Causal Identification of the Effects of POMDP Actions with non-Random Treatment Assignment

Ima Pseudonym
pseudonym@dev.null

Abstract

Response-to-intervention (RTI) is an educational framework for placing students into an appropriate level of support. The ability of the students is measured at several time points and the lowest performing students are placed into supplemental Tier 2 instruction. This framework is naturally modeled with a partially observed Markov decision process (POMDP), but using the POMDP model for planning requires an estimate of the effect of the Tier 2 instruction. This can be estimated from historical data, but unless the mechanism by which the treatments were assigned in the historical data are causally independent, the estimate will be biased. In particular, if the treatment (action) assignment is made purely on the basis of observed data, then the causal effect can be identified, but if teacher judgment, or some other unrecorded variable, is used to determine the treatment, the data will not meet the backdoor criteria for causal identifiability. This paper explores the implications of the lack of causal identifiability through a simple simulation study.

Key words: POMDP, Causal Identification, Application (Education)

1 Introduction

Partially observed Markov decision processes (POMDP Boutilier, Dean, & Hanks, 1999) provide a variety of algorithms for finding an optimal sequence of decisions; however, these algorithms all rely on having an estimate of the effects of various actions that can be taken at each time point. If the effects are unknown, they must be estimated from an existing database of measurements. However, in many databases, the mechanism by which actions is assigned is not randomized, thus the treatment (action) assignment may depend on observed or even unobserved variables. In this case, the effects of the action may not be causally identified.

As an example, consider the educational policy of response-to-intervention (RTI Fuchs & Vaughn, 2012). In RTI students are split into two (or three) tiers based on their scores on a pretest: Tier I, regular classroom instruction, and Tier II regular instruction supplemented by small group instruction. Clearly Tier II is more expensive, and the goal is to maximize student performance subject to a budget constraint on how many Tier II seats are available. The weakly coupled POMDP algorithms of Boutilier and Liu (2016) seem ideal for this, but the algorithm requires an estimate of the treatment effect for the Tier II instruction. Instructors are supposed to monitor student performance for Tier II students, adapting the instruction if the student is not properly responding to the intervention. Here the ability of POMDP models to produce forecasts under different policies seems like a tool that would be helpful.

RTI has been implemented in classrooms for quite some time, and there exist suites of tools, such as easyCBM (Alonzo, Tindal, Ulmer, & Glasgow, 2006), which support teachers implementing an RTI program. Consequently, large databases of student measurements exist, but as these are gathered over a variety of schools and districts different policies are used for assigning students to Tier II. While many schools use a strictly mechanical rule based on the screening test scores, others allow teacher's discretion (Mellard, McKnight, & Woods, 2009; Jenkins, Schiller, Backorby, Thayer, & Tilly, 2012). Worse, in the real data there is often a mixture of policies, and establishing what the policy is for each school and district is prohibitively expensive.

This paper looks at the issue of identification of the effects of actions in a POMDP when there exists a potential unmeasured relationship between the treatment assignment and the latent variable. It attempts to find bounds on the causal effects using a sensitivity analysis.

2 A Simplified Response-to-Intervention Model

RTI is method for delivering educational interventions which has been shown to be effective in closing achievement gaps (Fuchs & Vaughn, 2012). Although the details vary, in a typical RTI situation, students are given a screening test three times in an academic year. On the basis of the screening test, students are assigned to one of three tiers of instruction. Tier 1 is continued whole class instruction, Tier 2 is small group instruction in addition to the whole class instruction, and Tier 3 is individual instruction. In some implementations, Tier 3 is an assignment to a special education classroom. For simplicity, this paper only considers Tiers 1 and 2.

There is considerable variability in how the assignment to the tiers is done (Mellard et al., 2009; Jenkins et al., 2012). Often it is implemented as a simple cut score on the screening test, but in some cases the teacher could use expert judgment to override the cut score. In many cases, there is a limit to how many students can be assigned to Tier 2 based on constraints such as the amount of time the teacher, aid, or specialist can spend on small group work. These limits could be set at the classroom, school or district level (for example, a reading specialist could be shared across several schools).

