Sequential Randomization to Develop Personalized and Optimized Interventions in Massively Open Online Courses: A Case Study

Timothy NeCamp*
Department of Statistics
University of Michigan
Ann Arbor, Michigan, USA
tnecamp@umich.edu

Joshua Gardner
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, WA, USA
jpgard@cs.washington.edu

Christopher Brooks
School of Information
University of Michigan
Ann Arbor, Michigan, USA
brooksch@umich.edu

*Presenting Author

1 Introduction

In typical experiments, each subject is randomized one time; once a subject has been assigned to a particular treatment option, they remain in that same treatment throughout the rest of the trial. In contrast, sequentially randomized trials are experiments in which a subject is randomized several times to different treatment options throughout the experiment. Sequential randomization is beneficial over one-time randomization because it informs the design of personalized and optimized interventions. Firstly, by re-randomizing, subjects receive a variety of intervention sequences (e.g., treatment A then treatment B or treatment B then treatment A). These various sequences can be compared to discover the optimal intervention sequence. Secondly, instead of only being able to assess the overall effect of getting one particular treatment, sequential randomization lets researchers discover effects at smaller time-scales (e.g., treatment A does better in week 2 of the course, but treatment B does better in week 3 of the course). These discoveries inform at what time points certain interventions are most effective. Thirdly, re-randomization permits the discovery of important variables measured throughout the trial for adapting and personalizing an intervention. As opposed to only being able to discover pre-trial personalization variables (e.g., treatment A works better for women), we can also discover mid-trial personalization variables (treatment A works better for subjects with a large amount of missing data streams).

The focus of this work is a sequentially randomized trial that was designed for and implemented in a Coursera-based Massively Open Online Course (MOOC), Introduction to Applied Data Science with Python. MOOCs are an ideal platform for this form of experimentation, especially to develop new course content and evaluate learning resources. Improving MOOC content is important because MOOCs provide education to over 81 million users in over 100 different countries. The diversity of MOOC learners makes personalization critical as well. MOOC experiments fall under the umbrella of digital experiments because they can be easily implemented at scale; new course content can be provided to the entire population with the click of a button. MOOCs permit constant monitoring of users and their interactions with the content. These streams of course activity data can be used to quickly examine intervention effects on several learning outcomes. The goal of the focal study, named the Problem-based Email

Reminder Coursera Study (PERCS), was to evaluate the impact of sending different types of emails--with culturally relevant data science problems delivered using various psychological framing methods supported by learning research--on a learner's propensity to re-engage with the course.

2 Overview of Sequentially Randomized Trials

Sequential randomization refers to the randomization of an individual multiple times throughout the course of the trial. Suppose there are two types of interventions, such as learners in a MOOC receiving videos taught by a female instructor (intervention A) or a male instructor (intervention B). The simplest example of a sequentially randomized trial would be: During week 1, users have a 50% chance of receiving intervention A and a 50% chance of receiving intervention B. During week 2, users are re-randomized to another treatment. They again have a 50% chance of receiving intervention A and a 50% chance of receiving intervention B, independent of what happened in week 1. Hence, about 25% of users will have received each sequence (A, A), (A,B), (B,A), (B,B), where the parenthetical notation means (week 1 treatment, week 2 treatment).

This simple sequential randomization scheme can be modified for both practical and scientific reasons. Common modifications include using different time durations (e.g., re-randomize every month), increasing the number of time points (e.g., each person is randomized 10 times instead of twice), changing the number of treatments (e.g., A vs B vs C instead of A vs B, or A vs B in week 1 and C vs D in week 2), and altering the randomization scheme (e.g., not having 50/50 randomization probabilities each week, or, in week 2, only re-randomizing inactive users).

Sequentially randomized trials [1, 2] have become relatively common in clinical settings [3] and are becoming more prevalent in education settings [4]. Two of the most common types of sequentially randomized trials are Sequential Multiple Assignment Randomized Trials (SMARTs) [5], and Micro-randomized Trials (MRTs) [6].

2.1 Comparing to Other Trial Designs

Sequentially randomized trials are distinctive from other types of common trial designs. Unlike sequentially randomized trials, in A/B tests (often called a 2-arm randomized controlled trial in healthcare and education), each subject is randomized one time to either intervention A or intervention B, and receive that same treatment for the entire length of the study.

Sequentially randomized trials are distinctive from adaptive trial designs [8], since they do not perform any online optimization. That is, sequentially randomized trials do not use data collected throughout the trial to change randomization schemes. After the trial is complete, sequentially randomized trial data is used to develop personalized and optimized interventions.

