Asymptotic Decision Theory

October 19, 2016

1 Motivation

Asymptotic decision theory can be thought of as the proper successor to modal decision theory using logical inductors. We will define notions of agents and decision problems which are strong enough to define most of the problems we normally think about, including multi-agent games.

The goal of asymptotic decision theory is to define a limit-computable decision procedure D such that for all agents A, D doesn't do predictably worse than A on any problem forever. The aim of this post is to make that statement precise, as well as to give an algorithm which satisfies the property under further assumptions.

2 Notation blurb

Except where I state otherwise, I'll be using the notation from Tsvi's post.

Given a measure μ and a subset A of $\mathbb{R}$, I will use $L^1_A(\mu)$ to refer to the space of functions whose image lies in A which are absolutely integrable with respect to μ under the L^1 norm.

In this post, we will write $\ell^\infty_t(V_t)$ to denote bounded sequences over a sequence of norm spaces V_t under the dependently typed translation of the sup norm.

For topological spaces A and B we will write $A \to B$ only to refer to continuous functions from A to B.

Lastly, we will refer to the types of definable and computable elements of a type A as $\mathrm{defn}(A)$ and $\mathrm{expr}(A)$, respectively. For purposes of continuity, both will be considered under the subspace topology.

3 Definitions

3.1 Agents

We will use the word *agent* to refer to sequences of distributions over actions. More precisely, we will define the type

$$\mathcal{A} := \mathrm{defn}(\ell^\infty_t(L^1(P_t, 2)))$$

We will write agents as A, and their time-t behavior as A_t.

Any enumerable sequence of decidable sentences ϕ_t can be thought of as an agent which at time t takes action 0 in worlds where ϕ_t is true and 1 where ϕ_t is false. In fact every (computable) agent has this form, since we can take ϕ_t to be the sentence $[\![A_t]\!] = 1$ and recover the behavior of A exactly.

3.2 Embedding functions

In this post we will consider decision problems given by an embedding function with the following type:

$$\mathcal{E} := \prod_{t:\mathbb{N}^+} (\mathrm{defn}(L^1(P_t, 2)) \to \mathrm{defn}(L^1(P_t)))$$

such that for any $F : \mathcal{E}$ we have

$$\sup_{t:\mathbb{N}^+} \sup_{v:\mathrm{defn}(L^1(\mathbb{P}_t,2))} \sup_{w:2^\omega} |F_n(v, w)| < \infty$$

and

$$-\infty < \inf_{t:\mathbb{N}^+} \inf_{\mathrm{defn}(v:L^1(\mathbb{P}_t,2))} \inf_{w:2^\omega} |F_n(v, w)|$$

with the added property that F is computable on computable inputs.

Note that each F induces a map

$$(v_t \mapsto F_t(v_t)) : \ell_t^\infty(\mathrm{defn}(L^1(P_t, 2))) \to \ell_t^\infty(L^1(P_t))$$

which we will also write as F.

In order for this map to be continuous, we will restrict ourselves to the uniformly continuous embedding functions. More formally, we will for any ε want there to be a δ such that for all t,

$$\mathbb{E}_t^{\mathsf{PA}}[|A_t - B_t|] < \delta$$

implies

$$\mathbb{E}_t^{\mathsf{PA}}[|F_t(A_t) - F_t(B_t)|] < \varepsilon$$

Some example problems

To get a better feel for how to define decision problems in this environment, we'll go through a couple of examples. 5 and 10 is a fairly simple problem to define, so we'll start there.

$$510_t(A) := 5 + 5A$$

We can also define prisoner's dilemma against a sequence of opponents $O_t : L^1(P_t, 2) \circlearrowleft$

$$\mathrm{PD}_t^O(A) := 5A - 10O_t(A)$$

And we can define

$$\mathrm{NicerBot}_t^\varepsilon(A) := \mathrm{flip}(\mathbb{E}_t^{\mathsf{PA}}[A] + \varepsilon)$$

Defining agent simulates predictor is similarly easy.

$$\text{asp}_t(A) := 10\mathbb{E}_t^{\mathsf{PA}}[A] - A$$

That this is an adequate formalization of agent simulates predictor might not be immediately obvious, but it turns out to test for the same aspects of a decision theory.

