Bandit assignment for educational experiments: Benefits to students versus statistical power

Anna N. Rafferty[1], Huiji Ying[1], and Joseph Jay Williams[2]

[1] Computer Science Department, Carleton College, Northfield, MN 55057 USA
[2] School of Computing, Department of Information Systems & Analytics, National University of Singapore, Singapore

Abstract. Randomized experiments can lead to improvements in educational technologies, but often require many students to experience conditions associated with inferior learning outcomes. Multi-armed bandit (MAB) algorithms can address this by modifying experimental designs to direct more students to more helpful conditions. Using simulations and modeling data from previous educational experiments, we explore the statistical impact of using MABs for experiment design, focusing on the tradeoff between acquiring statistically reliable information and benefits to students. Results suggest that while MAB experiments can improve average benefits for students, at least twice as many participants are needed to attain power of 0.8 and false positives are twice as frequent as expected. Optimistic prior distributions in the MAB algorithm can mitigate the loss in power to some extent, without meaningfully reducing benefits or further increasing false positives.

Randomized controlled experiments are common in educational technologies. These experiments typically assign half of students to one version of technology components and half to another, investigating questions like whether video or text hints will be better. This approach is indifferent to benefits for learners: even if one condition is clearly ineffective, half of students experience it.

Using multi-armed bandit (MAB) algorithms in experimental designs could benefit learners by considering the utility of different versions of content. These algorithms learn a dynamically changing policy for choosing actions, balancing exploiting information already collected with exploring actions to collect additional information. Educational experimentation can be viewed as a MAB problem by treating condition assignments as action choices, with the dependent outcome serving as the reward. For example, in an experiment comparing hint types, the reward (outcome) might be 1 if the attempt after the hint was correct and 0 otherwise. Rather than assigning half of students to each condition, MABs sequentially assign students to conditions based on the rewards for previous students; more students can thus be assigned to better conditions. MABs have been used in education to discover what version of a system to give to learners [8, 9].

However, using MABs in experiment design creates a tension between benefits for students and information gained about differences between conditions [4, 6]. Because MABs assign students to conditions unevenly and change assignment proportions based on previous results, some conditions can be under-sampled and systematic measurement errors occur [4], limiting the inferences that can be drawn from results. We investigate the tradeoff between benefits to students and scientific gain, focusing on a systematic exploration of how MAB assignment impacts inferential statistics, such as the effects on power.

1 Statistical consequences of MAB-assigned conditions

We use simulations of two-condition experiments to investigate the statistical consequences of assigning conditions via Thompson sampling, a MAB algorithm with logarithmic bounds on regret growth [1] that performs well in practice [2]. We focus on Thompson sampling as a typical regret-minimizing MAB algorithm, where regret is incurred by choosing actions with lower benefit to students; we expect trends in results to hold for other regret-minimizing MAB algorithms.

1.1 Simulation methods

All simulations were repeated 500 times and across simulations, we varied:

- Method of condition assignment: MAB versus uniformly at random.
- Reward type: Binary (e.g., whether a student completes an activity) versus real-valued rewards (e.g., time to finish a problem). For MAB assignment, real-valued rewards were assumed to be normally distributed, and conjugate priors were used.
- True effect size: Zero and non-zero effect sizes were included. Non-zero effect sizes used thresholds for small, moderate, and large effects (binary: Cohen's $w = 0.1, 0.3, 0.5$; normally-distributed: Cohen's $d = 0.2, 0.5, 0.8$) [3]. Binary reward simulations fixed the average reward across conditions to 0.5, and normally-distributed reward simulations used fixed means and adjusted the variances across effect sizes.
- Number of participants (sample size): Sample sizes were $0.5m$ (lowest power), m, $2m$, and $4m$ (highest power) simulated students, where m is the sample size for 0.8 power with equally balanced conditions given false positive rate of 0.05. The same sample sizes were used when effect size was zero.
- Prior distributions (MAB): *Prior between* had a mean between the two conditions.[3] *Prior above* is optimistic about condition effectiveness, with the mean above both conditions. *Prior below* is pessimistic, placing the mean below both conditions.