2.2 Advantages of Sequentially Randomized Trials in Digital Experiments

Sequentially randomized trials are particularly advantageous in digital experiments because many of the barriers to performing these experiments are often eliminated in the digital setting. In non-digital settings (e.g., clinical trials in healthcare or school-level experiments in education) re-randomization and reassignment of treatment options adds a layer of complexity, making intervention delivery difficult. In many digital environments, where interventions can be delivered quickly and at scale, the burden of re-randomization and delivering a variety of treatments is minimized. Also, using sequentially randomized trials to inform mid-trial personalization requires the constant collection of data for personalization. Unlike non-digital settings, digital environments typically provide a constant source of data at minimal

cost to the experimenter (e.g., mobile phone sensor streams or webpage interactions). This information can potentially be used for personalization. Lastly, digital environments typically provide a large number of possible modifications and interventions. Sequentially randomized trials are useful as exploratory trials (as opposed to confirmatory), to help researchers narrow down intervention options and discover the best (which can then be further validated).

3 Motivation and Design of PERCS

3.1 Motivation behind PERCS

A well-known challenge in MOOCs is low completion rates. While there are many factors contributing to MOOC dropout [9], the goal of PERCS was to see if dropout could be ameliorated by using weekly email reminders to motivate learners to engage with course content. These emails sent to learners may have contained one or more of several factors intending to impact learner engagement: (i) The email could have contained a motivating data science problem to challenge the user to learn the upcoming week's content. (ii) The email might also have contained a location specific primer and a data science problem relevant to that user's specific culture (e.g., an Indian user might receive a problem about Bollywood or weather patterns in India). (iii) the email may have utilized growth mindset framing [10], a psychological framing method used to support learning. See Figure 1 for an overview of all the different kinds of emails.

Given all these different types of possible emails, the main research questions were:

1. Which sequence of emails causes the highest rate of completion?
2. Which type of data science problem email (no email, no problem, global data science problem, or culturally relevant data science problem) is most effective, on average, for bringing learners back to the course during each week?
3. Are certain data science problem emails more or less effective for active learners?
4. Does growth mindset framing improve an emails ability to bring learners back to the course?

3.2 Design of PERCS

A sequentially randomized factorial trial design was an effective method to jointly address the main research questions of PERCS. Prior to the start of weeks two, three, and four of the four-week long MOOC, learners were randomly assigned to receive one of four different email categories: an email message with a problem that reflects their geo-cultural situation based on IP address analysis (cultural problem condition), an email with a generic non-culture specific problem (global problem condition), an email with no problem (no problem condition), or no email at all (no email condition). These learners were also randomly assigned to have their email be framed with a growth mindset or without a growth mindset. The growth mindset piece crossed with the three email categories makes each week of PERCS a 2x3 factorial design with an additional control condition of no email.

The most novel aspect of PERCS is the sequential randomization. That is, a particular learner does not get randomized to one email condition and remain in that treatment condition for all 3 weeks (as is typically done in experiments). Instead, each week, a learner is re-assigned (with the same randomization probabilities) to a different email condition. The randomizations across weeks are independent. See Figure 1.

3.3 Participation in PERCS

All learners who signed up for the Introduction to Applied Data Science with Python course on Coursera between the times of April 1, 2018 to June 10, 2018 participated in PERCS. A total of 15,037 unique learners were sent 28,446 emails.

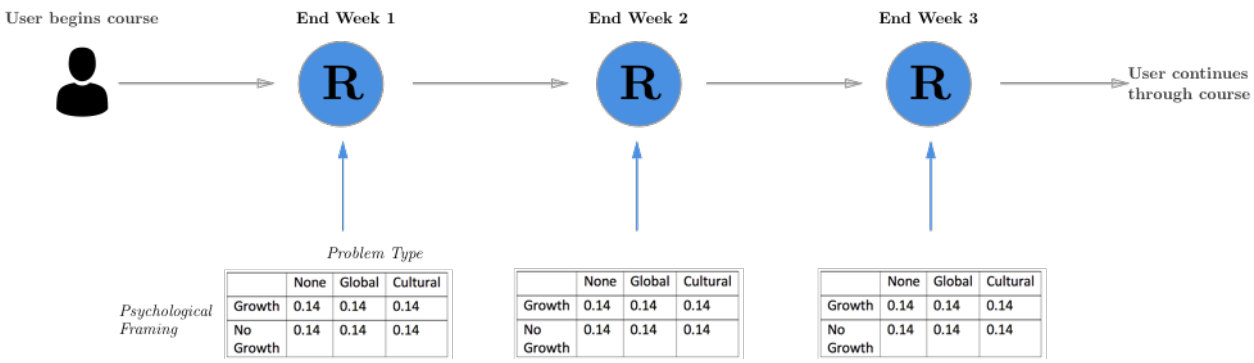

Figure 1: A schematic of PERCS's weekly randomization scheme. The tables contain the weekly randomization probabilities to different email types. There is also a 0.14 chance of receiving no email each week.