We can also define coordination problems. Let $O_t : L^1(P_t, 2) \circlearrowleft$. Then

$$\text{CProb}_t^O(A) := |A - O_t(A)|$$

3.3 Decision Theories

While these agents can have arbitrary (definable) behaviors, what they *cannot* do is react to being placed into different environments.

To talk about decision theories, we will have to define a notion of agent which is allowed to take its decision problem as input.

We will define the type of *decision theories* as follows:

$$\mathcal{DT} := \prod_{t:\mathbb{N}^+} (\text{defn}(L^1(P_t)) \to L^1(P_t, 2))$$

where the first argument is code for the utility function of the theory.

In order to embed a decision theory in a universe, we will need to make use of the diagonal lemma to thread together the embedding function and the decision theory. We will define a function pack and another function embed by mutual recursion as follows:

$$\begin{aligned} \text{pack}(D : \text{defn}(\mathcal{DT}), F : \text{defn}(\mathcal{E})) &:= \ulcorner D(\underline{\text{embed}(D, F)})\urcorner \\ \text{embed}(D : \text{defn}(\mathcal{DT}), F : \text{defn}(\mathcal{E})) &:= F(\text{pack}(D, F)) \end{aligned}$$

where operations are all applied dimensionwise.

4 Fairness

Informally, a fair decision problem is one which does not punish you for details of your code that don't affect your action or change the difficulty of predicting your action.

We will call an embedding function F *fair* iff for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|A_t - B_t|] < \delta$$

implies

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|F_t(A_t) - F_t(B_t)|] < \varepsilon$$

for all $A, B : \mathcal{A}$

To see that this corresponds to the notion of fairness I gave above, observe that fairness implies that for any $A, B : \mathcal{A}$, if

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|A_t - B_t|] = 0$$

then

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|F_t(A_t) - F_t(B_t)|] = 0$$

The fact that strong fairness looks so much like a continuity property is important. In fact, fairness is a consequence of continuity.

Theorem 1. *All (continuous) embedding functions are fair.*

Proof. Take $\varepsilon > 0$. Let $A, B : \mathcal{A}$ and $F : \mathcal{E}$.

By the continuity of F, there must be some $\delta > 0$ such that, for any $C, D : \mathcal{A}$,

$$\|C - D\| < \delta \to \|F(C) - F(D)\| < \varepsilon \tag{4.1}$$

Taking this for our δ, we may assume

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|A_t - B_t|] = \eta < \delta \tag{4.2}$$

Take $\theta > 0$ such that $\eta + \theta < \delta$ By (4.2), there must exist some n_θ such that for all $t > n_\theta$,

$$\mathbb{E}_t^{\mathsf{PA}}[|A_t - B_t|] < \eta + \theta$$

from which we can define

$$A_t^\theta := \begin{cases} A_t & \text{for } t > n_\theta \\ 0 & \text{otherwise} \end{cases}$$

with B^θ defined analogously, such that

$$\|A^\theta - B^\theta\| < \eta + \theta$$

Thus by (4.1), we have

$$\|F(A^\theta) - F(B^\theta)\| < \varepsilon$$

which implies

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|F_t(A_t) - F_t(B_t)|] = \lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|F_t(A_t^\theta) - F_t(B_t^\theta)|] \leqslant \|F(A^\theta) - F(B^\theta)\| < \varepsilon$$

□

5 Optimality Conditions

We would like a general criterion by which to measure the performance of a decision theory. In this section, we will discuss some such criteria, as well as In our setting, this sort of measure turns out to be definable.

We will use the term *asymptotic dominance* for several different properties of the form "A doesn't lose to any B forever".

We will also say that an agent A asymptotically dominates some class of agents $\mathcal{B}$ on an embedding function F iff for any $B \in \mathcal{B}$,

$$\liminf_{t \to \infty} \mathbb{E}_t^{\mathsf{PA}}[F_t(A_t) - F_t(B_t)] \geqslant 0$$

Similarly, we will say that a definable decision theory D asymptotically dominates $\mathcal{B}$ on $\mathcal{F}$ iff $\text{pack}(D, F)$ asymptotically dominates $\mathcal{B}$ on F for every definable $F \in \mathcal{F}$.

