{n-2}) \rangle)\)-minimal, compute

\[ v_\tau = M^\tau(\langle f(q_0), \ldots, f(q_{n-2}) \rangle) \]

Note that since \( M \) is categorical, by the above lemma, all such \( \tau \) can be found effectively. Also, the list may be empty, since \( M \) might need only a proper initial segment of \( \sigma \).
Step 3: If for some \( \tau \) we have \(|v_\tau - f(q_{n-1})| \geq 1/k\), let \( \sigma_{n+1} = G|_n \), otherwise, let \( \sigma_{n+1} = \sigma_n \).

End of Construction.

If \( f \in \mathcal{C} \), we claim that \( \hat{M} \) makes only finitely many queries to \( G \). Since \( f \) is \( NV_{1/k} \)-inferred by \( M^G \), there is an \( N_f \) so that for each \( n > N_f \), \( |M^G((f(q_0, \ldots, f(q_{n-1}))) - f(q_n))| < 1/k \). Let

\[
\mathcal{P} = \{ B \mid \forall n \geq N_f \ |M^B((f(q_0, \ldots, f(q_{n-1}))) - f(q_n))| < 1/k \}.
\]

Then \( \mathcal{P} \) is a \( \prod_1^0 \) class, and \( G \in \mathcal{P} \), so there is a \( \sigma \prec G \) with \( I(\sigma) \subseteq \mathcal{P} \). Thus, the consequent of step 3 is invoked only finitely often, for otherwise, there is stage \( n > N_f \) such that \( \sigma_n \succ \sigma \), and a \( \tau \succ \sigma_n \) such that \(|v_\tau - f(q_{n-1})| \geq 1/k\). But then \( I(\sigma) \not\subseteq \mathcal{P} \), a contradiction. Thus, \( \hat{M} \) queries \( G \) only finitely often.

It remains to show that \( \hat{M} \) infers \( f \). Let \( N > N_f \) be so large that the consequent of step 3 is not invoked after stage \( N \), and fix \( n > N \) (note that \( \sigma_n = \sigma_N \)). If \(|a_n - f(q_n)| \geq 1/k\), then there are two cases to consider. If some \( \tau \prec \sigma_N \) was used to compute \( a_n \), then \( a_n = M^G((f(q_0, \ldots, f(q_{n-1}))) \), contradicting that \( n > N_f \). On the other hand, if some \( \tau \succ \sigma_N \) was used, then at stage \( n + 1 \) we have \( v_\tau = a_n \), so that the consequent of step 3 is invoked at this stage, contradicting that \( n > N \).

**Definition 68.** Let \( M \) be any inference machine which outputs indices. The amalgamation procedure, \( AM \), is defined from \( M, \sigma, (x_0, \ldots, x_{n-1}) \) as follows:

1. **Case 1:** \((M^\sigma((x_0, \ldots, x_{n-1})) \downarrow)\). Output \( M^\sigma((x_0, \ldots, x_{n-1})) \).
2. **Case 2:** \((M^\sigma((x_0, \ldots, x_{n-1})) \uparrow)\). For all \((M, (x_0, \ldots, x_{n-1}))\)-minimal \( \tau \succ \sigma \), compute \( e_\tau = M^\tau((x_0, \ldots, x_{n-2})) \). Output \( e \) defined by

\[
\Phi_e(x) = \begin{cases} 
\Phi_{e_\tau}(x) & \text{for the first } \tau \text{ s.t. } \Phi_{e_\tau}(x) \downarrow, \\
\tau, & \text{if } \Phi_{e_\tau}(x) \uparrow \text{ for each } \tau.
\end{cases}
\]

End of Procedure.

**Theorem 69.** If \( G \) is generic, then \( BC_{1/k}[G^*] = BC_{1/k}[G] \).

**Proof.** Let \( C \in BC_{1/k}[G] \) via \( M^G \). We construct \( \hat{M} \) to \( BC_{1/k}[G^*] \)-infer \( C \). Let \( f \in \mathcal{QREC} \), and initialize \( \sigma_0 \) to 0. Then \( \hat{M} \) works as follows:

Stage \( n \):

1. **Step 1:** Output \( e_n \), some uniformly chosen index of

\[
AM(\sigma_n, (f(q_0), \ldots, f(q_{n-1}))).
\]