1.2 Results

Conditions differ: When conditions have different benefits for students, the goal is to detect that the difference is reliable and assign more students to the better condition. MAB assignment without an optimistic or pessimistic prior (*prior between*) decreased power from an expected 0.80 to 0.54 for binary rewards and 0.51 for normally-distributed rewards (Figure 1a). Doubling the sample size raised power to 0.78 and 0.69, but increasing sample size is less effective over time as evidence for the superiority of one condition leads to assigning few students to the alternative (Figure 1b). Type S errors [5] were rare ($< 0.15\%$), and no

[3] For zero effect size, the mean was equal to the mean of each condition.

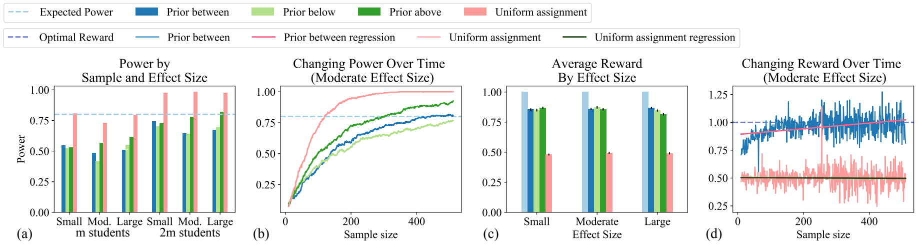

Fig. 1. Power and rewards by assignment type for normally-distributed rewards; binary rewards showed a similar pattern. Error bars represent one SE.

difference by assignment type was detected. An optimistic prior (*prior above*) led to higher power and more accurate effect sizes due to more equal sampling across conditions initially that provided better evidence for statistical inferences.

MAB assignment obtained greater rewards than uniform: longer experiments are offset by a larger proportion of students in the better condition. Expected reward per student approached the mean of the more effective condition (Figure 1d), and was only modestly decreased with more optimistic priors (Figure 1c). *Conditions do not differ:* MAB assignment increased false positives from an expected rate of 5% to 9.7% of simulations using MAB assignment. Thus, analyzing data collected via MAB assignment and using typical statistical tests may lead to higher false positives than expected based on setting α (the expected Type I error rate). Type I error rate was slightly higher for normally-distributed rewards than for binary, primarily due to insufficient exploration with small variances.

2 MAB-assignment in educational experiments

To understand how effects found in simulation might translate to real educational experiments, we analyzed MAB assignment in the context of ten significant/marginal results from twenty-two randomized experiments [7]. These experiments included both binary outcomes (whether a student *completed* an assignment by solving three consecutive problems correctly) and real-valued outcomes (the *problem count* for completion and logarithm of the problem count).

Parameter simulations used measured means (and variances) from the experiments to generate samples, allowing unlimited students but assuming rewards are accurately modeled by a given distribution. *Outcome* simulations directly sampled a student in the chosen condition from the data set (without replacement) and using their measured outcome for the reward. *Parameter* simulations had sample sizes equal to the original experiments, while *outcome* simulations terminated when no students remained in a chosen condition.

2.1 Results

Parameter simulations: As shown in Figure 2a, MAB assignment resulted in small improvements on average reward per student across all outcome measures ($t(9989) = 5.10$, $p < .0001$; median effect size $d = 0.70$). Figure 2b shows that

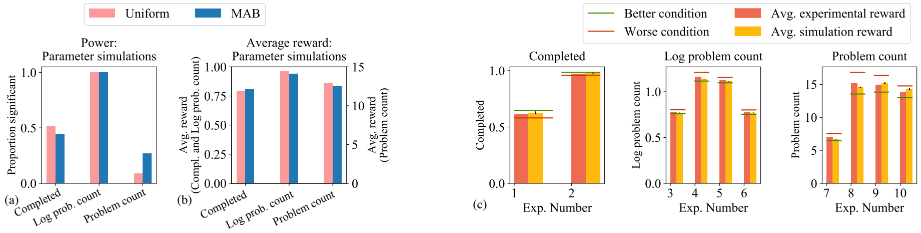

Fig. 2. Results based on educational experiments. (a-b) Power (a) and reward or cost per step (b) averaged across the *parameter* simulations. For rewards, higher is better for *completed*; lower is better for the other measures. (c) *Outcome* simulation rewards. "Better" and "worse" are observed experimental rewards for each condition.

