Adaptive Experiments for Customer Segmentation*

Kevin Laughren and Henry Schneider

November 14, 2023

DRAFT: NOT FOR CITATION OR DISTRIBUTION

1 Introduction

Government agencies, businesses, and other organizations must conduct experiments, formally or informally, if they are to optimize their policies. Broadly speaking, the focus of this optimization can be in-sample or out-of-sample optimization, or some combination of the two. In the language of machine learning, an in-sample focus is known as “exploit” – for example, allocating a fixed budget for a job training program to maximize employment; while an out-of-sample focus is known as “explore” – for example, identifying the most profitable long-run advertising strategy for a company’s products (Berger-Tal *et al.*, 2014).

Identifying an optimal policy among many options can often be framed as a multi-armed bandit. The standard multi-arm bandit (Gittins *et al.*, 2011) involves dynamically allocating a fixed budget across multiple lotteries with unknown payouts and sequential draws – proverbially speaking, pulling the arms of a set of slot machines sequentially to learn about their payouts, possibly shifting future pulls to machines showing more favorable returns in earlier pulls.

Adaptive sampling algorithms are a powerful approach to solving multi-arm bandit policy problems. Our paper introduces a novel adaptive sampling approach for higher-dimensional multi-arm bandits, which we believe has broad real-world applicability. We then demonstrate the approach through an experiment with a startup company and begin to explore its properties using the experimental data. In the remainder of the Introduction, we provide some background on adaptive sampling, and then overview our approach, experiment, and results.

Adaptive sampling is designed for experimental setting in which data are collected sequentially

*We thank David Freeman, Steve Lehrer, and Stephanie Wang for helpful feedback.

such that earlier data from the experiment can inform later treatment assignment. In contrast to randomized controlled trials (RCT), adaptive sampling embraces non-random treatment assignment. For example, if interim data suggests that women better respond to treatment than men and the objective is in-sample optimization, ongoing sampling could be diverted to women. Adaptive sampling is well known to some disciplines – e.g., clinical trial design (Chow, 2014; FDA, 2019) – but is only starting to receive meaningful attention in the social sciences. This lag may be due in part to the challenges of incorporating causal inference into non-random experimental design, but recent work by Athey (2015), Athey and Imbens (2016), Wager and Athey (2018) and others have made significant progress on this capability.

When the objective is exploit, a class of adaptive sampling algorithms called "greedy" (Edmonds, 1971) often works well. Greedy algorithms sample exclusively from the single action the accumulating interim data suggests may work best, at the expense of learning about the effectiveness of other actions and possibly missing better performing actions.

When the objective is explore, more versatile adaptive sampling approaches that balance the exploit-explore trade-off are available. Thompson sampling (Thompson, 1933) is among the most popular of these balanced approaches, sampling both from actions that have performed well to that point (exploit) and from actions with significant uncertainty over performance (explore). Caria *et al.* (forthcoming) is an important extension of Thompson sampling that includes a tempering parameter $\gamma \in [0, 1]$ that assigns a fraction of sampling to be fully randomized, allowing experimenters to tune the amount of exploration. This "Tempered Thompson Algorithm" (TTA) is a hybrid of RCT and Thompson sampling and expands the causal inference capabilities of Thompson sampling via this randomization component.

Introducing some basic notation, a generic sampling algorithm assigns every subject $x \in \mathcal{X}$ to a treatment $a \in \mathcal{A}$, where x is a covariate vector of subject attributes and a is a vector of treatment characteristics. The action space is $\mathcal{X} \times \mathcal{A}$. For example, a marketing strategy might assign every x to an a, where $\mathcal{X}$ represents consumer attributes in a marketing database and $\mathcal{A}$ represents possible promotions and ad copy.

A limitation of existing adaptive sampling methods, including Thompson sampling, is the need for large amounts of data, often prohibitively so, to learn about large action spaces. That is, a high-dimensional $\mathcal{X} \times \mathcal{A}$ can limit the use of existing adaptive sampling methods. A standard solution to this dimensionality problem is to restrict the covariate-treatment space upfront based on the experimenter's prior beliefs. However, this workaround can cause actions that are productive in unanticipated ways to be missed.