2. **Step 2:** For each \( \tau \succ \sigma \), and each \( m \) with \(|\sigma_n| < m < n \) such that \( \tau \) is \((M, (f(q_0), \ldots, f(q_m)))\)-minimal compute

\[
e_\tau^m = M^\tau((f(q_0), \ldots, f(q_m))).
\]

and for each \( x < n \), run \( v_\tau^m(x) = \Phi_{e_\tau}(q_x) \) for \( n \) steps.
Step 3: If for some \( \tau, m, \) and \( x \) we have \( v_\tau^m(x) \downarrow \), with \( |v_\tau^m(x) - f(q_x)| \geq 1/k \), let \( \sigma_{n+1} = G_{|n} \). Otherwise, let \( \sigma_{n+1} = \sigma_n \).

End of Construction.

As in Theorem 67, for \( f \in \mathcal{C} \), \( M \) makes only finitely many queries to \( G \). Since \( f \) is \( BC_{1/k} \)-inferred by \( M^G \), there is an \( N_f \) so that for each \( n > N_f \),

\[
\| \Phi_{M^G((f(q_0), \ldots, f(q_{n-1}))]} - f \|_\infty < 1/k.
\]

Let \( e^*_B \) denote \( M^B((f(q_0), \ldots, f(q_{n-1}))) \). Let

\[
\mathcal{P} = \{ B | \forall n > N_f \forall m \text{ s.t. } (N_f < m < n) \forall x \Phi_{e^*_B}(x) \downarrow \Rightarrow |\Phi_{e^*_B}(x) - f(x)| < 1/k \}.
\]

Then \( \mathcal{P} \) is a \( \Pi^0_1 \) class, and \( G \in \mathcal{P} \), so there is a \( \sigma \in G \) with \( I(\sigma) \subset \mathcal{P} \).

Thus, the consequent of step 3 is invoked only finitely often, for otherwise, there is a stage \( n > N_f \), \( x < n \) and \( \tau > \sigma_n \), with \( \sigma_n > \sigma \), such that \( |v_\tau(x) - f(q_x)| \geq 1/k \). But this contradicts that \( I(\sigma) \subset \mathcal{P} \). Thus, \( M \) queries \( G \) only finitely often.

Finally, we show that \( M \) infers \( f \). Let \( \bar{\sigma} = \lim_n \sigma_n \), \( N = \max\{N_f, |\bar{\sigma}|\} \), and fix \( n > N \). If \( \| \Phi_{e^*_B} - f \|_\infty \geq 1/k \), then there is an \( x \) with \( |\Phi_{e^*_B}(q_x) - f(q_x)| \geq 1/k \). Thus, by definition of \( \mathcal{AM} \), there is a \( \tau > \bar{\sigma} \) and an \( s \) such that \( \Phi_{e^*_B}(q_x) \downarrow \), but \( |\Phi_{e^*_B}(q_x) - f(q_x)| \geq 1/k \). Then at stage \( n^* = \max\{n, s + 1\} \), the consequent of step 3 is invoked, contradicting that \( n^* > N \).

Definition 70. Let \( M^{(1)} \) be a categorical oracle T.M., let \( A \in 2^\omega \), and let \( \mathcal{C} \subset \mathcal{QREC} \).

\( \varepsilon : I_\mathcal{Q}^{<\omega} \to \omega \) is an \( NV_\infty \)-modulus for the pair \( (M^A, \mathcal{C}) \) if \( \mathcal{C} \) is \( NV_\infty \)-inferred by \( M^A \), and for each \( f \in \mathcal{C} \), \( \forall N \forall k > N \forall n, \)

\[
n > \varepsilon((f(q_0), \ldots, f(q_k))) \Rightarrow |M^A((f(q_0), \ldots, f(q_{n-1}))) - f(q_n)| < 1/k.
\]

\( \varepsilon : I_\mathcal{Q}^{<\omega} \to I_\mathcal{Q} \) is a \( BC_\infty \)-module for the pair \( (M^A, \mathcal{C}) \) if \( \mathcal{C} \) is \( BC_\infty \)-inferred by \( M^A \), and for each \( f \in \mathcal{C} \), \( \forall N \forall k > N \forall n, \)

\[
n > \varepsilon((f(q_0), \ldots, f(q_k))) \Rightarrow \|M^A((f(q_0), \ldots, f(q_{n-1}))) - f\|_\infty < 1/k.
\]

Without loss of generality, we may assume that \( \varepsilon \) is increasing.