6 Optimality for Convergent Problems

One interesting result is that, for any enumerable class of problems and agents satisfying a certain condition, it is possible to define a decision theory that dominates those agents on those problems.

Informally, we will do well on problems that are essentially one-shot against agents that are essentially liar sentences. To formalize this, we will instead speak of *convergent* agents and problems. We will call an agent A convergent iff $\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[A]$ exists. We will call a problem F convergent iff $\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[F_t(A_t)]$ exists for all convergent A.

6.1 Soft-Argmax and defining optimal agents

Defining an agent that does well is fairly simple. Just let

$$\text{optimal\_agent}_t(F, \mathcal{A}) := \arg\max_{a \in A} \mathbb{E}_t^{\mathsf{PA}}[F_t(A_t)]$$

The problem with this definition is that $\arg\max$ is not continuous, and thus we won't necessarily be able to define problems which include this agent in a meaningful way. Since we are interested in multi-agent as well as single-agent problems, this won't quite do. We will instead define a continuous analogue to $\arg\max$ which takes a fuzziness parameter.

Let F be a problem and $\mathcal{A}$ be a finite collection of agents. We will define

$$\text{sadt}_\varepsilon(F, \mathcal{A}) := \mathcal{A}[\text{soft\_argmax}_\varepsilon(i \mapsto \mathbb{E}^{\mathsf{PA}}[F(\mathcal{A}_i)])]$$

Where convenient, we will refer to $\lim_{\varepsilon\to 0} \text{sadt}_\varepsilon$ as sadt.

Theorem 2. $\text{sadt}(F, \mathcal{A})$ *asymptotically dominates* $\mathcal{A}$ *on* F.

Algorithm 1: Definition of soft_argmax$_\varepsilon$

Input: Q, a list of n rational numbers
Output: A distribution over $m < n$
d := $\delta(0)$
for $1 \leqslant i < n$ **:**
  let exp = $\sum_{j=0}^{i-1} d_j Q_j$
  let pq = $\text{Ind}_\varepsilon(Q_i > \text{exp})$
  d := pq × $\delta(i)$ + (1 - pq) × d
return d

Proof. Consider the list $\mathcal{A}'$ formed by quotienting $\mathcal{A}$ over the relation

$$A \sim B \leftrightarrow \lim_{t\to\infty} \left|\mathbb{E}_t^{\mathsf{PA}}[F_t(A_t) - F_t(B_t)]\right| = 0$$

Then for $[A]$ and $[B]$ in $\mathcal{A}'$, there will be some $\delta_{A,B} > 0$ such that

$$\delta < \lim_{t\to\infty} \left|\mathbb{E}_t^{\mathsf{PA}}[F_t(A_t) - F_t(B_t)]\right|$$

and, since there are finitely many $[A] \in \mathcal{A}'$ there must be a least such δ for all A, B, which we will call η. Let $[A]$ be the class with maximal limiting expecting utility. It must then be the case that

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[F_t(A_t) - F_t(B_t)] > \eta$$

for every $A \in [A]$, $B \notin [A]$. Let A be the first element of $[A]$ in $\mathcal{A}$. Since every class will be separated by at least η in the limit, $\text{sadt}_\eta(F, \mathcal{A})$ will eventually be a distribution over just $[A]$. And, since $A \sim A'$ for every $A, A' \in [A]$, by the definition of soft_argmax it must be the case that

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|\text{sadt}_\eta(F, \mathcal{A})_t - A_t|] = 0$$

which by fairness implies

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|F_t(\text{sadt}_\eta(F, \mathcal{A})_t) - F_t(A_t)|]$$

And, since A is a member of the maximal class $[A]$, $\text{sadt}_\eta(F, \mathcal{A})$ can do no worse in the limit than any other agent in $[A]$.

