Optimal Testing in the Experiment-rich Regime

Sven Schmit
Stitch Fix, Inc

Virag Shah
Stanford University

Ramesh Johari
Stanford University

Abstract

Motivated by the widespread adoption of large-scale A/B testing in industry, we propose a new experimentation framework for the setting where potential experiments are abundant (i.e., many hypotheses are available to test), and observations are costly; we refer to this as the experiment-rich regime. Such scenarios require the experimenter to internalize the opportunity cost of assigning a sample to a particular experiment. We fully characterize the optimal policy and give an algorithm to compute it. Furthermore, we develop a simple heuristic that also provides intuition for the optimal policy. We use simulations based on real data to compare both the optimal algorithm and the heuristic to other natural alternative experimental design frameworks. In particular, we discuss the paradox of power: high-powered "classical" tests can lead to highly inefficient sampling in the experiment-rich regime.

1 INTRODUCTION

In modern A/B testing (e.g., for web applications), it is not uncommon to find organizations that run hundreds or even thousands of experiments at a time (Kaufman et al., 2017; Tang et al., 2010a; Kohavi et al., 2009; Bakshy et al., 2014). Increased computational power and the ubiquity of software have made it easier to generate hypotheses and deploy experiments. Organizations typically continuously experiment using A/B testing. In particular, the space of potential experiments of interest (i.e., hypotheses being tested) is vast; e.g., testing the size, shape, font, etc., of page elements, testing different feature designs and user flows, testing different messages, etc. Artificial intelligence techniques are being deployed to help automate the design of such tests, further increasing the pace at which new experiments are designed (e.g., Sensei, Adobe's A/B testing product, is being used in Adobe Target).[1]

This abundance of potential experiments has led to an interesting phenomenon: despite the large numbers of visitors arriving per day at most online web applications, organizations need to constantly consider the most efficient way to allocate these visitors to experiments. For many experiments, baseline rates may be small (e.g., a low conversion rate), or more generally effect sizes may be quite small even relative to large sample sizes. For example, large organizations may be seeking relative changes in a conversion rate of 0.5% or less, potentially necessitating millions of users allocated to a single experiment to discover a true effect. (See Tang et al. (2010a,b); Deng et al. (2013) and Azevedo et al. (2018), where these issues are discussed extensively.) Since organizations have a plethora of hypotheses of interest to test, there is a significant *opportunity cost*: they must constantly trade off allocation of a visitor to a current experiment against the potential allocation of this visitor to a new experiment.

In this paper, we study a benchmark model with the feature that experiments are abundant relative to the arrival rate of data; we refer to this as the *experiment-rich regime*. A key feature of our analysis is the impact of the opportunity cost described above: whereas much of optimal experiment design takes place in the setting of a *single* experiment, the experiment-rich regime fundamentally requires us to trade off the potential for discoveries across *multiple experiments*. Our main contribution is a complete description of an optimal discovery algorithm for our setting; the development of an effective heuristic; and an extensive data-driven simulation analysis of its performance against more classical techniques commonly applied in industrial A/B testing.

We present our model in Section 2. The formal setting we consider mimics the setting of most industrial A/B

Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

[1] `https://www.adobe.com/marketing-cloud/target.html`

testing contexts. The experimenter receives a stream of observational units and can assign them to an infinite number of possible experiments, or *alternatives*, of varying quality (*effect size*). We consider a Bayesian setting where there is a prior over the effect size of each alternative, which is natural in a setting with an infinite number of experiments.

We focus on the objective of finding an alternative that is at least as good as a given threshold s as fast as possible. In particular, we call an alternative a *discovery* if the posterior probability that the effect is greater than s is at least $1 - \alpha$, and the goal is to minimize the expected time per discovery. This is a natural criterion: good performance requires finding an alternative that is actually delivering practically significant effects (as measured by s). Adjusting s and α allows the experimenter to trade off the quality and quantity of discoveries made. Note that under this criterion any optimal policy is naturally incentivized to find the "best" experiments, because the discovery criterion is easiest to be met for those alternatives.

In Section 3 we present an optimal policy for allocation of observations to experiments. Since observations arrive sequentially, the problem can be equivalently formulated as minimizing the cumulative number of observations until until a discovery is made. We characterize a dynamic programming approximation of this problem, and show this method converges to the optimal policy in an appropriate sense. We also develop a simple heuristic that approximates and provides insight into the optimal policy.

In Section 4 we use data on baseball players' batting averages as input data for a simulation analysis of our approach. Our simulations demonstrate that our approach delivers fast discovery while controlling the rate of false discoveries; and that our heuristic approximates the optimal policy well. We also use the simulation setup to compare our method to "classical" techniques for discovery in experiments (e.g., hypothesis testing). This comparison reveals the ways in which classical methods can be inefficient in the experiment-rich regime. In particular, there is a *paradox of power*: efficient discovery can often lead to low power in a classical sense, and conversely high-powered classical tests can be highly inefficient in maximizing the discovery rate.

Due to space constraints, all proofs are given in the appendix.

1.1 Related work

The literature on sequential testing goes back many decades. Originally, Wald and Wolfowitz (1948) propose an optimal test, the sequential probability ratio test, or SPRT for short, for testing a simple hypothesis. See also Chernoff (1959). For a more thorough overview of sequential testing, we refer the interested reader to Siegmund (2013), Wetherill and Glazebrook (1986), Shiryayev (1978) and Lai (1997). None of these approaches consider the opportunity cost associated with having multiple experiments.

Recently, there has been an increased interest in sequential testing due to the rise in popularity of A/B testing (Deng et al., 2017; Kaufman et al., 2017; Kharitonov et al., 2015; Goldberg and Johndrow, 2017), and the ubiquity of peeking (Johari et al., 2017; Balsubramani and Ramdas, 2016; Deng et al., 2016). A recent paper by Azevedo et al. (2018) discusses how the tails of the effect distribution affect the assignment strategy of observations to experiments and complements this work nicely.

There is also a strong connection to the multi-armed bandit literature (Gittins et al., 2011; Bubeck and Cesa-Bianchi, 2012), especially the pure exploration problem (Bubeck et al., 2009; Jamieson et al., 2014; Russo, 2016), where the goal is to find the best arm. The case with infinitely many arms is studied by Carpentier and Valko (2015); Chaudhuri and Kalyanakrishnan (2017); Aziz et al. (2018). Locatelli et al. (2016) studies the setting of finding the set of arms (out of finitely many) above a given threshold in a fixed time horizon. Ramdas et al. (2017) consider a setting that combines the multi-armed bandit problems with sequential tests.