In the sequel, for a fixed \( f \), we will often abuse notation, denoting \( \varepsilon((f(q_0), \ldots, f(q_n))) \) by \( \varepsilon(n) \).

Theorem 71. Let \( G \) be generic, and let \( \mathcal{C} \in NV_\infty[G] \) via \( M^G \). If \( M \) has a recursive \( NV_\infty \)-module, then \( \mathcal{C} \in NV_\infty[G^*] \).

Proof. Suppose \( M^G \) \( NV_\infty \)-infers \( \mathcal{C} \) with recursive modulus \( \varepsilon \). We construct \( M \) to \( NV_\infty[G^*] \)-infer \( \mathcal{C} \). Let \( f \in \mathcal{QREC} \), initialize \( \sigma_0 \) to \( 0 \), and denote by \( k_n \) the largest
k < n such that n > ε(k). Note that the sequence \( \{k_n\} \) is recursive, uniformly in \( f \).

Then \( \hat{M} \) works as follows:

Stage n:

**Step 1**: Output \( a_n = M_{0}(f(q_0), \ldots, f(q_{n-1})) \).

**Step 2**: For each \( \tau > \sigma_n \) such that \( \tau \) is (\( M, \langle f(q_0), \ldots, f(q_{n-2}) \rangle \))-minimal, compute

\[
v_\tau = M^\tau(f(q_0), \ldots, f(q_{n-2})).
\]

**Step 3**: If for some \( \tau \) we have \( |v_\tau - f(q_{n-1})| \geq 1/k_n \), let \( \sigma_{n+1} = G|_n \), otherwise, let \( \sigma_{n+1} = \sigma_n \). End of Construction.

Suppose \( f \in \mathcal{C} \). We claim that \( \hat{M} \) makes only finitely many queries to \( G \). Since \( f \) is \( N\mathcal{V} \)-inferred by \( M^G \), there is a \( K_f \) such that for each \( k > K_f \) and each \( n, n > E(k) \),

\[
n > \varepsilon(k) \Rightarrow |M^G(f(q_0), \ldots, f(q_{n-1})) - f(q_n)| < 1/k.
\]

Let

\[
\mathcal{P}_k = \{ B \mid \forall n > \varepsilon(k) \, |M^B(f(q_0), \ldots, f(q_{n-1})) - f(q_n)| < 1/k \}.
\]

Then \( \mathcal{P}_k \) is a \( \prod^0_1 \) class. Since \( \mathcal{P} = \bigcap_{k > K_f} \mathcal{P}_k \) is an effective intersection of \( \prod^0_1 \) classes, and \( G \in \mathcal{P} \), there is a \( \sigma < G \) with \( I(\sigma) \subseteq \mathcal{P} \).

Thus, the consequent of step 3 is invoked only finitely often, for otherwise, there is a \( k > K_f \), a stage \( n > \varepsilon(k) \), with \( |\sigma_n| > k \) so that \( \sigma < \sigma_n \), and a \( \tau > \sigma_n \), such that \( |v_\tau(x) - f(q_\tau)| \geq 1/k \). But then \( I(\sigma) \not\subseteq \mathcal{P}_k \), a contradiction, since \( k > K_f \). Thus, \( \hat{M} \) queries \( G \) only finitely often.

It remains to show that \( \hat{M} \) infers \( f \). Let \( k > K_f \), and \( N > \varepsilon(k) \) be so large that the consequence of step 3 is not invoked after stage \( N \), and fix \( n > N \) (note that \( \sigma_n = \sigma_N \)). If \( |a_n - f| \geq 1/k \), then there are two cases to consider. If some \( \tau < \sigma_N \) was used to compute \( a_n \), then \( a_n = M^G(f(q_0), \ldots, f(q_{n-1})) \), contradicting that \( n > \varepsilon(k) \) and \( k > K_f \). On the other hand, if some \( \tau > \sigma_N \) was used, then \( a_n = M^\tau(f(q_0), \ldots, f(q_{n-1})) \), so at stage \( n + 1 \), we have that \( |v_\tau - f(q_{n-1})| \geq 1/k \), so that the consequent of step 3 is invoked, again a contradiction. \( \square \)

**Theorem 7.2.** Let \( G \) be generic, and let \( \mathcal{C} \in BC_\infty[G] \) via \( M^G \). If \( M^G \) infers \( \mathcal{C} \) with recursive \( BC_\infty \)-modulus, then \( \mathcal{C} \in BC_\infty[G^\ast] \).