4 The Benefits of Sequentially Randomized Trials Digital Experiments

Re-randomization permits sequence optimization (research question 1). For the average learner in the course, there may be an optimal sequence of emails that will increase the chance of course completion the largest. For example, learners may prefer no email the first two weeks, and then a personalized problem email in the last week (because email reminders might annoy users if they are sent too early). If every user received the same treatment throughout the whole study, one can only compare sequences where users received the exact same email for all 3 weeks (e.g., receiving no email for 3 weeks vs receiving problem email with growth mindset for all 3 weeks). However, in the re-randomization case, users will be randomly assigned to each possible sequence of emails. We can compare these different sequences and find the optimum.

Sequentially randomized trials permit the estimation of average treatment effects at various time points (research question 2). For example, we could estimate the average effect of culturally personalized problem emails in week 3 compared to global problem emails in week 3. If learners were only randomized one time, we could not estimate the average effect of a treatment option in week 3. In this case, because week 3 treatment is the same treatment delivered in weeks 1 and 2, the week 3 effect will be confounded with effects of prior treatment. By re-randomizing individuals, we break the confounding from previous time points and can separately estimate average effects at each time point.

Re-randomization also helps experimenters understand which mid-trial variables interact with treatment to improve outcomes (research question 3). In the context of PERCS, re-randomization permits the discovery of mid-course moderators, variables measured during the course that change the efficacy of the treatment. For example, one may discover that personalized problem emails in week 3 only benefit students who were inactive in the course in week 2. In this case, week 2 course activity *moderates* the efficacy of week 3 personalized problem emails. Statistically, due to bias introduced when including post-randomization variables in analyses, one can only discover moderator variables measured prior to randomization. If all learners were only randomized once, potential moderators measured after the first sent email cannot be discovered; one can only discover moderators prior to the first email (e.g., learner's gender, learner's location). By re-randomizing, potential moderators measured before each of the randomizations (which now includes mid-course data) can be discovered. Discovering mid-course moderators promotes personalized intervention delivery by informing which emails should be sent to which users, based on their current activity in the course.

5 Analysis of Sequentially Randomized Trials

Answering questions about average treatment effects, sequence optimization, and mid-trial personalization requires slight modifications of standard statistical methods. Using these methods to analyze PERCS data, we discovered several initial findings of interest. For research question 1, we found that email sequences which send emails in the first two weeks outperform email sequences which only send emails in the third week. For research question 2, we found that average weekly effects of problem based emails were relatively weak. For research question 3, we found that in week 1, problem based emails do better for learners with prior course activity, however, in week 3, problem based emails do worse for learners with course activity in week 2 (prior course activity is a *moderator* in week 1 and 3).

6 Conclusion

Sequential randomization is a useful experimental design for digital experiments. Sequential randomization allows researchers to discover ways to personalize and optimize sequences of treatments. The benefits of sequential randomization were exemplified via PERCS, a large digital experiment aiming to diminish online course dropout in MOOCs.

References

[1] P.W. Lavori and R. Dawson. 2000. A design for testing clinical strategies: biased adaptive within-subject randomization. Journal of the Royal Statistical Society: Series A (Statistics in Society) 163, 1 29–38.

[2] Philip W Lavori and Ree Dawson. 2004. Dynamic treatment regimes: practical design considerations. Clinical Trials 1, 1 (2004), 9–20.

[3] Huitian Lei, I Nahum-Shani, K Lynch, D Oslin, and SA Murphy. 2011 . A "SMART" Design for Building Individualized Treatment Sequences. 8, 21–48.

[4] Daniel Almirall, Connie Kasari, Daniel F. McCaffrey, and Inbal Nahum-Shani. 2018. Developing Optimized Adaptive Interventions in Education. Journal of Research on Educational Effectiveness 11, 1 (2018), 27–34.

[5] Susan A Murphy. 2005. An experimental design for the development of adaptive treatment strategies. Statistics in medicine 24, 10 (2005), 1455–1481.

[6] Predrag Klasnja, Eric B Hekler, Saul Shiffman, Audrey Boruvka, Daniel Almirall, Ambuj Tewari, and Susan A Murphy. 2015. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychology 34, S (2015), 1220.

[7] William E. Pelham Jr., Gregory A. Fabiano, et al. (2016). Treatment sequencing for childhood ADHD: A multiple-randomization study of adaptive medication and behavioral interventions. *Journal of Clinical Child & Adolescent Psychology*, *45*(4), 396-415.

[8] Shein-Chung Chow and Mark Chang. 2008. Adaptive design methods in clinical trials – a review. Orphanet Journal of Rare Diseases 3, 1 (02 May 2008), 11.

[9] Jeffrey A. Greene, Christopher A. Oswald, and Jeffrey Pomerantz. 2015. Predictors of Retention and Achievement in a Massive Open Online Course. American Educational Research Journal 52, 5 (2015), 925–955.

[10] Carol S Dweck. 2008. Mindset: The new psychology of success. Random House Digital, Inc.