Methods to control of the false discovery rate in the sequential hypothesis setting are discussed by Foster and Stine (2007), Javanmard and Montanari (2016) and Ramdas et al. (2017). The connection between with multi-armed bandits is made by Yang et al. (2017). However, the Bayesian framework we propose does not require multiple testing corrections. These works take the results of the tests as given and provide methods to adjust for multiple comparisons in a sequential manner, rather than helping the experimenter to decide what experiments to run.

The heavy-coin problem (Chandrasekaran and Karp, 2012; Malloy et al., 2012; Jamieson et al., 2016; Lai et al., 2011) is another closely related research area. Here, a fraction of coins in a bag is considered heavy, while most are light. The goal is to find a heavy coin as quickly as possible. These approaches rely on likelihood ratios, as there are only two alternatives, and there is a connection to the CUSUM procedure (Page, 1954). Recently, these ideas have been adapted to crowdfunding applications Jain and Jamieson (2018).

Optimal stopping rules have been studied extensively, often under the umbrella of the secretary problem (Freeman, 1983; Samuels, 1991). There, the focus is on

comparing across alternatives.

2 MODEL AND OBJECTIVE

In this section we describe the model we study and the objective of the experimenter.

Experiments. We consider a model with an infinite number of experiments, or alternatives, indexed by $i \in \{1, 2, \ldots\}$. Each experiment is associated with a parameter $\mu_i \in M \subset \mathbb{R}$ drawn independently from a common (known) prior π that completely characterizes the distribution of outcomes corresponding to that experiment. Throughout our analysis, the experimenter is interested in experiments with higher values of μ_i.

Actions and outcomes. At times $t = 1, 2, \ldots$, the experimenter selects an alternative I_t and observes an independent outcome X_t drawn from a distribution $F(\mu_{I_t})$. Note, in particular, that opportunities for observations arrive in a sequential, streaming fashion. We also assume that observations are independent across experiments.

We assume that $F(\mu_i)$ is described by a single parameter natural exponential family, i.e. the density for an observation can be written as:

$$f_X(x \mid \mu) = h(x) \exp\left(\mu S(x) - A(\mu)\right), \qquad (1)$$

for known functions S, h, and A. Let $S_t^i = \sum_{t:I_k=i} S(X_t)$ be the canonical sufficient statistic for experiment i at time t. Note that in particular, this model includes the conjugate normal model with known variance and the beta-binomial model for binary outcomes.

Policies. Let $\mathcal{F}_t = \sigma\{X_1, I_1, X_2, I_2, \ldots, X_t, I_t\}$ denote the σ-field generated. A *policy* is a mapping from $\mathcal{F}_t$ to experiments.

Discoveries. The experimenter is interested in finding *discoveries*, defined as follows.

Definition 1 (Discovery)**.** *We say that alternative i is a* discovery *at time t, given s and α, if*

$$\mathbb{P}(\mu_i < s \mid \mathcal{F}_t) < \alpha. \qquad (2)$$

Here s and α are parameters that capture the experimenter's preferences, i.e., the level of aggressiveness and risk that she is willing to tolerate. (Note that this is more stringent than the related false discovery rate guarantees (Benjamini and Hochberg, 2007).)

We assume that the prior satisfies $\mathbb{P}(\mu_i < s \mid \emptyset) \in (\alpha, 1)$ to avoid the trivial scenarios that all or none of the alternatives is a discovery before trials begin.

Objective: Minimize time to discovery. As motivated in the introduction, informally the objective is to find discoveries as fast as possible. We formalize this as follows: The goal of the experimenter is to design a policy (i.e., an algorithm to match observations to experiments) such that the number of observations until the first discovery is minimized.

In particular, define the time to first discovery τ as:

$$\tau = \min\{t : \exists i^* \text{ s.t. } \mathbb{P}(\mu_{i^*} < s \mid \mathcal{F}_t) < \alpha\}. \qquad (3)$$

Then the goal is to *minimize* $\mathbb{E}[\tau]$ over all policies. Given this goal, the only decision the experimenter needs to make at each point in time till the first success is whether to reject the current experiment or to continue with it.

Discussion. We conclude with three remarks regarding our model.

(1) *Posterior validity.* Note that at the (random) stopping time τ, the posterior is computed based on the potentially adaptive matching policy used by the experimenter. The following lemma shows that when the experimenter computes the posterior and decides to stop the experiment at time t when the condition $\mathbb{P}(\mu_{i^*} < s \mid \mathcal{F}_t)$ is met, the decision to stop does not invalidate the discovery.

Lemma 1. *The posterior for the discovered experiment i^* at time τ satisfies*

$$\mathbb{P}(\mu_{i^*} < s \mid \mathcal{F}_\tau) < \alpha \qquad (4)$$

almost surely.

(2) *Fixed cost per experiment.* In some scenarios, starting a new experiment has a cost; e.g., there may be a cost to implementing a new variant, or results may need to be analyzed on a per experiment basis. We can incorporate such a cost in the objective, and our results and approach generalize accordingly. Formally, let c be the cost of starting a new experiment, and let $m_t = |\{i : \exists t' \leq t : I_{t'} = i\}|$ be the cumulative number of matched experiments up to time t. We can include the per experiment cost by considering instead the problem of *minimizing* $\mathbb{E}[\tau + cm_\tau]$.

(3) *Fast rejection of experiments.* The optimal policy, described in the next section, tends to reject many experiments very quickly, sometimes even after a single sample. That is not surprising, given that related CUSUM procedures (Page, 1954) share this characteristic. This happens since we are studying an asymptotic regime where the number of experiments are infinite. As we will see, the limiting case of the experiment-rich regime provides a valuable lens that challenges common perceptions, such as the importance of power.

In practice, an experimenter may wish to avoid such an aggressive rejection since starting experiments may

be costly. Including a cost term for each experiment in our setup can help alleviate the problem, which is easy to do as described above.

3 OPTIMAL POLICY

In this section, we characterize the structure of the optimal policy, show that it can be approximated arbitrarily well by considering a truncated problem, and give an algorithm to compute the optimal policy of the truncated problem. Finally, we present a simple heuristic that approximates the optimal policy remarkably well.