The name *response-to-intervention* comes about because of what happens within Tier 2. In Tier 2 students are given more frequent (often weekly) progress monitoring tests. If students are not making "adequate progress" the intervention should be changed; possibly changing the intensity (meeting more frequently, for longer periods or with smaller group sizes) or the curriculum or approach changed. In extreme cases, the student might be moved to Tier 3 or, if the student did unexpectedly well, returned to Tier 1. The definition of adequate progress is vague, and it is clear that a planning system could help educators forecast the effects of changing the educational plan for a student.

For the purposes of this paper, the RTI process will be simplified by simply considering the Tier 1 and Tier 2 assignments without looking at the progress monitoring. Also, for simplicity only one intensity of Tier 2 treatment will be considered. Finally, additional screening tests will be included so that there are more decision points. The goal is to estimate the effect of the Tier 2 assignment so that it can be used in planning.

2.1 Common Data Layout

Let I be the number of students, and T_i be the number of measurements made on Student i. Let $T_{max} = \max_{i \in I} T_i$.

Let $obs_{t,i}$ be the observation for Student i on the tth measurement occasion. Let $Time_{t,i}$ be the elapsed time between measurement occasion t and $t+1$ for Student i and let $Dose_{t,i}$ be the dosage of treatment received by Student i between times t and $t+1$. In general the dose will be the treatment intensity multiplied by the elapsed time.

Let $\theta_{t,i}$ be the proficiency of Student i at measurement occasion t. For simplicity, both $\theta_{t,i}$ and $obs_{t,i}$ will be taken as unidimensional even though the multidimensional case is more interesting (e.g., if the overall proficiency is reading, the students ability to decode works and comprehend sentences could be separate measures and addressed with different interventions).

Note that the indexes are backward from the usual description so these can be described as a one-dimensional array of vectors in Stan (Stan Development Team, 2013).

3 Common Evidence Model

Another problem that arises with the educational context is that the measurement instruments are different at each time point. In contrast, in a model trying to find the position of a robot, the same instruments (with the same measurement properties) are used to measure the robot's position at each time point. In an educational setting the instrument is a test, but the same test cannot be used repeatedly. For example, if the same reading passage was used over and over increases in comprehension or reading fluency could be due to familiarity with the specific passage. Therefore, the measurement models consist of a collection of instruments for each time point, each with potentially different relationships to the target latent variable, θ.

Almond, Tokac, and Al Otaiba (2012) illustrate another possible identification issue which arises if both the average growth rate and the difficulty (negative intercept in a regression model) and discrimination (slope in a regression model) of the instrument must be estimated from the same data. If on average the students score higher at Time 2 than at Time 1, it

is impossible to tell if an observed difference is score is due to student improvement, or a difference between the forms of the test administered at Time 1 and Time 2, or some combination. Almond, Goldin, Guo, and Wang (2014) identify two approaches to this problem: (1) perform some kind of data collection designed to put all of the instruments on a common scale, and (2) assume that the average growth is the same as the average change in difficulty and examine deviations from stationarity.

The easyCBM product (Alonzo et al., 2006) uses the first approach. A separate calibration study was done where a number of different forms of the progress monitoring instruments were given to students at about the same time so that they could be placed on a common scale. Furthermore, this initial calibration study establishes the parameters that link the observations to the latent variable. This is the approach taken in this simulation.

The relationship between the latent variable and the observation is assumed to be a simple latent regression:

$$Obs_{t,i} \sim N(obs_{int} + obs_{slope}\theta_{t,i}, res_{std}) \qquad (1)$$

The three parameters which control equation 1 are further defined in terms of other parameters. In Psychometrics, the *reliability* of an instrument is defined as the correlation between two different readings from an instrument taken under identical conditions. Let obs_{rel} be the reliability of the instrument, $obs_{std,1}$ be the standard deviation of the scores at the first measurement occasion, and $obs_{mean,1}$ be the mean of those scores. To identify the latent scale, $\theta_{i,1}$ is assume to have a standard normal distribution. Therefore,

$$\begin{aligned} obs_{int} &= obs_{mean,1} \\ obs_{slope} &= obs_{std,t}\sqrt{obs_{rel}} \\ res_{std} &= obs_{std,t_1}\sqrt{1 - obs_{rel}} \end{aligned}$$

This should ensure that the scale at the initial time point is properly identified.

3.1 Variable Slopes Model

The model used in this study assumes that students' abilities grow according to a Wiener process with drift. That is, between each time point there is an independent increment to each student's ability, and those increments accumulate over time. The process is assumed to have drift as the students are actively receiving instruction, and the average trend will depend on the instruction received.