**Proof.** Let \( M^G \) have modulus \( s \). We construct \( \hat{M} \) to \( BC_\infty[G^\ast] \)-infer \( C \). Let \( f \in QREC \), initialize \( \sigma_0 \) to \( \emptyset \), and denote by \( k_n \) the largest \( k < n \) such that \( n > \varepsilon(k) \). Then \( \hat{M} \) works as follows:

Stage n:

**Step 1**: Output \( e_n \), some uniformly chosen index of

\[
AM(\sigma_n, f(q_0), \ldots, f(q_{n-1})).
\]
Step 2: For each \( \tau \succ \sigma \), and each \( m \) with \( |\sigma_n| < m < n \) such that \( \tau \) is \((M, f(q_0), \ldots, f(q_m)))\)-minimal, compute
\[
e^{\sigma}_\tau = M^\tau((f(q_0), \ldots, f(q_m))),
\]
and for each \( x < n \), run \( v^{\sigma}_\tau(x) = \Phi^{\tau}_{e^\sigma}(q_x) \) for \( n \) steps.

Step 3: If for some \( \tau, m, \) and \( x \) we have \( v^{\sigma}_\tau(x) \downarrow \), with \( |v^{\sigma}_\tau(x) - f(q_x)| \geq 1/k_m \), let \( \sigma_{n+1} = G|_n \). Otherwise, let \( \sigma_{n+1} = \sigma_n \).

End of Construction.

Suppose \( f \in \mathcal{C} \). We claim that \( \hat{M} \) makes only finitely many queries to \( G \). Since \( f \) is \( BC_\infty \)-inferred by \( M^G \), there is an \( K_f \) such that for each \( k > K_f \) and each \( n \),
\[
n > \varepsilon(k) \Rightarrow \|\Phi_{M^G((f(q_0), \ldots, f(q_{n-1})))} - f\|_\infty < 1/k.
\]
Let \( e^\sigma_\eta \) denote \( M^\eta((f(q_0), \ldots, f(q_{n-1}))) \). Let
\[
\mathcal{P}_k = \{ B \mid \forall n > \varepsilon(k) \forall m \text{ s.t. } (n < m < n) \forall x
\phi^{e^\sigma_\eta}_B(x) \downarrow, |\phi^{e^\sigma_\eta}_B(x) - f(x)| < 1/k \}.
\]
Then \( \mathcal{P}_k \) is a \( \prod^0_1 \) class. Since \( \mathcal{P} = \bigcap_{k > K_f} \mathcal{P}_k \) is an effective intersection of \( \prod^0_1 \) classes, and \( G \in \mathcal{P} \), there is a \( \sigma < G \) with \( I(\sigma) \subset \mathcal{P} \).

Thus, the consequent of step 3 is invoked only finitely often, for otherwise, there is a \( k > K_f \), a stage \( n > \varepsilon(k) \), with \( |\sigma_n| > k \) so that the consequent of step 3 is invoked at this stage. So, there is an \( x < n \), and \( m \) with \( |\sigma_n| < m < n \), such that \( \sigma < \sigma_n \), and a \( \tau \succ \sigma_n \), such that \( |v^{\sigma}_\tau(x) - f(q_x)| \geq 1/k_m \). But then \( I(\sigma) \not\subset \mathcal{P}_k \), a contradiction, since \( k_m \geq k > K_f \). Thus, \( \hat{M} \) queries \( G \) only finitely often.

It remains to show that \( \hat{M} \) infers \( f \). Let \( k > K_f \), and \( N > \varepsilon(k) \) be so large that the consequent of step 3 is not invoked after stage \( N \), and fix \( n > N \) (note that \( \sigma_n = \sigma_N \)). If \( \|\Phi_{e^\sigma_\eta}(q_x) - f\|_\infty \geq 1/k \), then for some \( x \), \( |\Phi_{e^\sigma_\eta}(q_x) - f(q_x)| \geq 1/k \). There are two cases to consider. If some \( \tau < \sigma_N \) was used to compute \( e_\eta \), then \( e_\eta = M^G((f(q_0), \ldots, f(q_{n-1}))) \), contradicting that \( n > \varepsilon(k) \) and \( k > K_f \). On the other hand, if some \( \tau \succ \sigma_N \) was used, then by definition of \( AM \), \( \Phi_{e^\sigma_\eta}(q_x) = \Phi_{e^\tau}(q_x) \) for some \( m \). Now, if \( m \geq n \), then we have \( \|\Phi_{e^\tau_\tau}(f(q_0), \ldots, f(q_{n-1}))) \|_\infty \geq 1/k \), so that \( I(\sigma) \not\subset \mathcal{P}_k \), a contradiction, whence \( m < n \). But in this case, since \( n - 1 > \varepsilon(k) \), we have that \( k_n \geq k \). Thus, since \( e_\tau = M^\tau((f(q_0), \ldots, f(q_{n-1}))) \), at stage \( n \) we have \( |v^{\sigma}_\tau(x) - f(q_x)| \geq 1/k_n \), so that the consequent of step 3 is invoked, which is again a contradiction, since \( n > N \).