3.1 Sequential policies

We start with a key structural result that simplifies the search for an optimal policy. The following lemma shows that we can focus on policies that only consider experiments *sequentially*, in the sense that once a new experiment is being allocated observations, no previous experiment will ever again receive observations.

Lemma 2. *There exists an optimal policy such that $I_{t+1} \geq I_t$ for all t almost surely.*

This result hinges on three aspects of our model: experiments are independent of each other, with identically distributed effects μ_i; there are an infinite number of experiments available; and observations arrive in an infinite stream. As a consequence, all experiments are *a priori* equally viable, and *a posteriori* once the experimenter has determined to stop allocating observations to an experiment, she need never consider it again.

Note in particular that this lemma also reveals that any optimal policy for the first discovery also straightforwardly minimizes the expected time until the k'th discovery, for any k.

3.2 Reformulating the optimization problem

Based on Lemma 2, we can reformulate and simplify the optimization problem faced by the experimenter as a sequential decision problem, where the only choice is whether or not to continue testing the *current* experiment.

We abuse notation to describe this new perspective. Let μ denote the effect size of the current experiment. In particular, let X_n be the n'th observation; let $\mathcal{F}_n$ be the σ-field generated by observations of the current experiment $(X_1, \ldots, X_n)$. Let $S_n = \sum_{k=1}^{n} S(X_k)$ denote the canonical sufficient statistic at state n. The *state* of the sequential decision problem is (n, S_n), the number of observations and the sufficient statistic of the current experiment.

If (n, S_n) has the property that $\mathbb{P}(\mu < s|S_n) < \alpha$, then a discovery has been found and so the process stops. The following lemma shows that this discovery criterion induces an *acceptance region* on the sufficient statistic S_n, i.e., a sequence of thresholds a_n such the current experiment is a discovery when $S_n \geq a_n$.

Lemma 3. *There exists a sequence $\{a_n\}_{n=1}^{\infty}$ such that $\mathbb{P}(\mu < s \mid S_n) < \alpha$ if and only if $S_n > a_n$.*

If $S_n < a_n$, then the experimenter can make one of two decisions:

1. *Continue* (i.e., collect one additional observation on the current experiment); or
2. *Reject* (i.e., quit the current experiment and collect the first observation of a new experiment).

If *Continue* is chosen, the state updates to $(n+1, S_{n+1})$. If *Reject* is chosen, the state changes to $(1, S_1)$ (where S_1 is an independent draw of the sufficient statistic after the first observation); and in either case, the process continues.

The goal of the experimenter is to minimize the expected time until the observation process stops, i.e., until a discovery is found. Let $V(n, S_n)$ be this minimum, starting from state (n, S_n). Then the Bellman equation for this process is as follows:

$$V(n, S_n) = 0, \quad S_n \geq a_n; \tag{5}$$

$$V(n, S_n) = 1 + \min\left\{\mathbb{E}[V(n+1, S_{n+1})|S_n], \mathbb{E}[V(1, S_1)]\right\}, S_n \leq a_n \quad n \geq 1. \tag{6}$$

The first line corresponds to the case where S_n is in the acceptance region, i.e., the process stops. In the second line, we consider two possibilities: continuing incurs a unit cost for the current observation, plus the expected cost from the state $(n+1, S_{n+1})$; rejecting resets the state with no cost incurred. The optimal choice is found by minimizing between these alternatives. The expected number of samples T^* till a discovery satisfies $T^* = 1 + \mathbb{E}[V(1, S_1)]$.

3.3 Characterizing the optimal policy

The following theorem shows that an optimal policy for the dynamic programming problem (5)-(6) can be expressed using a sequence of rejection thresholds on the sufficient statistic. That is, for each n there is an r_n such that it is optimal to *Continue* if $S_n \geq r_n$, and to *Reject* if $S_n < r_n$.

Theorem 4. *There exists an optimal policy for* (5)-(6) *described by a sequence of rejection thresholds $\{r_n\}_{n=1}^{\infty}$ such that, after n observations,* Reject *is declared if $S_n < r_n$,* Continue *is declared if $r_n \leq S_n \leq a_n$, and the process stops with a discovery if $S_n > a_n$.*

The remainder of the section is devoted to computing the optimal sequence of rejection thresholds.

3.4 Approximating the optimal policy via truncation

In order to compute an optimal policy, we consider a *truncated* problem. This problem is identical in every respect to the problem in Section 3.2, except that we consider only policies that must choose *Reject* after k observations. We refer to this as the *k-truncated* problem.

Let $V_k(n, S_n)$ denote the minimum expected cumulative time to discovery for the k-truncated problem, starting from state (n, S_n). The Bellman equation is nearly identical to (5)-(6), except that now $V_k(k, S_k) = 1 + \mathbb{E}[V_k(1, S_1)], S_k \leq a_k$, and we add the additional constraint that $n < k$ to (6). We have the following proposition.

Theorem 5. *There exists an optimal policy for the k-truncated problem described by a sequence of rejection thresholds $\{r_n^k\}_{n=1}^{\infty}$ such that, after n observations,* Reject *is declared if $S_n < r_n^k$,* Continue *is declared if $r_n^k \leq S_n \leq a_n$, and* Accept *is declared is if $S_n > a_n$.*

Further, let $T_k^* = \mathbb{E}[V_k(1, S_1)] + 1$ be the optimal expected number of observations until a discovery is made. Then for each n, $r_n^k \to r_n$ as $k \to \infty$; and $T_k^* \to T^*$ as $k \to \infty$.

3.5 Computing the truncated optimal policy

The truncated horizon brings us closer to computing an optimal policy, but it is still an infinite horizon dynamic programming problem. In this section we show instead that we can compute the truncated optimal policy by iteratively solving a single-experiment truncated problem with a fixed rejection cost κ. Let $W_k(n, S_n|\kappa)$ be the optimal expected cost for this problem starting from state (n, S_n). We have the following Bellman equation.

$$W_k(n, S_n|\kappa) = 0, \quad S_n > a_n; \tag{7}$$

$$W_k(k, S_k|\kappa) = \kappa, \quad S_k \leq a_k; \tag{8}$$

$$W_k(n, S_n|\kappa) = 1 + \min\left\{\mathbb{E}[W_k(n+1, S_{n+1}|\kappa)|S_n], \kappa\right\}, \quad n < k, S_n \leq a_n. \tag{9}$$

For any terminal cost κ, this dynamic programming problem is easily solved using backward induction to find the rejection boundaries. The following theorem shows how we can use this solution to find an optimal policy to the truncated problem.

