Distributions Allowing Tiling of Staged Subjective EU Maximizers

Eliezer Yudkowsky

May 11, 2014, revised May 31

Abstract

This is a brief technical note summarizing some work done at the May 2014 MIRI workshop. We consider expected utility maximizers making a staged series of sequential choices, and replacing themselves with successors on each time-step (to represent self-modification). We wanted to find conditions under which we could show that a staged expected utility maximizer would replace itself with another staged EU maximizer (representing stability of this decision criterion under self-modification). We analyzed one candidate condition and found that the "Optimizer's Curse' implied that maximization at each stage was not actually optimal. To avoid this, we generated an extremely artificial function η that should allow expected utility maximizers to tile. We're still looking for the exact necessary and sufficient condition.

1 Setup

All maximization will be assumed to take place on a tree X with root x_0. $x_{n+1} \in x_n$ will range over the children of node x_n. $U(x_N)$ is the utility of the leaf node x_N. $\mathbb{E}[f]$ is the expected value of the function f given some assumed probability distribution $\mathbb{P}$. $\mathbb{E}[U|\phi]$ is the expected utility if the proposition ϕ is true. $\mathbb{E}[U|x_i]$ is the expected utility of choosing node x_i in the tree, presumably after already arriving at x_i's parent. The agent A_n will be at some node x_n occupying level n of the tree and be choosing from among its options $x_{n+1} \in x_n$.

Informally, what we want to show is that if A_n is also choosing to construct its successor agent A_{n+1}, then expected utility maximizers with well-calibrated probabilistic beliefs ought to construct expected utility maximizers with the same utility function and well-calibrated probabilistic beliefs. A key issue in showing this formally is that the expected utility of choosing x_n depends on what sort of agent will choose at x_{n+1}—which node has greatest expected utility now depends on whether you expect your successor to choose optimally, pessimally, randomly, etc. Ideally, we would also would like A_n to reason about offspring A_{n+1} that have updated on additional information, or possibly even thrown away low-value information (like not remembering every pixel of a webcam input) while still maintaining well-calibrated beliefs.

Most of our thinking before this workshop focused on probabilistic reflection principles in which $\mathbb{E}_0[f]$ relates to $\mathbb{E}_1[f]$ in some fashion that lets an agent using $\mathbb{E}_0$ conclude that if an agent using $\mathbb{E}_1$ expects high utility, then $\mathbb{E}_0$ expects high utility.

This paper is about a different angle in which we try to construct one particular probability distribution and expected utility function which will be apt to tiling maximizers reasoning about it, instead of trying to start from the general properties of $\mathbb{E}$. This should let us construct at least one tiling maximizer for the special case, which we can then examine further, although this has not yet been done formally. First, however, we'll introduce the problem and show why other distributions don't work.

2 Vingean blinders

In discussions prior to the workshop, the following equation was suggested:

$$\mathbb{E}[\max_{x_{n+1} \in x_n} \mathbb{E}[U|x_{n+1}]] = \max_{x_{n+1} \in x_n} \mathbb{E}[U|x_{n+1}]$$

This says that the expected utility of a node x_n is the maximum of the expected utilities of x_n's children. This is how a traditional maximization tree search would work on the parts of the tree that had been exhaustively searched.

The problem with this equation is that if it is true over every part of the search tree, it means that an expected utility maximizer is guaranteed to find the global maximum of the tree, i.e., the leaf with the highest expected utility among all leaves. Bounded agents will not have enough computing power to search the whole tree; they cannot *globally maximize.* Even if the leaf is an uncertain estimate, a bounded agent will usually not be able to search the whole tree to find the maximum among all uncertain estimates at leaf nodes.

We generally term the *Vingean principle* the notion that an agent should not be able to predict the exact actions of its successors. (Otherwise the first agent in a sequence of self-improvers would be able to predict out all the policies of all its successors out to the end of time).

We therefore want to evaluate *staged* subjective expected utility maximizers in which each node is only a subjective estimate of the maximum subjective expected utility available at the next stage, if the agent goes down that branch of the tree. Subjective estimates at one stage should be updated on additional information at the next stage, and sometimes be wrong. Bounded agents should not be guaranteed to find the global maximum; they should do what they guess to be the best thing at each next step (taking into account that their future selves will likewise do the apparently best thing on each successive step while taking into account that their successors will do the same, etc.) That this is a guess implies that it should not be assumed to be perfectly accurate.