MAB assignment decreased power for the *completed* measure. Counterintuitively, MAB assignment increased power for *problem count* by oversampling highly variable conditions, leading to more confident estimates of effectiveness. Average Type S error rates were small (uniform assignment: 0.3%; MAB: 0.4%).
Outcome simulations: MAB assignment achieved small improvements on average reward for eight out of ten experiments (Figure 2c); these rewards were almost as good as the better condition, which is the maximum possible.

For the nine experiments that had a significant effect, 65% of simulations found a significant difference between conditions, which compares favorably to the 0.55 power for uniform assignment in the *parameter* simulations.

3 Discussion

Experiments using uniform random assignment can identify more effective educational strategies, but there are ethical concerns about their impact on students. Our simulations demonstrate MABs can assign a greater proportion of students to the better condition, but can also lead to higher Type I error rates than expected and the need for doubled sample sizes to achieve expected power when results are analyzed using traditional inferential statistics. These results were generally confirmed in our experimental modeling, but were less extreme: power was increased in some cases due to differences in variability across conditions, and relatively small differences between conditions in the original experiments meant there was limited potential for MAB assignment to increase rewards.

There are several limitations to this work. First, we focused only on experiments with two conditions. Second, we focused on a regret-minimizing algorithm. While exploring the statistical consequences of other objectives is important future work, our goal is to illustrate how standard MAB algorithms impact conclusions for researchers who may be excited by the potential benefits to students. We hope this will lead to careful consideration of how to achieve *both* research and pedagogical aims, and that our focus on statistical significance shows that MAB assignment can lead to erroneous generalizations in addition to measurement error. MAB assignment is one way to mitigate costs to students as educational experiments become more ubiquitous, but caution must be used when interpreting results and applying standard statistical methods.

References

1. Agrawal, S., Goyal, N.: Analysis of thompson sampling for the multi-armed bandit problem. In: Mannor, S., Srebro, N., Williamson, R.C. (eds.) Proceedings of the 25th Annual Conference on Learning Theory. vol. 23, pp. 39.1–39.26. PMLR, Edinburgh, Scotland (2012)
2. Chapelle, O., Li, L.: An empirical evaluation of thompson sampling. In: Advances in neural information processing systems. pp. 2249–2257 (2011)
3. Cohen, J.: Statistical power analysis for the behavioral sciences. Routledge, 2 edn. (1988)
4. Erraqabi, A., Lazaric, A., Valko, M., Brunskill, E., Liu, Y.E.: Trading off rewards and errors in multi-armed bandits. In: International Conference on Artificial Intelligence and Statistics (2017)
5. Gelman, A., Carlin, J.: Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. Perspectives on Psychological Science 9(6), 641–651 (2014)
6. Liu, Y.E., Mandel, T., Brunskill, E., Popovic, Z.: Trading off scientific knowledge and user learning with multi-armed bandits. In: Educational Data Mining 2014 (2014)
7. Selent, D., Patikorn, T., Heffernan, N.: Assistments dataset from multiple randomized controlled experiments. In: Proceedings of the Third (2016) ACM Conference on Learning@ Scale. pp. 181–184. ACM (2016)
8. Whitehill, J., Seltzer, M.: A crowdsourcing approach to collecting tutorial videos–Toward personalized learning-at-scale. In: Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale. pp. 157–160. ACM (2017)
9. Williams, J.J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K.Z., Lasecki, W.S., Heffernan, N.: Axis: Generating explanations at scale with learnersourcing and machine learning. In: Proceedings of the Third (2016) ACM Conference on Learning@ Scale. pp. 379–388. ACM (2016)