Theorem 6. *If $\kappa = T_k^*$, then the optimal policy for* (7)-(9) *with rejection thresholds $\overline{r}_n^k$ found by backward induction satisfies $\overline{r}_n^k = r_n^k$ for all $n \leq k$. Furthermore, let $f(\kappa) = 1 + \mathbb{E}[W_k(1, S_1|\kappa)]$ be the optimal cost. Then if $\kappa > T_k^*$, $f(\kappa) < \kappa$, and if $\kappa < T_k^*$, then $f(\kappa) > \kappa$.*

Thus, to find approximately optimal rejection thresholds, select k suitably large, and start with an arbitrary κ. Then iteratively compute the corresponding thresholds $\overline{r}_n^k$ and the cost $f(\kappa)$, using bisection to converge on T_k^*, and thus the corresponding optimal thresholds.

We note that the same program we have outlined in this section can be used to compute an optimal policy with a per experiment fixed cost c, by using rejection cost $\kappa + c$ instead of κ. Empirically, this leads to only slightly lower rejection thresholds; due to space constraints, we omit the details.

3.6 HEURISTIC APPROXIMATION

We have seen that the optimal policy is easy to approximate by solving dynamic programs iteratively. However, this does not give us direct insight into the structure of the solution, and in certain cases a quick rule-of-thumb that provides an approximate policy might be all that is required. In this section, we show that there exists a simple heuristic that performs remarkably well.

The approximate rejection boundary at time n is found as follows. Let $\hat{\mu}$ be the MAP estimate of μ for sufficient statistic $S_{n+T^*} = a_{n+T^*}$. Then reject the current experiment if S_n is not plausible under $\hat{\mu}$. That is, the heuristic boundary $\overline{r}_n$ is, for a suitably chosen β,

$$\mathbb{P}(S_n \leq \overline{r}_n \mid \mu = \hat{\mu}) = \beta. \tag{10}$$

Of course, this heuristic is not practical as is, as in general we do not know T^* unless we compute the optimal policy. But often a_{n+t} varies only little in t so a reasonable approximate choice T_h is sufficient. In Figure 1 we plot the discovery and rejection boundaries, along with the heuristic outlined above (with $T_h = T^*$), for the normal and Bernoulli models.

The heuristic and optimal policies clearly exhibit aggressive rejection regions, cf. Figure 1. THe interpretation is as follows: to continue sampling from the current experiment, we do not just want its quality to be s, but substantially better than s, since $a_n > s$ for all n. If not, it would take too many additional observations to verify the discovery.

4 CASE STUDY: BASEBALL

We now empirically analyze our testing framework based on a simulation with baseball data. First, we demonstrate empirically that the proposed algorithm leads to fast discoveries, and behaves differently from traditional testing approaches. Second, we show that

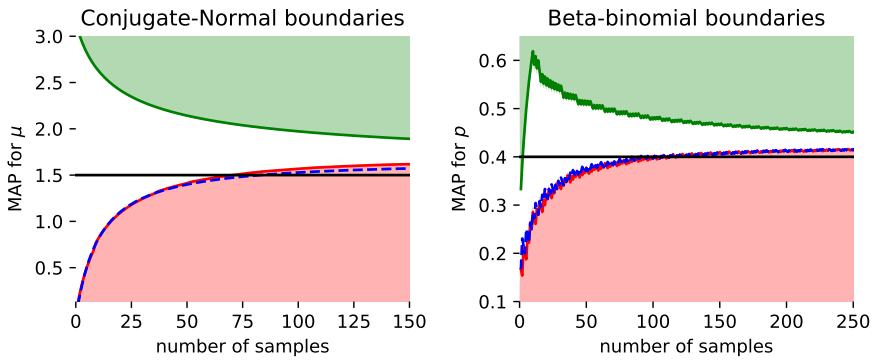

Figure 1: Acceptance and rejection regions for the conjugate Normal and the Beta-Binomial models. The dashed blue line gives the heuristic rejection boundary, while the red line corresponds to the optimal rejection thresholds. Note that the boundaries are shown in terms of the MAP estimate.

the rule-of-thumb heuristic performance is close to that of the optimal policy.[2]

Data We use the baseball dataset with pitching and hitting statistics from 1871 through 2016 from the Lahman R package. The number of At Bats (AB) and Hits (H) is collected for each player, and we are interested in finding players with a high batting average, defined as $b_i = \text{Hits}_i/\text{At Bats}_i$. We consider players with at least 200 At Bats, which leaves a total of 5721 players, with a mean of about 2300 At Bats. In the top left of Figure 2, we plot the histogram of batting averages, along with an approximation by a beta distribution (fit via method of moments). We note that these fit the data reasonably well, but not perfectly. This discrepancy helps us evaluate the robustness to a misspecified prior.

Simulation setup To construct the testing problem, we view the batters as alternatives, with empirical batting average b_i of batter i treated as ground truth. We want to find alternatives with $b_i > s$. We draw a Bernoulli sample of mean b_i to simulate an observation from alternative i. These samples are then used to test whether $b_i > s$. We set $\alpha = 0.05$, and vary s between 0.25 and 0.32. For each simulation, we iterate through each batter and repeat it 1000 times to reduce variance. This allows us to compare methods fairly, ensuring that each procedure is run on exactly the same test cases.

4.1 Benchmarks

To assess performance, we compare several testing procedures. Note that the non-traditional setup of our testing framework does not allow for easy comparison with other methods, in particular frequentist approaches, as they give different guarantees. Thus, we restrict attention to Bayesian methods that provide the same error guarantee. All of the benchmarks use the same beta prior computed above.

Optimal policy First we study the optimal policy based on the beta-binomial model, computed using the bisection and backward induction approach in Section 3.5, where we truncate after $k = 5000$ samples.

Heuristic policy Next, we include the heuristic rejection thresholds that approximate the optimal policy for truncation $k = 5000$ samples. The heuristic policy requires setting two parameters: T_h, i.e., how far to look into the future to find the acceptance boundary, which is ideally set close to T^*; and the rejection region β. To demonstrate the the insensitivity to T^*, we use $T_h = 2000$ and $\beta = 0.2$ for all simulations. (Note that T^* varies dramatically as we change the threshold s.