The novel adaptive sampling algorithm that we introduce extends Thompson sampling (and TTA) to larger action spaces, allowing a more complete set of covariates and treatments to be considered. Our approach partitions this larger action space into a coarse grouping of actions with similar response rates using classification trees (Breiman *et al.*, 1984) based on the accumulating interim data. Instead of sampling from all actions in the action space, as with traditional Thompson sampling, we sample from groups of actions created by this partition, thereby addressing the dimensionality problem. We use classification trees as opposed to ensemble methods like random forest because they create a single partition of the action space from which to sample.[1] We call our approach "Partitioned Thompson Algorithm" (PTA).

A summary of PTA is as follows, with details in Sections 3, 4, and 5:

(1) Define the experimental design and machine learning parameters, including maximum sample size, maximum number of rounds over which to sample, outcome measure of interest, and treatment variations.

(2) Perform an initial round of sampling using fully random treatment assignment.

(3) Conduct recursive partitioning on the data collected to this point using a classification tree (Breiman *et al.*, 1984).

(4) Consider each leaf of the classification tree as an action and perform Thompson sampling on these actions. Append the new data from this Thompson sampling to the existing sample.

(5) Repeat steps (3) and (4), where each repetition is a round, until the maximum sample size, maximum number of rounds, or an early stopping condition (described later) is reached.

Step (3) is the innovation of PTA: Sampling from leaves of a classification tree rather than from every individual action makes it feasible to apply adaptive sampling to higher-dimensional action spaces. Our focus is Thompson sampling because of the popularity and intuitive nature of this method (Russo *et al.*, 2018). However, our classification tree-partitioning approach can extend other adaptive sampling methods to higher-dimensional settings as well, including Upper Confidence Bound (UCB) algorithms and epsilon-Greedy algorithms. Doing so simply requires replacing Thompson sampling in Step (4) with another of these algorithms.

[1] Classification trees are also more "explainable" and hence more accessible to real-world practitioners. Explainable machine learning models are less powerful than state-of-the art models like deep learning but excel at being "explainable, interpretable, and transparent" (Roscher *et al.*, 2020; Belle and Papantonis, 2021).

After introducing PTA, we explore its properties using data we collected from an marketing field experiment in collaboration with a direct-to-consumer startup company. Startups often seek to make immediate sales to stay afloat (exploit) but also to learn the optimal customer base to whom to apply future marketing resources (explore). This exploit-explore dichotomy, combined with the potential for very large covariate and treatment spaces, makes startup marketing a natural testing ground for our adaptive sampling approach. In this marketing context, each leaf of the classification tree represents a distinct consumer segment and the classification tree represents a consumer segmentation map.

Our analysis represents an initial and somewhat informal exploration of the properties of PTA, and we speculate about how PTA may apply beyond our setting. Our experiment varies prices and ad copy across online ad impressions presented to Facebook and Instagram users. We record clicks, landing page views, email registrations, and sales as conversion measures. We then compare the performance of PTA to three counterfactual sampling approaches constructed from the experimental data: an RCT, a Greedy algorithm, and an Oracle, which we describe briefly below and in more detail in Section 6.3.

To create the RCT counterfactual, we sample with replacement from the already-collected experimental data such that each action (i.e., covariate-treatment combination) has equal probability of being sampled. To create the Greedy algorithm counterfactual, we choose in each round $t \geq 2$ the single action with the highest expected payoff based on the experimental data sampled through $t-1$. To construct the Oracle counterfactual, we choose the single action with the highest expected payoff based on the complete experimental data. We then calculate the differences in in-sample performance between PTA, RCT, and Greedy, and consider their trade-offs, with the Oracle serving as a best-case benchmark. Our measure of in-sample performance is regret, a common machine learning performance measure, while our primary measure of out-of-sample performance is a variation of root mean square error, weighted by the financial consequence of each error, which we call "money loss," defined in Section 6.3.

We conduced our experiment in collaboration with Two Tails Story Co., which produces artistic alphabet and counting books for children. The purpose of the experiment was to demonstrate PTA and explore its properties in a real-world application. In the experiment, we presented online advertisement impressions to Facebook and Instagram users over three rounds of PTA sampling. In total there were 414,470 impressions, 7,382 clicks, 3,071 company website visits (not all clicks generate a website visits), and 61 Two Tails email list sign-ups. We elaborate on the experimental procedures and outcomes in Section 7.