The average growth (or drift) has two components a natural growth component and a treatment effect. Students in Tier I receive the normal instruction and only exhibit normal growth. Students in Tier II receive both normal instruction and some kind of supplemental instruction; thus, their growth with have both natural and treatment effects. The variable $Dose_{t,i}$ indicates how much supplemental instruction each student receives between measurement points t and $t+1$. It is zero for students in Tier I and positive for students in Tier II.

Using this decomposition for the average learning gain, the change in the latent proficiency can be decomposed as:

$$\theta_{t+1,i} = \theta_{t,i} + slope_i * Time_{t,i} + treat_{eff} * Dose_{t,i} + \epsilon_{t,i} \qquad (2)$$

In this equation, the natural growth rate, $slope_i$, varies by person, but the treatment effect does not. Also, it is assumed that the treatment effect and natural growth rate are additive. Finally, to make this a Wiener process, the variance of the innovation term, $\epsilon_{t,i}$ depends on the elapsed time, $Time_{t,i}$; in particular, $\epsilon_{t,i} \sim N(0, \sqrt{var_{innov} Time_{t,i}})$.

Willett (1988) notes that there is often a correlation between the slope and the initial value in growth curves.[1] This is because the first measurement occasion is often not the true time zero. Consider a growth curve for reading in Kindergarten students. Most students will have received some kind of pre-reading instruction either through home or pre-school. So even if the first measurement occasion is the first day of class, they still will have received prior instruction. Students who naturally grow at a faster rate are likelier to then be at a higher level when first measured. Students entering Kindergarten vary considerably in the amount of pre-school they may have attended and the number of reading related activities that they do in their home life, so the effective time zero may vary from student to student.

To capture this idea, the slope distribution is characterized with three parameters, $slope_{mu}$, $slope_{std}$ and $slope_{r2}$. The last parameter is the correlation between the $slope_i$ and $\theta_{i,1}$. To capture this relationship, the slopes are made dependent on the initial proficiencies as follows:

$$slope_i = slope_{mu} + slope_{std}(\sqrt{1 - slope_{r2}^2}\phi_i + slope_{r2} * \theta 1, i, \qquad (3)$$

where both $\theta 1, i$ and ϕ_i have unit normal distributions.

[1] XXX, personal communication, has indicated that she has found this correlation to be both positive and negative across many studies involving pre-school children.

3.2 Tier assignment policies

If the goal is to identify the treatment effect, $treat_{eff}$, then ideally the treatment would be randomly assigned. This is often done in trials for specific interventions. However, collecting data under controlled conditions is fairly expensive, especially when considering that often tests with pre-schoolers require human administration and strict fidelity checks are needed to ensure uniformity of the treatment. Even a study with a million dollar budget can usually only afford to measure several hundred students on 3 time points in a year.[2]

The alternative is to go to databases of student measurement that are gathered through normal educational applications of an RTI system. There are two problems. First, as no fidelity checks are done on the treatment, there is likely considerable variability in the efficacy of the implementation. Second, different districts, schools and classrooms may use different policies for assignment into the tier groups.

Examine two different policies. The first will be based on a simple cut score model. The second will allow the teacher to override the cut score with expert judgment.

Cut Score Policy. This is the easiest policy to implement: if $obs_{t,i} < cut_t$ then Student i is assigned to Tier 2, otherwise to Tier 1. Mellard et al. (2009) and Jenkins et al. (2012) surveyed a number of schools and found a fair number of them using variations on this policy. Often the cut score is set to allow a certain number of students into Tier 2. In the case of the simulation study described here, the cut score for each time point is set to catch students who are one standard deviation down from the expected observation score at each time point.

Cut Score with Override Policy. This policy is meant to emulate the situation where the cut score rule is in place, but the teacher may use expert judgment to override the scoring rule. In this scoring rule, the teacher uses personal observation of the student to assess the student's value of $\theta_{t,i}$. If the teacher chooses to override, then the student is assigned to Tier 2 if $theta_{t,i} < cut_t$. It is assumed that the teacher overrides with a certain probability $override_p$, and that the override decision is made independently for each student (and independently of θ).