Fixed sample size test Our next benchmark is a simple fixed sample size test. For each experiment, we gather N observations, and claim a discovery if $P(\mu_i < s \mid Y_i) < \alpha$ where Y_i is the number of Hits of alternative (batter) i. We focus our attention on using $N = 1000$ samples per test, as this seems to perform best when compared to other sample sizes, but any differences are immaterial for our conclusions.

Fixed sample size test with early stopping This benchmark is similar to the fixed sample size test, except that we stop the experiment early if the discovery criterion is met. Thus, we can quantify the gains from being able to discover early.

Bayesian sequential test Now we consider a sequential test that also rejects early. In particular, we reject the current experiment if $\mathbb{P}(b_i > s \mid \mathbf{S}_t^i) < \beta$. We also reject an alternative after 4000 samples. This ap-

[2] Code to replicate results will be made publicly available.

proach also requires careful tuning of β. In particular, if β is too large, say larger than the prior probability $\mathbb{P}_0(b_i > s)$, then the test is too aggressive and rejects all alternatives outright. Instead, we found empirically that setting $\beta = 0.9\mathbb{P}_0(b_i > s)$ leads to good performance across all values of s.

4.2 Results

Average time to discovery The average number of observations until a discovery is shown in the top right plot of Figure 2. As expected, the fixed sample test performs worst. Early stopping leads to slightly better performance, but this method is still not effective as most of the gains come from early rejection. The Bayesian sequential test demonstrates this effect and shows substantial gains over the fixed tests. The heuristic policy, despite lack of parameter tuning, performs very well, essentially matching the performance of the optimal algorithm for most thresholds.

False discovery proportion and robustness Next, we compare the *false discovery proportion* (FDP) (Benjamini and Hochberg, 2007), i.e., the fraction of discoveries that in fact had true $b_i < s$. If the prior is correctly specified, the methods we consider satisfy $\mathbb{E}(\text{FDP}) \leq \alpha$.[3] Indeed, we observe that the guarantee holds for most thresholds and algorithms in the bottom left plot of Figure 2. There is some minor exceedance of the FDP for thresholds around $s \approx 0.28$, which can be explained by the fact that the prior does not fit the empirical batting averages perfectly. Since there are few rejections for thresholds beyond $s = 0.3$, the FDP estimate has higher variance in that regime. Across all simulations, the optimal policy has an FDP of $0.048 < \alpha$. Finally, we see that the lack of early stopping makes the fixed test rather conservative.

The paradox of power Finally, we compare *power*, i.e., the fraction of alternatives i with $b_i > s$ that are declared a discovery. Power comparisons across the algorithms are plotted in the bottom right of Figure 2. The most surprising insight from the simulations is the *paradox of power*. Algorithms that are effective have very low power. This is counter-intuitive: how can algorithms that make many discoveries have only a small chance of picking up true effects? The main driver of good performance for an algorithm is the ability to quickly reject unpromising alternatives. Some unpromising alternatives are "barely winners": i.e., b_i is only slightly above s. In the experiment-rich regime, such alternatives should be rejected quickly, because it takes too many observations to get enough concentration around the posterior to claim a discovery. This effect leads to low power, but fast discoveries.

[3]However, note that this is different from frequentist FDR guarantees, which these methods do not provide.

5 CONCLUSION

We consider an experimentation setting where observations are costly, and there is an abundance of possible experiments to run — an increasingly prevalent scenario as the world is becoming more data-driven. Based on backward induction, we can compute an approximately optimal algorithm that allocates observations to experiments such that the time to a discovery is minimized. Simulations validate the efficacy of our approach, and also reveal discuss the *paradox of power*: there is a tension between high-powered tests, and being efficient with observations.

Our paradigm has several additional practical benefits. First, we can leverage knowledge across experiments through the prior. Second, adaptive matching of observations to experiments does not preclude valid inference, and thus outcomes can thus continuously be monitored. Finally, the framework also provides an easy "user interface": it directly incorporates the desired effect size, and leads to guarantees that are easy to explain to non-experts.

ACKNOWLEDGEMENTS

We would like to thank Johan Ugander, David Walsh, Andrea Locatelli, and Carlos Riquelme for their helpful comments and suggestions. This research was done while Sven Schmit was at Stanford University. The research was supported by the National Science Foundation under grants 1544548 and 1839229.

References

E. M. Azevedo, A. Deng, J. M. Olea, J. M. Rao, and E. G. Weyl. A/b testing. In *Proceedings of the Nineteenth ACM Conference on Economics and Computation*. ACM, 2018.

M. Aziz, J. Anderton, E. Kaufmann, and J. A. Aslam. Pure exploration in infinitely-armed bandit models with fixed-confidence. In *ALT*, 2018.

E. Bakshy, D. Eckles, and M. S. Bernstein. Designing and deploying online field experiments. In *Proceedings of the 23rd international conference on World wide web*, pages 283–292. ACM, 2014.

A. Balsubramani and A. Ramdas. Sequential nonparametric testing with the law of the iterated logarithm. *CoRR*, abs/1506.03486, 2016.

A. Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. 2007.