5.2. Inference with recursive moduli

Definition 73. We say \( \mathcal{C} \in MNV_\infty[A] \) if there is a categorical OTM. \( G^\epsilon \) such that \( \mathcal{C} \in NV_\infty[A] \) via \( G^\epsilon \), with recursive \( NV_\infty \)-modulus \( \epsilon \). We define \( MBC_\infty[A] \) similarly. We say \( \mathcal{C} \in MNV_\infty \) if \( \mathcal{C} \in MNV_\infty[A] \) for some recursive \( A \), and define \( MB \infty \) similarly.

We show in the sequel that the classes defined above lie strictly between the corresponding standard and approximate inference classes.
Theorem 74. \( NV \not\subseteq MNV_{\infty} \).

Proof. Note that there is a single recursive function which works as an \( NV_{\infty} \)-modulus for all \( NV \)-inference machines \( M \), and \( \mathcal{C} \in NV \), namely
\[
\epsilon(a_0, a_1, \ldots, a_{n-1}) = n.
\]
Thus, \( NV \subseteq MNV_{\infty} \). Now, for \( f \in QREC \), define \( \hat{f} \) by \( \hat{f}(x) = f(q_n)/(n+1) \). Let \( \mathcal{C} = \{ f \mid f \in QREC \} \). Then \( \mathcal{C} \in MNV_{\infty} \) via \( M \) which outputs 0 on any input. The modulus function is given by
\[
\epsilon(a_0, a_1, \ldots, a_{n-1}) = n.
\]
However, it is easy to see that \( \mathcal{C} \notin NV \). Suppose \( \mathcal{C} \in NV \) via \( M \). Then we can use \( M \) to build a machine \( \hat{M} \) as follows:
\[
\hat{M}((f(q_0), \ldots, f(q_{n-1}))) = (n+1) * M((\hat{f}(q_0), \ldots, \hat{f}(q_{n-1}))).
\]
It is easy to see that \( \hat{M} \) \( NV \)-infers \( QREC \). □

Theorem 75. \( BC \not\subseteq MBC_{\infty} \).

Proof. As with \( NV \)-inference,
\[
\epsilon(a_0, a_1, \ldots, a_{n-1}) = n.
\]
works as a \( BC_{\infty} \)-modulus for all \( BC \)-inference machines \( M \), and \( \mathcal{C} \in BC \). Thus, \( BC \subseteq MBC_{\infty} \). For \( f \in QREC \), define \( \hat{f} \) by linear interpolation of the following points:
\[
\hat{f}(1/n) = (1/n) * f(q_n), \text{ with } \hat{f}(0) = 0, \text{ and } \hat{f}(1) = 1.
\]
Let \( \mathcal{C} = \{ f \mid f \in QREC \} \). Then \( \mathcal{C} \subseteq RUC \), so \( \mathcal{C} \in BC_{\infty} \) by Theorem 21 (via a machine which at stage \( n \) outputs the linear interpolation of the \( n \) inputs). A recursive modulus function is given by
\[
\epsilon(a_0, a_1, \ldots, a_{n-1}) = k,
\]
where \( k \) is greatest such that \( 1/k \in \{q_0, q_1, \ldots, q_{n-1}\} \). However, it is easy to see that \( \mathcal{C} \notin BC \). Suppose \( \mathcal{C} \in BC \) via \( M \). We use \( M \) to build a machine to \( \hat{M} \) defined as follows: on input \( ((f(q_0), \ldots, f(q_{n-1}))) \), \( \hat{M} \) outputs an index of the function \( \phi_n(x) \) defined by
\[
\phi_n(q_m) = m * \Phi_M((f(q_0), \ldots, f(q_{n-1}))) (1/m).
\]
Then \( \hat{M} \) \( BC \)-infers \( QREC \). □

Theorem 76. \( MNV_{\infty} \not\subseteq NV_{\infty} \).