The assumptions in the cut score with override policy are unrealistic, but this is more or less designed to be a worst case scenario for causal identification. Also, the cut score policy is a special case of the cut score with override policy with the override probability set to zero, which is convenient for implementation.

[2] XXX, personal communication.

4 A simple simulation study

To assess whether or not the treatment effect could be recovered under ideal conditions, a simulation study was performed. Data was simulated for 400 student at 10 time points using both of the two policies (simulation code is in the accompanying file `varSlopesSim2a.R`). Then the model (`varSlopes2.stan`) was fit using Stan (Stan Development Team, 2013). Five chains were run for 2000 interactions each (with 1/2 used for warm up). The usual tests indicated that for both simulations the chains had reached the stationary distribution. (The accompanying file `varSlopesRun2.R` shows the model fitting and checking code.)

In both simulations, the treatment effect was set to .25, corresponding to growth of 1/4 of a standard deviation over an academic year (a fairly typical effect size for an educational intervention). In the simulation using the simple cut score policy, the mean treatment effect posterior was .13 with a standard deviation of .09, a median of .11, and a 95% credibility interval of .01 to .34; which contains the true simulation value. For the cut score with override policy, the override probability was set to .5. In this simulation, the posterior mean was .04, the standard deviation, .03, the median, .03 and the 95% interval .00 to .12, clearly an under estimate.

5 Causal Identification

So why does the policy without override produce an apparently unbiased estimate, and allowing the teacher to override produce a biased estimate? The answer can be found by trying to see whether or not the effect of the treatment is causally identified by the data (Pearl, 2009). Examine Figure 1(a), which corresponds to the cut score without override policy. In this case, as obs_i is observed, there is no backdoor path to θ_2 or obs_2 from $Dose_1$, so its effects are causally identified.

For the cut score with override policy, Figure 1(b), there is an extra dashed edge from θ_1 to $Dose_1$. This introduces a backdoor path which destroys the causal identification. Thus, the estimates from this model are biased.

Note that this could also be cast as a model misspecification problem rather than a causal identification problem. In particular, if the mechanism corresponding to the dashed arrow were known, and added to the MCMC model, the act of dosing becomes another observation. The policy parameters corresponding to the dashed line (e.g., the override probability) are not estimable from data, but a sensitivity analysis could be

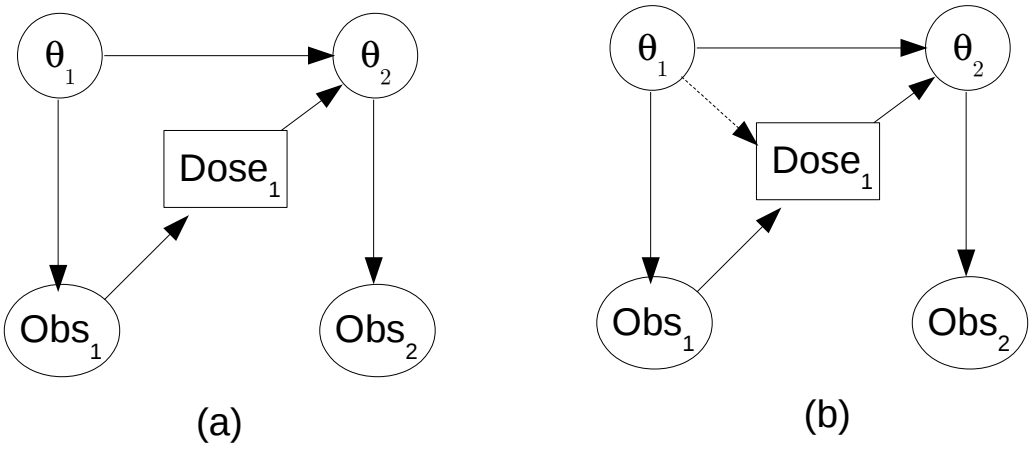

Figure 1: Two time points of the model under the cut score policy with (b) and without (a) override.

performed by trying a range of parameters for the override mechanism. This would at least produce bounds for the size of the treatment effect. This is obviously the next step for this research.