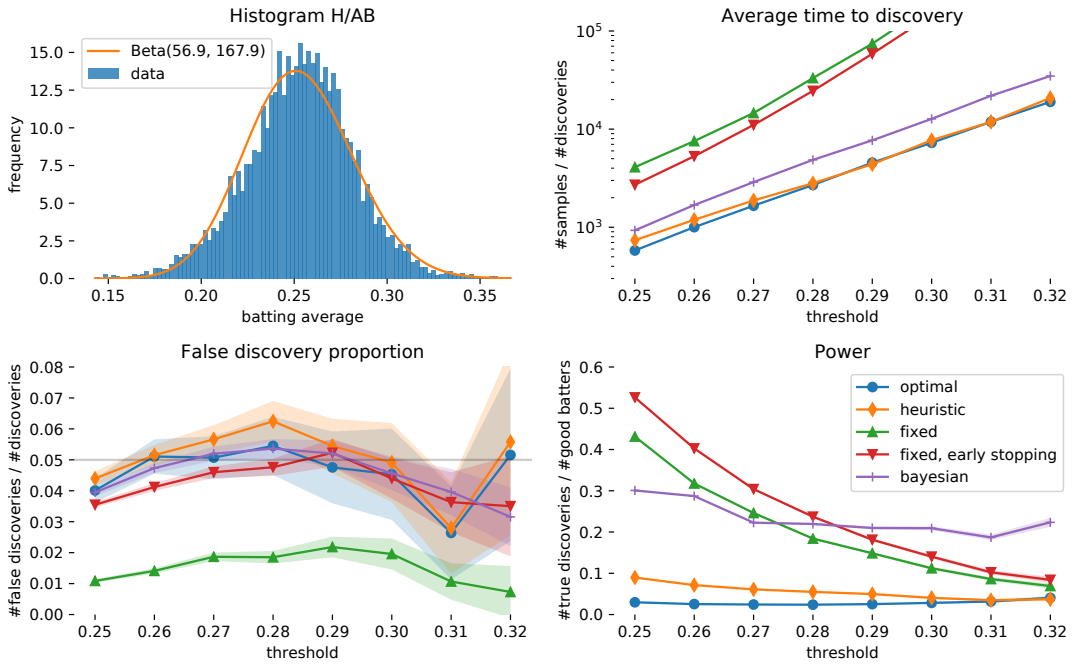

Figure 2: Top left: histogram of batting averages. Top right: Efficacy of of algorithms. Bottom left: Plot the false discovery proportion across thresholds. Bottom right: Plot of the empirical power of algorithms. Note the *paradox of power* effect: the most efficient algorithms have low power.

S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5:1–122, 2012.

S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in multi-armed bandits problems. In *ALT*, 2009.

A. Carpentier and M. Valko. Simple regret for infinitely many armed bandits. In *ICML*, 2015.

K. Chandrasekaran and R. M. Karp. Finding the most biased coin with fewest flips. *CoRR*, abs/1202.3639, 2012.

A. R. Chaudhuri and S. Kalyanakrishnan. Pac identification of a bandit arm relative to a reward quantile. In *AAAI*, pages 1777–1783, 2017.

H. Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770, 1959.

A. Deng, Y. Xu, R. Kohavi, and T. Walker. Improving the sensitivity of online controlled experiments by utilizing pre-experiment data. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 123–132. ACM, 2013.

A. Deng, J. Lu, and S. Chen. Continuous monitoring of a/b tests without pain: Optional stopping in bayesian testing. *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 243–252, 2016.

A. Deng, J. Lu, and J. Litz. Trustworthy analysis of online a/b tests: Pitfalls, challenges and solutions. In *WSDM*, 2017.

D. P. Foster and R. A. Stine. Alpha-investing: A procedure for sequential control of expected false discoveries. 2007.

P. Freeman. The secretary problem and its extensions: A review. *International Statistical Review/Revue Internationale de Statistique*, pages 189–206, 1983.

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

D. Goldberg and J. E. Johndrow. A decision theoretic approach to a/b testing. *arXiv preprint arXiv:1710.03410*, 2017.

L. Jain and K. Jamieson. Firing bandits: Optimizing crowdfunding. In *International Conference on Machine Learning*, pages 2211–2219, 2018.

K. G. Jamieson, M. Malloy, R. D. Nowak, and S. Bubeck. lil' ucb : An optimal exploration algorithm for multi-armed bandits. In *COLT*, 2014.

K. G. Jamieson, D. Haas, and B. Recht. The power of adaptivity in identifying statistical alternatives. In *NIPS*, 2016.

A. Javanmard and A. Montanari. Online rules for control of false discovery rate and false discovery exceedance. *CoRR*, abs/1603.09000, 2016.

R. Johari, P. Koomen, L. Pekelis, and D. Walsh. Peeking at a/b tests: Why it matters, and what to do about it. In *KDD*, 2017.

R. L. Kaufman, J. Pitchforth, and L. Vermeer. Democratizing online controlled experiments at booking.com. *arXiv preprint arXiv:1710.08217*, 2017.

E. Kharitonov, A. Vorobev, C. MacDonald, P. Serdyukov, and I. Ounis. Sequential testing for early stopping. 2015.

R. Kohavi, T. Crook, R. Longbotham, B. Frasca, R. Henne, J. L. Ferres, and T. Melamed. Online experimentation at microsoft. 2009.

L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis. Quickest search over multiple sequences. *IEEE Transactions on Information Theory*, 57:5375–5386, 2011.

T. L. Lai. On optimal stopping problems in sequential hypothesis testing. *Statistica Sinica*, 7(1):33–51, 1997.

A. Locatelli, M. Gutzeit, and A. Carpentier. An optimal algorithm for the thresholding bandit problem. In *ICML*, 2016.

M. Malloy, G. Tang, and R. D. Nowak. Quickest search for a rare distribution. *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2012.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

A. Ramdas, F. Yang, M. J. Wainwright, and M. I. Jordan. Online control of the false discovery rate with decaying memory. In *Advances in Neural Information Processing Systems*, pages 5655–5664, 2017.

D. Russo. Simple bayesian algorithms for best arm identification. In *Conference on Learning Theory*, pages 1417–1418, 2016.

S. M. Samuels. Secretary problems. *Handbook of sequential analysis*, 118:381–405, 1991.

A. N. Shiryayev. Optimal stopping rules, volume 8 of applications of mathematics, 1978.

D. Siegmund. *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media, 2013.

D. Tang, A. Agarwal, D. O'Brien, and M. Meyer. Overlapping experiment infrastructure: More, better, faster experimentation. In *Proceedings 16th Conference on Knowledge Discovery and Data Mining*, pages 17–26, Washington, DC, 2010a.

D. Tang, A. Agarwal, D. O'Brien, and M. Meyer. Overlapping experiment infrastructure: More, better, faster experimentation (presentation). 2010b. URL `https://static.googleusercontent.com/media/research.google.com/en//archive/papers/Overlapping_Experiment_Infrastructure_More_Be.pdf`.

A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, pages 326–339, 1948.

G. B. Wetherill and K. D. Glazebrook. Sequential methods in statistics, 1986.

D. Williams. *Probability with martingales*. Cambridge university press, 1991.

F. Yang, A. Ramdas, K. G. Jamieson, and M. J. Wainwright. A framework for multi-a (rmed)/b (andit) testing with online fdr control. In *Advances in Neural Information Processing Systems*, pages 5959–5968, 2017.