Proof. We suppose that \( RUC \) is \( NV_{\infty} \)-inferred by \( M \), with recursive \( NV_{\infty} \) modulus \( \epsilon \), for a contradiction. The proof uses a diagonalization construction: use \( M, \epsilon \) to construct
f in RUC for which the modulus e fails infinitely often. We construct f as follows (set f(0) = f(1) = 0, and let $\hat{q}_0 = 0$):

Stage n:

Step 1: If $q_n \leq \hat{q}_n$, then $f(q_n)$ is already defined, so do nothing.

Step 2: Otherwise, let $v_n = M((f(q_0), \ldots, f(q_{n-1})))$, and let $k$ be greatest such that $n > \varepsilon((f(q_0), \ldots, f(q_k)))$.

Step 3: If $v_n \geq 1/k$, then we violate $\varepsilon((f(q_0), \ldots, f(q_k)))$ by extending $f$ by zero to $q_n$ (i.e. set $f(q) = 0$ for all $q \in [\hat{q}_n, q_n]$). On the other hand, if $v_n < 1/k$, we violate $\varepsilon((f(q_0), \ldots, f(q_k)))$ by extending $f$ with a “hat” of height $2/k$. Thus, we define $f(q)$ for $q \in [\hat{q}_n, q_n]$ by linear interpolation of the three points $(\hat{q}_n, 0)$, $((q_n - q)/2, 2/k)$, and $(q_n, 0)$.

Step 4: Set $\hat{q}_n = \max\{q_0, \ldots, q_n\}$.

End of Construction.

Then $f \in RUC$, but by the construction, e fails infinitely often on f. □

Theorem 77. $MBC_\infty \subseteq BC_\infty$.

Proof. We suppose that RUC is $BC_\infty$-inferred by $M$, with $BC_\infty$ modulus $e$, for a contradiction. As above, we use $M$, $e$ to construct $f$ in RUC for which the modulus $e$ fails infinitely often. We construct f as follows (set $f(0) = f(1) = 0$, and let $\hat{q}_0 = 0$):

Stage n:

Step 1: If $q_n \leq \hat{q}_n$ then $f(q_n)$ is already defined, so do nothing.

Step 2: Otherwise, $q_n > \hat{q}_n$. Let $\hat{f}$ be defined by

$$
\hat{f}(q) = \begin{cases} 
  f(q) & \text{if } q \leq \hat{q}, \\
  0 & \text{otherwise},
\end{cases}
$$

and let $k$ be greatest such that $n > \varepsilon((f(q_0), \ldots, f(q_k)))$.

Step 3: Dovetail computations of $\{\Phi_{M(f_n)}(q_n)\}_{n \geq 1}$ until we find a convergent computation $v_n = \Phi_{M(f_n)}(q_n)$.

Step 4: Let $r = \max\{q_n, q_n\}$. If $v_n \geq 1/k$, then we violate $\varepsilon((f(q_0), \ldots, f(q_k)))$ by extending $f$ by zero to $r$ (i.e. set $f(q) = 0$ for all $q \in [\hat{q}_n, r]$). On the other hand, if $v_n < 1/k$, we violate $\varepsilon((f(q_0), \ldots, f(q_k)))$ by extending $f$ with a “hat” of height $2/k$. Thus, we define $f(q)$ for $q \in [\hat{q}_n, r]$ by linear interpolation of the three points $(\hat{q}_n, 0)$, $((r - q)/2, 2/k)$, and $(r, 0)$.

Step 5: Set $\hat{q}_n = \max\{q_0, \ldots, r\}$.

End of Construction.

Then $f \in RUC$, but by the construction, e fails infinitely often on f. □
6. Summary and conclusions

6.1. Summary and open problems

We have utilized the standard metric on $\mathbb{Q}$ to extend the basic notions of inductive inference in a natural way, allowing us to infer a larger class of functions, and in particular, to infer classes of continuous functions. We have explored the relationships among these new notions of approximate inference, as well as between these notions and the basic notions $NV$, $EX$, and $BC$. Specifically, we gave precise inclusions between the new inference notions and those in the standard inference hierarchy. We also explored weaker notions of approximate inference, leading to inference hierarchies analogous to the $EX_0$ and $BC_0$ hierarchies. Oracle inductive inference was also considered, and we gave sufficient conditions under which approximate inference from a generic oracle $G$ is equivalent to approximate inference with only finitely many queries to $G$. Whether these conditions are also necessary remains an open question.