Since η was arbitrary beneath the least $\delta_{A,B}$ this must also hold for every $\varepsilon < \eta$, and thus by fairness sadt must have the same asymptotic properties. □

7 Learning the Embedding Function

We will again assume convergence of agents and embedders in this section. While sadt achieves dominance, it only does so with the input of an embedding function determining its counterfactuals. We would like to design a more naturalized

Algorithm 2: Definition of ditto decision theory (written ddt)

Input: A rational number ε
Input: A list of convergent embedders $\mathcal{F}$
Input: A list of convergent agents $\mathcal{A}$
Input: A universe U
let $\text{me} = \ulcorner \text{ddt}_\varepsilon(\mathcal{F}, \mathcal{A}, U) \urcorner$
let $P_i = \text{Ind}_\varepsilon(\mathbb{E}^{\mathsf{PA}}[|F(\text{me}) - U|] < \varepsilon)$
let $A_i = \text{sadt}_\varepsilon(\mathcal{F}_i, \mathcal{A})$
let $\Delta = \text{soft\_argmax}_\varepsilon(i \mapsto P_i \mathbb{E}^{\mathsf{PA}}[\mathcal{F}_i(A_i)])$
A_Δ

algorithm, which can learn the correct way to locate itself in the environment. In this section we will exhibit such an algorithm and prove optimality.

Intuitively, this algorithm searches for all the possible ways of embedding itself such that it would expect to get the same utility in that world that it expects to actually get, and picks the one where it's possible to do the best.

We will again use ddt to refer to $\lim_{\varepsilon \to 0} \text{ddt}_\varepsilon$ where convenient.

Theorem 3. $\text{ddt}(\mathcal{F}, \mathcal{A})$ *asymptotically dominates* $\mathcal{A}$ *on* $\mathcal{F}$.

Proof. (Note: pieces of this proof are a bit suspect, but I expect the result to still hold)

Let ρ be the function

$$\rho(i) = \lim_{t \to \infty} \mathbb{E}_t^{\mathsf{PA}}[F_i(\text{sadt}(F_i, \mathcal{A}))]$$

Let $\prec$ be the well-order defined by

$$i \prec j \leftrightarrow (\rho(i) > \rho(j)) \lor (\rho(i) = \rho(j) \land i < j)$$

The proof will proceed by induction on $\prec$. Take $F_i \in \mathcal{F}$. Suppose $\text{ddt}(\mathcal{F}, \mathcal{A})$ dominates $\mathcal{A}$ on all F_j with $j \prec i$. We will show by cases that $\text{ddt}(\mathcal{F}, \mathcal{A})$ dominates $\text{sadt}(F_i, \mathcal{A})$ on F_i.

First, suppose $\rho(j) > \rho(i)$ for all $j \prec i$. Then since there are finitely many such j, we can pick an ε small enough that

$$\rho(j) - \rho(i) > \varepsilon$$

for every $j \prec i$. Then either we do better than $\text{sadt}_\varepsilon(F_i, \mathcal{A})$, or their plausibility will approach 0 and Δ will converge to a δ-distribution on i, which by fairness implies we perform exactly as well as $\text{sadt}_\varepsilon(F_i, \mathcal{A})$ in the limit.

On the other hand, let $\mathcal{J}$ be the set of $j \prec i$ such that $\rho(j) = \rho(i)$. If F_j has zero plausibility for every $j \in \mathcal{J}$, then Δ will converge to a δ-distribution on i.

On the other hand, suppose F_j has positive plausibility for some $j \in \mathcal{J}$. By hypothesis, we dominate $\text{sadt}_j(F_j, \mathcal{A})$ on F_j. It must then be the case that

$$\lim_{t \to \infty} \mathbb{E}_t^{\mathsf{PA}}[F_j(\text{ddt}_\varepsilon(\mathcal{F}, \mathcal{A}))] \geqslant \rho(j) = \rho(i)$$

Thus in order for F_j to have nonzero plausibility, it must be the case that

$$\lim_{t\to\infty} \mathbb{E}_t^{\mathsf{PA}}[|F_i(\mathrm{ddt}_\varepsilon(\mathcal{F},\mathcal{A})) - F_j(\mathrm{ddt}_\varepsilon(\mathcal{F},\mathcal{A}))|] < \varepsilon$$

and thus in the limit as ε goes to 0 we must do as well in F_i as in F_j. □