A PROOFS

A.1 Proofs from Section 2

Proof of Lemma 1. The result relies on τ being a stopping time. Recall that i^* indicates the discovered experiment. Then we find

$$\begin{aligned}\mathbb{P}(\mu_{i^*} < s \mid \mathcal{F}_\tau) &= \sum_{t=1}^{\infty} \mathbb{P}(\mu_i^* < s \mid \mathcal{F}_\tau \cap \{\tau = t\})\mathbb{P}(\tau = t) \\ &= \sum_{t=1}^{\infty} \mathbb{P}(\mu_i^* < s \mid \mathcal{F}_t)\mathbb{P}(\tau = t) \\ &\leq \alpha \sum_{t=1}^{\infty} \mathbb{P}(\tau = t) = \alpha\end{aligned}$$

where we use that $F \in \mathcal{F}_\tau$ if $F \cap \{\tau = t\} \in \mathcal{F}_t$ (Williams, 1991)[p.219]. □

Proof of Lemma 2. Note that due to independence we can assume without loss of generality that the index of the arm corresponds to the order in which alternatives are first considered. Thus the result follows if we show that for any t, action $I_t < I_{t-1}$ cannot be strictly better than $I_t = I_{t+1}$. Assume to the contrary that $I_t = y$ is optimal (and strictly better than $I_t = I_{t-1} + 1$ for some $y < I_t$. Consider the last time alternative y was selected: $t' = \max\{k < t : I_k = y\}$. At that time it was at least as good to consider a new alternative, and subsequently the posterior for alternative y has not changed due to independence. Due to the infinite time horizon, it is thus at least as good to consider a new alternative. □

A.2 Proofs from Section 3

Proof of Lemma 3. Let $n \geq 1$. We can rewrite the discovery criterion as

$$\mathbb{P}(\mu < s \mid S_n = t) = \frac{\int_{-\infty}^{s} \prod_{i=1}^{n} h(X_i) \exp\left(\mu S(X_i) - A(\mu)\right) d\pi(\mu)}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} h(X_i) \exp\left(\mu S(X_i) - A(\mu)\right) d\pi(\mu)} \tag{11}$$

$$= \frac{\int_{-\infty}^{s} \exp\left(\mu S_n - nA(\mu)\right) d\pi(\mu)}{\int_{-\infty}^{\infty} \exp\left(\mu S_n - nA(\mu)\right) d\pi(\mu)} \tag{12}$$

$$= \frac{\int_{-\infty}^{s} \exp\left(\mu t - nA(\mu)\right) d\pi(\mu)}{\int_{-\infty}^{\infty} \exp\left(\mu t - nA(\mu)\right) d\pi(\mu)} \tag{13}$$

We show that this is decreasing in t.

Now take the logarithm and the derivative with respect to t to obtain

$$\frac{d}{dt} \log(\mathbb{P}(\mu < s \mid S_n = t)) = \frac{\int_{-\infty}^{s} \mu \exp(\mu t - nA(\mu)) d\pi(\mu)}{\int_{-\infty}^{s} \exp(\mu t - nA(\mu)) d\pi(\mu)} \tag{14}$$

$$- \frac{\int_{-\infty}^{\infty} \mu \exp(\mu t - nA(\mu)) d\pi(\mu)}{\int_{-\infty}^{\infty} \exp(\mu t - nA(\mu)) d\pi(\mu)} \tag{15}$$

$$= \mathbb{E}_{f_t}(\mu \mid \mu < s) - \mathbb{E}_{f_t}(\mu) < 0 \tag{16}$$

where the expectations in the last line is taken with respect to the distribution with density

$$f_t(\mu) = \frac{\exp(\mu t - nA(\mu)) d\pi(\mu)}{\int_{-\infty}^{\infty} \exp(\mu t - nA(\mu)) d\pi(\mu)} \tag{17}$$

Note that the last inequality holds, because, in general

$$\mathbb{E}(\theta) = \mathbb{E}(\theta \mid \theta < s)\mathbb{P}(\theta < s) + \mathbb{E}(\theta \mid \theta \geq s)\mathbb{P}(\theta \geq s) > \mathbb{E}(\theta \mid \theta < s)\mathbb{P}(\theta < s) + s\mathbb{P}(\theta \geq s) > \mathbb{E}(\theta | \theta < s) \tag{18}$$

Now the lemma follows: if $\mathbb{P}(\mu < s \mid S_n = t) < \alpha$, then $\mathbb{P}(\mu < s \mid S_n = t') < \alpha$ for all $t' > t$, and similarly if if $\mathbb{P}(\mu < s \mid S_n = t) > \alpha$, then $\mathbb{P}(\mu < s \mid S_n = t') > \alpha$ for all $t' < t$. □

To prove the theorems in Section 3 we use the following lemmas, which are proven at the end of this section.

Lemma 7. *The optimal policy for the truncated problem can be characterized by a rejection threshold. That is, the optimal policy rejects the current experiment if $S_n < r_n^k$ for a sequence r_n^k, and collects another observation for the current experiment otherwise, until a discovery is made.*

Write T_k for the expected number of observations required for a discovery for the optimal policy of the truncated problem. Then we can show that both T_k and r_n^k converge.

Lemma 8. *Both T_k and r_n^k converge as $k \to \infty$.*

Proof of Theorem 4. Lemma 7 shows that the truncated problem has an optimal policy that has the form of a threshold. Next, lemma 8 shows that both the thresholds and the optimal cost converge.

Recall $T^* = \lim_{k\to\infty} T_k^*$ and $r_n = \lim_{k\to\infty} r_n^k$. Now we show that limiting policy r_n with corresponding cost T^* is optimal.

