Discovering ‘Tough Love’ Interventions Despite Dropout

Joseph Jay Williams
Harvard University
125 Mt Auburn St
Cambridge, MA 02138
joseph_jay_williams@harvard.edu

Anthony Botelho
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609 USA
abotelho@wpi.edu

Adam Sales
University of Texas
Statistics
Austin, TX 78712
asales@utexas.edu

Neil Heffernan
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609 USA
nth@wpi.edu

Charles Lang
New York University
239 Greene St
New York, NY, 10003
charles.lang@nyu.edu

ABSTRACT

This paper reports an application to educational intervention of Principal Stratification, a statistical method for estimating the effect of a treatment even when there are different rates of dropout in experimental and control conditions. We consider the potential value for using principal stratification to identify “Tough Love Interventions” – interventions that have a large effect but also increase the propensity of students to drop out. This method allowed us to generate an estimate of the treatment effect in an RCT without the selection bias induced by differential attrition by restricting analysis to just the inferred “stratum” of students who would not drop out in either condition. This paper provides a case study of how to appropriate the method of principal stratification from statistics and medical research fields to educational data mining, where it has been largely absent despite increasing relevance to online learning.

Keywords

principal stratification; selection bias; statistics; attrition; noncompliance; randomized controlled trial; experiment; online education

1. INTRODUCTION

A persistent problem in interpreting randomized experimental comparisons in learning environments is that the frequency of student dropout may vary between conditions . This is known as *differential attrition*, and causes problems with statistical inference [3] regarding the magnitude and direction of differences between treatment and control conditions. In cases where student completion is the metric of interest, such differences in condition are easily measured by the number of students to complete each; a problem arises, however, when performance is the metric of interest, as if less students drop out of one condition than the other, it is over-represented in the analysis causing unreliable results.

Differential attrition can mask the existence of what we label **“tough love”** interventions (TLIs). A TLI describes an intervention which introduces a treatment condition with features that cause some students to drop out, but has beneficial effects for students who persist. It is important to know *how much* such interventions impact a potential outcome in order to perform a cost-benefit comparison against the dropout rate. We believe that principal stratification is one tool that can be used to measure the effect of conditions in the presence of differential attrition and help identify TLIs.

2. ILLUSTRATIVE EXPERIMENT: IMPACT OF QUESTIONS ABOUT CONFIDENCE

In the preliminary data presented here, we consider a randomized controlled experiment (RCE) conducted within ASSISTments, a K-12 online and blended learning platform, reported in EDM 2015 [4]. Students were randomly assigned to either a condition of Treatment, where students were asked about their confidence in solving problems, or Control, where students were asked about technology usage. The data set used for analysis consists of 712 12-14 year olds in the eighth grade of a school district in the North East of the United States with 5,861 log records collected while students were solving math problems. The goal here is to estimate how the conditions differ in their impact on Mastery Speed, the number of problems needed to reach three consecutive correct responses indicating a sufficient level of understanding. It is important to note that a lower value in this metric indicates better performance.

3. ANALYTIC STRATEGY

Principal stratification [2, 5] is an approach to modeling causal effects for a subset of subjects defined subsequently to treatment assignment. For instance, it applies when issues of noncompliance, censoring-by-death, and surrogate outcomes within conditions have occurred. It uses two models, labeled here as the *Attrition* and *Outcome* models, to first stratify students and then estimate effects on a single stratum. Our

Attition model identifies four strata based on a student's likelihood to attrite: 1) **AA or Always Attriters:** Students who drop out regardless of condition. 2) **AC**: Students who complete if assigned to Treatment but drop out if assigned to control group. 3) **CA:** Students who only complete if assigned to Control. 4) **CC or "Never-Attriters":** Students who always complete regardless of condition; this is the stratum of interest for our work here, as it is the only group for which a treatment effect is well-defined.