Through the experiment and counterfactuals, we investigate the exploit and explore performance of PTA in relation to the full-exploit approach of Greedy and full-explore approach of RCT. By considering a modest-sized action space, we are also able to investigate whether our partitioned action-space approach causes any performance degradation relative to traditional Thompson sampling, which allows sampling from every action in the action space. While the primary reason for using a partitioned action space is to extend adaptive sampling to higher-dimensional action spaces, we still wish to learn about potential performance loss that comes with this flexibility.

We calculate the performance of PTA in relation to the counterfactual sampling algorithms (RCT, Greedy, Oracle) for a range of parameter vectors. For each algorithm and parameter vector, we resample with replacement from the original experimental data 500 times to obtain reliable averages. As mentioned, in some settings, in-sample performance is the priority, which occurs when one wishes to optimize the benefits of fixed resource during the trial itself; while in other settings, out-of-sample performance is the priority, to focus on exploring a broader set of actions during the trial to increase the chance of identifying the best long-run policy.

First we consider in-sample performance. We find that Greedy outperforms PTA and RCT when there are fewer rounds of sampling or larger sample sizes. This is as expected because Greedy is designed to exploit in-sample data at the expense of exploration and more data increases the ability of the algorithm to accurately focus the ongoing sampling. Perhaps surprisingly, PTA outperforms Greedy when there are more rounds of sampling or smaller sample sizes. In these sparser-data settings, we find the exploration capabilities of PTA can identify better performing actions, whereas Greedy can get stuck on a weaker performing action in cases where an action over-performs its population average in earlier sampling rounds.

Next we consider out-of-sample performance. While RCT has the lowest RMSE (and Greedy the highest), RMSE is an inadequate measure when one cares asymmetrically about accurately predicting a particular response type (e.g., a positive response), as is often the case in real-world settings. Instead we focus on a money-loss error measure that assigns a particular monetary loss to each type of prediction error. For example, in our setting, the money-loss from failing to predict a positive response (e.g., not showing an ad impression that would have generated an email sign-up) might be hundreds of times larger than the money-loss from failing to predict a non-response (e.g., showing an ad impression than does not generate a response). Because PTA is designed to sample from higher-performing actions, the money-loss error under PTA seems to be much smaller compared to RCT, which has better prediction accuracy in general but is less helpful in practice when a primary aim is to identify the most productive actions, rather than obtaining the most accurate estimates of all actions, productive or not.

Finally, we investigate whether performance suffers due to the partitioned nature of the sampling space of PTA, which can be far coarser than the sampling space of traditional Thompson sampling. In fact, PTA generates lower out-of-sample money loss compared to traditional Thompson sampling, suggesting our partitioning approach may address not only the dimensionality challenges of adaptive sampling with larger action spaces but may also provide more accurate out-of-sample predictions. We report the full results in Section 8.

References

Athey, S. (2015). Machine learning and causal inference for policy evaluation. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 5–6.

— and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, **113** (27), 7353–7360.

Belle, V. and Papantonis, I. (2021). Principles and practice of explainable machine learning. *Frontiers in big Data*, p. 39.

Berger-Tal, O., Nathan, J., Meron, E. and Saltz, D. (2014). The exploration-exploitation dilemma: a multidisciplinary framework. *PloS one*, **9** (4), e95693.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classification and regression trees. *Cole Statistics/Probability Series*.

Caria, S., Gordon, G., Kasy, M., Quinn, S., Shami, S. and Teytelboym, A. (forthcoming). An adaptive targeted field experiment: Job search assistance for refugees in Jordan. *Journal of the European Economic Association*.

Chow, S.-C. (2014). Adaptive clinical trial design. *Annual review of medicine*, **65**, 405–415.

Edmonds, J. (1971). Matroids and the greedy algorithm. *Mathematical programming*, **1**, 127–136.

FDA (2019). Adaptive designs for clinical trials of drugs and biologics: Guidance for industry. *Draft Guidance*, p. 33.

Gittins, J., Glazebrook, K. and Weber, R. (2011). *Multi-armed bandit allocation indices*. John Wiley & Sons.

Roscher, R., Bohn, B., Duarte, M. F. and Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *Ieee Access*, **8**, 42200–42216.

Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., Wen, Z. *et al.* (2018). A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, **11** (1), 1–96.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, **25** (3-4), 285–294.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, **113** (523), 1228–1242.