6 Discussion

Intuitively, finding an optimal policy for a POMDP requires first finding good estimates of the probable effects of the various actions. However, unbiased estimates of those effects depend on the mechanism by which actions are assigned in the training data. In particular, a problem might exist if all of the variables used to assign the action are not observed. In particular, to create unbiased estimates of the action effects one of two conditions must hold: (1) actions are assigned only on the basis of observed variables, or (2) the mechanism by which the action assignment is related to the latent variables is explicitly modeled. In the latter case, a suitable parameterization of the model can allow bounds for the causal effect of the action to be calculated using a sensitivity analysis.

The good news is that in a typical POMDP policy, the action selection is made on the basis of a function of the sequence of observations. Even though this is more complex than the simple example provided here, it is still sufficient to satisfy the no backdoor path criterion, and so the causal identification holds. The problem comes when the database used for estimate is based on historical records where the mechanism for action assignment was not recorded. Here the potential use of expert judgment could open a backdoor that would cause a problem with the causal identification.

There are several other issues with these data that have not yet been addressed. The first is structural missingness. Students who are assigned to Tier 2 are typically measured more often than students who are assigned to Tier 1. For students in Tier 2 the model is effectively estimating $slope_i + treat_{eff}$, while for Tier 1 it is only estimating $slope_i$, but with fewer time points. It is unclear if this will cause problems (an increase in the posterior variance is likely).

A second issue is that it was assumed that the Tier 2 effect was uniform and did not vary from person to person or depend on the state of $\theta_{t,i}$. Almond (2007) suggested that the effect of an educational intervention was likely to be highest for students at or near the proficiency level for which it was defined. Using a more sophisticated model for the treatment effect is probably appropriate.

A third issue is that the Tier 2 treatment is applied to a small group. As the Tier 2 treatment is applied to small groups (and sampling is usually done at the classroom or higher level so that all of the students in the group are included in the sample), this calls into question the stable unit treatment value assumption. Again, a more complicated model is needed to model this dependency. Still, the procedure explored here provides a way to start to approach the problem of applying AI planning techniques to classroom decision making.

References

Almond, R. G. (2007). Cognitive modeling to represent growth (learning) using Markov decision processes. *Technology, Instruction, Cognition and Learning (TICL)*, *5*, 313–324. Retrieved from http://www.oldcitypublishing.com/TICL/TICL.html

Almond, R. G., Goldin, I., Guo, Y., & Wang, N. (2014). Vertical and stationary scales for progress maps. In J. Stamper, Z. Pardoz, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th international conference on educational data mining* (pp. 169–176). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/long20papers/169_EDM-2014-Full.pdf

Almond, R. G., Tokac, U., & Al Otaiba, S. (2012, August). Using POMDPs to forecast kindergarten students reading comprehension. In J. M. Agosta, A. Nicholson, & M. J. Flores (Eds.), *The 9th Bayesian modelling application workshop at UAI 2012.* Catalina Island, CA. Retrieved from http://www.abnms.org/uai2012-apps-workshop/papers/AlmondEtal.pdf

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM online progress monitoring assessment sytem* (Tech. Rep.). Behavioral Research and Teaching. Retrieved from http://easyCBM.com/

Boutilier, C., Dean, T., & Hanks, S. (1999). Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, *11*, 1-94. Retrieved from citeseer.ist.psu.edu/boutilier99decisiontheoretic.html

Boutilier, C., & Liu, T. (2016). Budget allocation using weakly coupled, constrained markov deicsion processes. In A. Ihler & D. Janzing (Eds.), *Uncertainty in artificial intelligence (UAI): Proceedings of the thirty-second conference* (pp. 52–61). Association for Uncertainy in Artificial Intelligence (AUAI). Retrieved from http://auai.org/uai2016/proceedings/papers/246.pdf

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*, 195-203. doi: 10.1177/0022219412442150

Jenkins, J. R., Schiller, E., Backorby, J., Thayer, S. K., & Tilly, W. (2012). Response to intervention in reading: Architecture and practices. *Learning Disabilities Quarterly*, *36*(1), 36–46.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research and Practice*, *24*, 186–191.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146. Retrieved from http://ftp.cs.ucla.edu/pub/stat_ser/r350.pdf doi: 10.1214/09-SS057

Stan Development Team. (2013). Stan: A C++ library for probability and sampling (Version 2.2.0 ed.) [Computer software manual]. Retrieved from http://mc-stan.org/

Willett, J. B. (1988). Questions and answers in the measurement of change. *Review of Research in Education*, *15*, 345-422.

Acknowledgments