A requirement that expresses this principle is:

$$\mathbb{E}[\max_{x_{n+1} \in x_n} \mathbb{E}[U|x_{n+1}]] \neq \max_{x_{n+1} \in x_n} \mathbb{E}[U|x_{n+1}] \tag{1}$$

(note the not-equals sign). Another way of expressing the principle is that within the calculations of an agent at level n, maxes ranging over all $n+1$-level children of an n-level node should only appear inside expectation operators, not outside them as in the right-hand-side of (1). The expectation operator acts as *Vingean blinders*—the 'parent' is not allowed to know the *exact* maximum EU of the next level down at the point that it chooses, only an expectation of this maximum.

This corresponds to a realistic, bounded agent which deals with a large search space by slicing up the problem into staged categories. At each stage, the best category of possible decisions is selected, that branch is then further investigated, the best category within that branch is taken, and so on until a leaf node is reached. (This corresponds well to sequential game playing in which the opponent's move is learned after each of our own moves, but it is also a potentially general way of slicing up search spaces.)

3 The Optimizer's Curse

An obvious next avenue of investigation is to hypothesize a function $\phi(x_n)$ which is an unbiased but noisy estimator of the maximum of the estimates available at the next stage:

$$\mathbb{E}[\phi(x_n)] = \max_{x_{n+1} \in x_n} \phi(x_{n+1}) \,.$$

To generate an example $\phi(n)$ with this property, we can start by labeling all leaves of the tree with their true utilities $U(x_N)$. We then add normally distributed random noise to obtain $\phi(x_N) = U(x_N) + N(0,1)$. We then recursively obtain ϕ of the parent nodes x_{n-1} by taking

$$\phi(x_{n-1}) = \max_{x_n \in x_{n-1}} \phi(x_n) \; + \; N(0,1)$$

ϕ has the desired property that $\phi(x_n)$ is an unbiased estimator of the maximum over $\phi(x_{n+1})$, with the leaf nodes $\phi(x_N)$ being unbiased estimators of the utility of x_N. An even weaker condition also fulfilled by $\phi(x_n)$, which was previously proposed as a possible sufficient condition for tiling, is:

$$\mathbb{E}[\phi(x_n) - \max_{x_{n+1} \in x_n} \phi(x_{n+1})] = 0$$

Note that in accordance with the Vingean principle, no maximization operators appear outside of expectation functions in the above equation.

ϕ however is subject to the Optimizer's Curse described by Smith and Winkler (2006). The Optimizer's Curse means that it is probably not prudent to choose the option with greatest ϕ at every stage.

Intuitively, the problem is that the max operator selects for positive noise. Suppose one node $x_{n,i}$ had a thousand children all with true utilities 0, while another node $x_{n,j}$ had ten children all with true utilities zero. Our expectation of

$\phi(x_{n,i})$ is much higher than our expectation of $\phi(x_{n,j})$ because our expectation of the maximum of a thousand draws from a normal distribution is higher than our expectation of the maximum of ten draws from a normal distribution, even though all this variance is pure noise and the true values are zero in every case. So if $\phi(x_{n,i})$ were only slightly higher than $\phi(x_{n,j})$, the prudent choice would be to go down the branch for $\phi(x_{n,j})$. This means that a prudent expected utility maximizer that understands how the noise in the unbiased estimator interacts with its own tendency to maximize, will not always go down the path of greatest ϕ. (This also shows why there are nontrivial issues with tiling subjective maximizers that pick the best action conditional on their successors also maximizing, rather than doing something else like taking the average action, etc.)

4 The η-distribution: Unbiased estimates instead of unbiased estimators.

One way of viewing the problem with ϕ is that if our estimator is noisy, then when we see a ϕ with high value, we expect to some degree that it has had high noise added. So when we follow the ϕ we are predictably disappointed. Smith and Winkler's suggested solution to the Optimizer's Curse for the special case they considered is to regress their Gaussian-noisy estimates toward the mean, which they show is equivalent to treating the "unbiased estimator" as evidence that combines with a Bayesian prior. Once this is done, Smith and Winkler show that expected disappointment after choosing the highest estimate is zero.

To generalize, the problem with a standard "unbiased estimator" V of a true value μ is that it merely has the property:

$$\forall x : \mathbb{E}[V|\mu = x] = x$$

But this is an insufficient condition for plugging V into decision-making—e.g., for treating V as an estimate of expected utility. A sufficient condition is:

$$\forall x : \mathbb{E}[\mu|V = x] = x \tag{2}$$

although this is a *very strong* condition (see below).