We have only begun to explore the area of approximate inductive inference. In the remaining sections, we offer some ideas for further research in this field.

6.2. Stability

Recall that the standard inference hierarchy is linear, that is

$$NV = PEX \subseteq EX \subseteq BC,$$

but the analogous relation does not hold for the approximate inference classes. In particular, it is the class $NV_\infty$ which "ruins" the analogy. We wish to explore ways to redefine the notion of $NV_\infty$ to remedy this situation. To show that $NV \subseteq PEX$, at stage $N$, one uses the outputs from an $NV$-machine $M$ at stages $n > N$ as the inputs to the following stages to create a function to use as a guess for the input function at the stage $N$. The following definition formalizes this procedure.

**Definition 78.** For any $M$, $\sigma = \langle a_0, \ldots, a_{m-1} \rangle$, we define $S_{M,\sigma}$ on $I_\mathbb{Q}$ by recursion as follows:

$$S_{M,\sigma}(q_n) = \begin{cases} 
M(\langle a_0, \ldots, a_{n-1} \rangle), & \text{if } n < m, \\
M(\langle a_0, \ldots, a_{m-1}, S_{M,\sigma}(q_m), \ldots, S_{M,\sigma}(q_{n-1}) \rangle), & \text{otherwise}.
\end{cases}$$

We now introduce our first notion of stability for $NV$.

**Definition 79.** We say that $\mathcal{C}$ is $NV_\infty$-stable if there is an $M$ which $NV_\infty$-infers $\mathcal{C}$, and for all $f \in \mathcal{C}$ there is a stage $L$ so that for all stages $l > L$,

$$\lim_{n \to \infty} |S_{M,\sigma_l}(q_n) - f(q_n)| = 0,$$

where $\sigma_l = \langle f(q_0), \ldots, f(q_l) \rangle$. Denote by $SNV_\infty$ the class of all such $\mathcal{C}$. 
It is then easy to see the following.

**Proposition 80.** $SNV_\infty = PBC_\infty$.

This does not quite get us the desired inclusion, however, since $PEX_\infty \subseteq PBC_\infty$. We need an even stronger notion of stability to achieve this.

**Definition 81.** We say that $\mathcal{C}$ is $NV_\infty$-superstable if there is an $M$ which $SNV_\infty$-infers $\mathcal{C}$, and for all $f \in \mathcal{C}$ there is a stage $L$ so that for all stages $k, l > L$,

$$S_{M,s_i} = S_{M,s_j}.$$

Denote by $SSNV_\infty$ the class of all such $\mathcal{C}$.

We then obtain the desired result:

**Proposition 82.** $SSNV_\infty = PEX_\infty$.

### 6.3. Other notions of approximate inference

Another scheme for defining notions of approximate inference is one in the style of Egorov’s Theorem. We desire our inference method to get “close” to the input function, except on a set of size $\varepsilon$.

**Definition 83.** We say that $M$ $ENV_\varepsilon$-infers $\mathcal{C}$ if there is a computable set $E \subset [0, 1]$ with $\mu(E) < \varepsilon$ such that for each $f \in \mathcal{C}$,

$$\lim_{k \to \infty} |M((f(q_0), \ldots, f(q_{n-1}))) - f(q_n)| = 0,$$

where $\{q_n\}$ denotes the subsequence of $\{q_n\}$ given by the elements $q_n \notin E$.

**Definition 84.** We say that $M$ $EEX_\varepsilon$-infers $\mathcal{C}$ if there is a computable set $E \subset [0, 1]$ with $\mu(E) < \varepsilon$ such that for each $f \in \mathcal{C}$, there is an index $e$ (of a total function) and a stage $N$ such that for all $n > N$, $M((f(q_0), \ldots, f(q_n))) = e$, and

$$\lim_{k \to \infty} |\Phi_e(q_n) - f(q_n)| = 0,$$

where $\{q_n\}$ denotes the subsequence of $\{q_n\}$ given by the elements $q_n \notin E$.