True stratum membership is never observed, but must be inferred by the Attrition model using observed covariates, for which this work uses only the student's prior percent correctness labeled as acc_i. As attrition for one condition is known for each student, only the likelihood that the student would complete the opposing condition is inferred as seen in the following equations:

$$logit(Pr(completes_{i,ctrl} = 1)) = \alpha_{ctrl} + \beta_{ctrl} * acc_i$$

$$logit(Pr(completes_{i,treat} = 1)) = \alpha_{treat} + \beta_{treat} * acc_i$$

The Outcome model then observes only students placed in to the "Never-Attriter" stratum to estimate treatment effects. The equation used here utilizes the same covariates as the Attrition model with the addition of a dichotomized value of condition and a class-level variance term:

$$masteryspeed_i = \beta_{0s} + \beta_{1s} * acc_i + \beta_2 * cond_i + \sigma_i$$

The model parameters were estimated with Markov Chain Monte Carlo (MCMC) using four chains over 16000 iterations of which the first 8000 are omitted as a burn-in period allowing for convergence. The *Rhat* value shown in Table 1 reflects the degree of convergence of the Markov Chains, with the values near 1 indicating proper convergence. The results of the analysis are also seen in that table, and indicate that a TLI is not found as the effects of condition are not significant, falling within the confidence interval.

	mean	sd	0.95 CI	Rhat
Constant	1.78	0.13	(1.52,2.04)	1
Prior_Percent_Correct	-0.14	0.18	(-0.49,0.21)	1
Treatment	0.02	0.05	(-0.08,0.11)	1

	mean	sd	0.95 CI	Rhat
Constant	2.95	0.31	(2.34,3.55)	1
Prior_Percent_Correct	-1.33	0.39	(-2.09,-0.56)	1
Treatment	0.02	0.06	(-0.1,0.14)	1

Table 1: Typical Analysis: Coefficients for outcome model that predicts Mastery Speed based on Condition and Prior Accuracy, without using principal stratification (top) versus those coefficients using principal stratification (bottom).

4. SIMULATION STUDY

As no significance was found for coefficients in either case, a further comparison of principal stratification to traditional methods was conducted to verify principal stratification is beneficial in identifying such interventions when ground truth is known. The data generating model was designed to capture a tough-love intervention in which reliable difference could be found between conditions for students who would never drop out. For each simulated student, we assumed two latent/unobserved variables, intended to capture notions of *Grit* and *Ability*. There were two observed covariates, *prior percent complete*, which was a function of grit, and *prior percent correct*, which was a function of ability. The Outcome Variable (which might correspond to a post-homework quiz score) was a continuous variable that was a linear function of Ability.

A similar methodology to that described on the non-simulated dataset was then conducted. The coefficient for condition gave us a treatment effect for the never-attritor stratum. For comparison, we also conducted a Typical Analysis that estimated a treatment effect using ordinary least squares regression on all the data *without* using principal stratification and after 500 runs of the simulation, the 95% confidence interval from OLS included the average treatment effect for the never-attritors 62% of the time. In contrast, the principal stratification credible intervals were more efficient/reliable, including the true treatment effect 91% of the time.

5. CONCLUSION

This paper presented an explanation and case study application of principal stratification, to illustrate its potential as a method for analyzing randomized experiments and interventions in digital learning environments. One example from our analysis was identifying "Tough Love Interventions", but differential attrition pose a wide range of challenges to analyzing data from experiments, especially as learners gain flexibility in online environments such as Massive Open Online Courses (MOOCs). This makes the reliable analysis of experiments with variable dropout and attrition of increasing importance to the educational data mining community.

6. ACKNOWLEDGMENTS

This work is partially supported by the United States National Science Foundation Grant #DRL-1420374 to the RAND Corporation.

7. REFERENCES

[1] J. D. Angrist, G. W. Imbens, and D. B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

[2] C. E. Frangakis and D. B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

[3] J. Heckman, N. Hohmann, J. Smith, and M. Khoo. Substitution and dropout bias in social experiments: A study of an influential social experiment. *Quarterly Journal of Economics*, pages 651–694, 2000.

[4] C. Lang, N. Heffernan, K. Ostrow, and Y. Wang. The impact of incorporating student confidence items into an intelligent tutor: A randomized controlled trial. In *Proceedings of the 8th International Conference on Educational Data Mining*, 2015.

[5] A. C. Sales and J. F. Pane. Exploring causal mechanisms in a randomized effectiveness trial of the cognitive tutor. In *Proceedings of the 8th International Conference on Educational Data Mining*, 2015.