The distribution η was constructed to fulfill this condition and make ηx_n an unbiased *estimate* (not estimator) of $\max_{x_{n+1} \in x_n} \eta(x_{n+1})$ regardless of which value of ηx_n is observed. It is constructed as follows:

$$
\begin{aligned}
\eta(x_0) &:= N(0,1) \\
\text{for } x_{n+1,k} \in x_n : & \\
i &:= \text{Random}(K(x_n)) \\
\eta(x_{n+1,i}) &:= \eta(x_n) + N(0,1) \\
\eta(x_{n+1,j\neq i}) &:= \eta(x_{n+1,i}) - |N(0,1)| \\
\\
U(x_N) &:= \eta(x_N) + N(0,1)
\end{aligned}
$$

We start with the root node and generate a random value of η. Then we generate children recursively as follows: For a parent with a known value, add noise with mean zero to this known value to get the maximum child that will appear on a random branch beneath that parent. Then assign random values less than the maximum child's value to all other children of that parent. When a leaf node is reached and assigned a value of η, assign it a random utility equal to η plus noise with mean zero.

Although this probability distribution over η and U was constructed in a somewhat perverse way (starting from the map and generating the territory), it makes $\eta(x_n)$ a correct estimate of the U which will be achieved by further staged maximization down the branch x_n. To achieve maximum expected U, assuming you can only look ahead to the nodes immediately below you, it is always wisest to choose the node with maximum η at your current stage. This remains true even if noise is much greater in some parts of the tree than others, or some parts of the tree are much wider or deeper than others.

η demonstrates one property a probability distribution or expectation function can have that seems like it should suffice for staged maximizers to tile over it. Equation (1) is fulfilled by η and thus it does not usually find global maximums of η or U. So η is a fair model of staged, subjective maximization rather than global, objective maximization.

A next step is to write out a tiling of a staged subjective EU maximizer for the first time, using η. This has not yet been done but looks easily reachable from here.

5 Problem: η obeys overly strong conditions.

A problem with η is that the condition (2) which η fulfills is extremely strong. (2) states that for every possible estimate x, conditioning on the estimator V returning the value x, we believe that x is in fact the best expected estimate for the underlying value μ. Suppose you were about to construct an offspring similar to yourself. Do you, considering the hypothetical case that your offspring has come to believe the best estimate of the mass of an electron is a million grams, believe that your own best estimate of the mass of the electron, conditioning on this fact, is a thousand grams? The condition (2) can be seen as demanding that, in whatever distribution we have over (impossible) possible worlds, the possibility of our offspring assigning an estimate $V = x$ corresponds to a group of possible worlds in which $V = x$ and the true value μ has an average value of x. This is a counterintuitively strong condition on possible worlds, if μ is the true mass of an electron and x is a million grams.

Note that by construction of η, the probability distribution for $\eta(x_n)$ has finite probability density at every possible real value $\eta(x_n) = y, -\infty < y < +\infty$. Even if $\eta(x_0) = 10^6$ or some other value with extremely tiny probability, nonetheless, among possible worlds where $\eta(x_0) = 10^6$, the expected maximum $\eta(x_1)$ that is found is 10^6.

The property (2) is also proving difficult to obtain from Christiano-style reflective probability distributions (Christiano et al. 2013), even with error terms added in.

Thus a key question is whether there's some intermediate guarantee between the too-weak

$$\mathbb{E}[\phi(x_n) - \max_{x_{n+1} \in x_n} \phi(x_{n+1})] = 0$$

and the very strong

$$\forall y \in \mathbb{R} : \mathbb{E}[\max_{x_{n+1} \in x_n} \eta(x_{n+1}) | \eta(x_n) = y] = y \,,$$

that will also enable staged EU maximizers to tile.

References

Christiano, Paul, Eliezer Yudkowsky, Marcello Herreshoff, and Mihaly Barasz. 2013. "Definability of Truth in Probabilistic Logic." Early Draft. http://intelligence.org/files/DefinabilityTruthDraft.pdf.

Smith, James E., and Robert L. Winkler. 2006. "The Optimizer's Curse: Skepticism and Postdecision Surprise in Decision Analysis." *Management Science* 52, no. 3 (March): 311–322. doi:10.1287/mnsc.1050.0451. https://faculty.fuqua.duke.edu/jes9/bio/The_Optimizers_Curse.pdf.