**Definition 85.** We say that $M$ $EBC_\varepsilon$-infers $\mathcal{C}$ if there is a computable set $E \subset [0, 1]$ with $\mu(E) < \varepsilon$ such that for each $f \in \mathcal{C}$,

$$\lim_{k \to \infty} \|\Phi_M((f(q_0), \ldots, f(q_{n-1}))) - f\|_E = 0,$$

where $\| \cdot \|_E$ indicates that the infimum is taken over $q \in I_q - E$.

These definitions yield inference hierarchies distinct from the previous ones.
6.4. Inference of non-recursive functions

As noted in the introduction, the criteria for successful inference in the standard classes $NV, EX,$ and $BC$ limit us to inference of recursive functions. In contrast, the ideas of approximate inference allow us to extend the notion of inductive inference to include non-recursive functions. For example, the linear interpolation procedure works to $NV\infty$- or $BC\infty$-infer the class of all uniformly continuous functions on $I_\mathbb{Q}$ (of which $RUC$ is a proper subset). Now, if we take our domain of inference to be the set of all $f : I_\mathbb{Q} \to I_\mathbb{Q},$ many interesting questions arise. Clearly, not all of these functions are inferable by any of the methods given. For example, fix any non-computable irrational $a$ in $[0,1].$ Then if we define $\chi : I_\mathbb{Q} \to I_\mathbb{Q}$ by

$$\chi(q) = \begin{cases} 0 & \text{if } q < a, \\ 1 & \text{if } q > a, \end{cases}$$

then the singleton $\{\chi\}$ is not $NV\infty$- or $BC\infty$-inferable: since $\chi$ is $0,1$-valued, if it is $NV\infty$ (resp. $BC\infty$) inferable, then it is $NV$ (resp. $BC$) inferable.

6.5. Inference of real-valued functions

Slightly generalizing the input procedure for the approximate inference classes will allow us to further extend the domain of inference to include all real-valued functions (for an alternate formulation see [2]). The actual machinery is only slightly changed. Suppose that $f$ maps $[0,1]$ into $[0,1].$ We assume some fixed enumeration $\{q_n\}$ of the elements of $I_\mathbb{Q},$ in which each rational appears infinitely often, and make guesses based on finite sequences of pairs of rationals $(q,r),$ where $r$ represents a rational approximation of $f(q)$ for the input function $f.$ We may then use these “updates” to the approximation of $f(x)$ for each $x$ to try to $NV\infty$- or $BC\infty$-infer $f.$ In fact, the usual linear interpolation procedure, modified to use at each stage the latest approximations given, works to infer the class of all continuous functions mapping $[0,1]$ into $[0,1].$ Since the continuous functions on $[0,1]$ are determined by their values on $I_\mathbb{Q} = [0,1] \cap \mathbb{Q},$ we only need to approximate $f(q)$ for rational $q.$ But $q$ appears infinitely often in our enumeration $\{q_n\},$ say as the subsequence $\{q_{n_k}\},$ so the interpolation procedure $M$ will produce approximations $a_k = M((f(q_0), \ldots , f(q_{n_{k-1}}))),$ whence $f(q) = \lim_k a_k.$

6.6. Inductive feature extraction

We may wish, for instance, to compute $f'$ or $\int f$ from $f.$ All of the inductive inference paradigms, standard as well as approximate, can be used as “feature extraction” tools to compute in this manner. For example, the standard inference classes can be used to compute “formal” derivatives of the class of polynomials over $\mathbb{N}.$ With the techniques of approximate inference we can do a bit more. For elements of $\mathbb{Q}REC,$ we can compute approximations to (true) derivatives. For example, if $\mathbb{C} \subset RUC$ is a class of functions $f$ for which $f'$ is continuous, we can use linear interpolation, along with
the mean value theorem, to construct a machine $M$ which, upon input $f(q_0), f(q_1), \ldots$ outputs functions which approximate $f'$ in the $NV_\infty$ or $BC_\infty$ sense. Note that it is not necessarily the case that $f'$ is an element of $\mathbb{Q}REC$, or that its range is contained in $L_0$. Thus, this type of feature extraction provides a natural setting in which to extend the domain of functions under consideration to ones which are non-recursive, and real-valued.

Acknowledgements

This paper is an extended version of a portion of the author’s Ph.D. dissertation [15]. Some of these results were announced in [7].

The author wishes to thank his advisor, Doug Cenzer, for his guidance, and Bill Gasarch for his suggestions and comments.

References

[18] P. Odifreddi, Inductive inference, manuscript.