Jan. 23, 2019

Toby,

I think that's good news that Grace is thinking of becoming a physical therapist. My impression, both as a patient and as someone involved in the medical scene here in Corvallis, is that physical therapy provides both valuable health benefits to the patients and job satisfaction to the therapist. For example, as compared to some other medical fields, a physical therapist can spend more time working with people and less time with other activities.

Dolly and I are pleased that you and Grace will be coming to our birthday party. It will be in late June—we'll let you know once we start to plan.

On your problem, most likely we are looking at the same thing but expressing our thoughts somewhat differently. In any case, here are some of my thoughts, beginning with some definitions, notation, and assumptions.

By a single-stage decision model, I will mean a structure having the following features: (1) a set of options, (2) a means of determining which options are possible, (3) for each option, an outcome or a probability distribution of outcomes, (4) a function to evaluate each outcome or probability distribution of outcomes. By a multi-stage decision model, I will mean a succession of single-stage decision models joined in some manner.

By a <u>solution</u>, I will mean a possible option or succession of options, and by an <u>optimal solution</u>, I will mean a possible option or succession of options such that no other possible option or succession of options has a better evaluation. And by an <u>algorithm</u>, I will mean a method of finding an optimal solution.

For the type of decision model that you consider, there are $T \geq 1$ times, $t = 1, \ldots, T$, and at each time t the decision maker chooses a <u>state</u> s from a set $S = \{ s_1, \ldots, s_m \}$. An option is a state s, and an outcome is a <u>value</u> v from a numerical set $\{ v_1, \ldots, v_n \}$. Assume that before choosing (and thus observing) a state s, the value of s is a random variable $\boldsymbol{v}(s)$ with a probability distribution $p_1(s), \ldots, p_n(s)$ on its possible values v_1, $\ldots$, v_n. Assume that after choosing (and thus observing) a state s the decision maker knows the value of s, to be denoted by $v(s)$. Assume that the probability distribution $p_1(s), \ldots, p_n(s)$ and the value $v(s)$ are the same for any time t.

Let $E(\boldsymbol{v}(s)) = p_1(s)v_1 + \ldots + p_n(s)v_n$ denote the expectation of the value of s before observing s (and as noted let $v(s)$ denote the value of s after observing it). Assume that the probability distributions $p_1(s), \ldots, p_n(s)$ are independent of one another, i.e., observing one state tells the decision maker nothing about the value of any other state.

Since there are m choices of a state at each of T times, there are m^T sequences of choices. Such a sequence of choices will be called a <u>choice</u>-<u>vector</u> and will be denoted by $\boldsymbol{s} = (s(1), \ldots, s(T))$. As in other multi-stage decision models, a choice at a time $t > 1$ may depend on what the decision maker has learned by his choices at the previous times.

At the initial time $t = 1$, the decision maker will evaluate a state s by the expected value $E(\boldsymbol{v}(s))$. At a later time $t > 1$, the decision maker will evaluate a state s that has not been previously chosen by $E(\boldsymbol{v}(s))$ and will evaluate a state s that has been previously chosen by $v(s)$. Assume that the evaluation of a choice-vector $\boldsymbol{s}$ is a strictly increasing function $U(s(1), \ldots, s(T))$ of the evaluations $E(\boldsymbol{v}(s))$ and $v(s)$ of the states $s(1), \ldots, s(T)$ in $\boldsymbol{s}$. In the single-stage case, $T = 1$, $U(s(1))$ is a strictly increasing function of $E(\boldsymbol{v}(s(1)))$. In the two-stage case, $T = 2$, $U(s(1), s(2))$ is a strictly increasing function of $v(s(1))$ and of either $v(s(2))$ or $E(\boldsymbol{v}(s(2)))$ depending on whether $s(2) = s(1)$ or $s(2) \neq s(1)$. And so forth.

This description completes our definition of a class of multi-stage decision models. Let's call a model in this class a <u>Harvey</u> <u>model</u>.

Normally, in consulting work the task is to construct a model that describes the situation at hand, and in research the task is to construct an algorithm that can solve a given class of models. But in our emails, we have intertwined these two tasks. That's what I meant when I said "backwards but cool."

As the next step in this project, I need to complete our definition of what I have called a maximize-explore algorithm. An algorithm to solve a Harvey model will be called a <u>maximize</u>-<u>explore</u> <u>algorithm</u> provided that at each time t the decision maker either chooses a state s that has been previously chosen such that $v(s)$ is a maximum for those states that have been previously chosen or chooses a state s that has not been previously chosen such that $E(\boldsymbol{v}(s))$ is a maximum for those states that have not been previously chosen.

In this email, let's consider only a special type of Harvey model, namely those Harvey models in which the evaluation function $U(s(1), \ldots, s(T))$ depends only on the state $s(T)$ and thus on the evaluation $v(s(T))$ or $E(\boldsymbol{v}(s(T)))$ of $s(T)$. Such a model will be called an <u>only-the-end-matters model</u>. I conjecture that the following maximize-explore algorithm will find an optimal solution of any only-the-end-matters model.

If $m < T$, then at each time $t \leq m$ choose a different state and at each time $t > m$ choose a state s whose value $v(s)$ is a maximum over all the states.

If $m \geq T$, then at each time $t < T$ choose a different state. Let $\max \{v(s)\}$ denote the maximum value of these observed states, and let $\max E(\boldsymbol{v}(s))$ denote the maximum expected value of the unobserved states. (There will be such states since $m > T–1$.) At time $t = T$, if $\max \{v(s)\text{: } t < T\} \geq \max E(\boldsymbol{v}(s))$ then choose a previously observed state whose value equals $\max \{v(s)\}$, and if $\max \{v(s)\} < \max E(\boldsymbol{v}(s))$ then choose a previously unobserved state whose expected value equals $\max E(\boldsymbol{v}(s))$.

More complicated Harvey models might be of interest for machine learning, but I have no insights here since I haven't thought about artificial intelligence in eons. I can, however, imagine the following types of models.

(1) Life before the end matters too: In these models, the evaluation function $U(s(1), \ldots, s(T))$ depends on all of the chosen states $s(1), \ldots, s(T)$.

(2) Observations are imperfect: In these models, observation of a state s does not tell the decision maker the value $v(s)$ of s. Instead, it provides a Bayesian updating of the probability distribution for the value of s.

(3) No state is an island: In these models, observation of a state provides an updating of the probability distributions for the values of all m states. In particular, observation of a state might provide information on which state to observe next.

Writing this up has been fun, and I hope the resulting story will be of interest to you. In any event, best wishes to you and Grace, and do let Dolly and I know when you might be in Oregon.

Mac