Suppose there exists an $\varepsilon > 0$ and a policy $\overline{\phi}$ with cost $\overline{T}$ such that $\overline{T} = T^* - \varepsilon$. Consider a policy with cost $\overline{T} < T^*$. Let $\overline{\tau}$ be the stopping time of this policy. We consider the truncated version of this policy, and show that it cannot be much worse. On the other hand, this truncated policy has a cost larger than T^*. The k-truncated policy, denoted by $\overline{\phi}_k$ rejects the current alternative after k samples, but is otherwise identical to $\overline{\phi}$. Let $\overline{\tau}$ and $\overline{\tau}_k$ be the stopping times corresponding to $\overline{\phi}$ and $\overline{\phi}_k$. Trivially, we have $\overline{T} = \sum_{k=1}^{\infty} \mathbb{P}(\overline{\tau} \geq k)$. Because $\overline{T}$ is finite, $\mathbb{P}(\overline{\tau} \geq k) = O((k \log k)^{-1})$. Because $\overline{\phi}$ and $\overline{\phi}_k$ are identical up to k observations, it follows that if $\overline{\tau} < k$, then $\overline{\tau}_k < k$, and thus we find that

$$\begin{aligned}\mathbb{E}(\overline{\tau}_k) &= \mathbb{P}(\overline{\tau} > k)\mathbb{E}(\overline{\tau}_k \mid \overline{\tau} > k) + \mathbb{E}(\overline{\tau}\mathbb{I}(\overline{\tau} \leq k)) \\ &\leq \mathbb{P}(\overline{\tau} > k)(k + \mathbb{E}(\overline{\tau}_k)) + \mathbb{E}(\overline{\tau}\mathbb{I}(\overline{\tau} \leq k))\end{aligned}$$

Thus, it follows that

$$\mathbb{E}(\overline{\tau}_k) \leq \frac{k\mathbb{P}(\overline{\tau} > k)}{1 - \mathbb{P}(\overline{\tau} > k)} + \frac{\overline{T}}{1 - \mathbb{P}(\tau > k)}. \tag{19}$$

Since $\mathbb{P}(\overline{\tau} > k) = O((k \log k)^{-1})$, $\mathbb{E}(\overline{\tau}_k) \to \overline{T}$ as $k \to \infty$.

However, $T^* \leq \mathbb{E}(\overline{\tau}_k)$ for all k, and thus $T^* \leq \lim_{k\to\infty} \mathbb{E}(\overline{\tau}_k) = \overline{T}$, which is a contradiction. □

Proof of Theorem 5. This is a direct consequence of lemmas 7 and 8. □

Proof of Theorem 6. Let τ_r denote the (random) hitting time of the boundary of the first alternative

$$\tau_r = \min\{n : S_n \geq a_n \text{ or } S_n < r_n\} \tag{20}$$

under rejection boundary $r = \{r_n\}_{i=1}^k$. Furthermore, let $q_r = \mathbb{P}(S_\tau < r_\tau)$ denote the rejection probability. Now note that $f(\kappa) = \min_r \mathbb{E}(\tau_r) + \kappa q_r$ Note that we can solve this minimization problem using backward induction, since the time horizon is fixed (k). First, we show that f has a unique fixed point which is equal to T_k^*.

Note that we have

$$T_r = \mathbb{E}(\tau_r) + T_r q_r \tag{21}$$

By definition, T_k^* minimizes $\min_r \mathbb{E}(\tau_r) + T_k^* q_r$, thus, it follows immediately that T_k^* is a fixed point of f.

Next, we show that $f(\kappa) > \kappa$ for each $\kappa < T_k^*$ and $f(\kappa) < \kappa$ for each $\kappa > T_k^*$.

First, fix $\kappa < T_k^*$. Suppose that $f(\kappa) \leq \kappa$. Thus, there exists r' such that $\mathbb{E}(\tau_{r'}) + \kappa q_{r'} \leq \kappa$. Thus, $\kappa \geq \frac{\mathbb{E}(\tau_{r'})}{1-q_{r'}} = T_{r'}$, where the last equality follows from (21). This, along with $\kappa < T_k^*$ implies that $T_{r'} < T_k^*$, a contradiction. Thus, we must have $f(\kappa) > \kappa$.

Finally, fix $\kappa > T^*$. We know that

$$T^* = \mathbb{E}(\tau_{r^*}) + T_k^* q_{r^*} < \mathbb{E}(\tau_{r^*}) + \kappa q_{r^*}. \tag{22}$$

Thus, there exists r (equal to r^*) such that $\mathbb{E}[\tau_r] + \kappa q_r < \kappa$. Thus, $f(\kappa) < \kappa$. □

A.3 Proofs of lemmas

Proof of Lemma 7. Based on Lemma 2, there exists a policy that can be characterized by a sequence of three sets

- *Discover* if $S_n \in A_n$, the experiment is a discovery
- *Continue* if $S_n \in D_n$, and
- *Reject* if $S_n \in R_n$

Now note that R_n^k is a threshold region for $n \geq k$ by definition. Assume $R_m^k = (-\infty, r_m^k]$ for all $m > n$. Further, from the Bellman equation for the truncated problem, it is clear that the optimal solution rejects the current experiment at the time n if

$$\mathbb{E}[V_k(n+1, S_{n+1}) \mid S_n] > \mathbb{E}[V_k(1, S_1)] = T_k^* - 1. \tag{23}$$

Note that

$$\mathbb{E}[V_k(n+1, S_{n+1}) \mid S_n = x] = \int_{y \in D_{n+1}} V_k(n+1, y) f(S_{n+1} = y \mid S_n = x) dy + T_k^* \mathbb{P}(S_{n+1} < r_{n+1} \mid S_n = x) \tag{24}$$

Then for n we note that $\mathbb{E}[V_k(n+1, S_{n+1}) \mid S_n = x]$ is decreasing in x. This follows since $V_k(n+1, y) < T_k^*$ for all $y \in D_{n+1} = [r_{n+1}, a_{n+1}]$, as for such y it is better to continue than to reject. Furthermore, arguing along the lines of the proof of Lemma 3, $\mathbb{P}(S_{n+1} < r_{n+1} \mid S_n = x)$ is decreasing in x. This implies we can write $R_n^k = (-\infty, r_n^k]$ for some r_n^k. □

Proof of Lemma 8. Due to increased degrees of freedom, it follows that T_k^* is decreasing. Since T_k^* is bounded below by 0, T_k^n converges. Let $T^* = \lim_k T_k^*$.

Next, we show that r_n^k is decreasing in k: Clearly, $r_k^{k+1} \leq r_k^k$. Now suppose $r_{n+1}^{k+1} \leq r_{n+1}^k$, then $r_n^{k+1} \leq r_n^k$, which follows from the fact that $\mathbb{E}[V_k(n+1, S_{n+1}) \mid S_n = r_n^k] + 1 = T_k^*$, and T_k^* is decreasing in k. It remains to show that r_n^k is bounded.

We construct a lower bound on r_n^k, for large k, as follows. Let $\varepsilon = \frac{1}{2T^*}$ and let x be such that $\mathbb{P}(\exists m > n \text{ s.t. } S_m > a_m \mid S_n = x) < \varepsilon$, by choosing x sufficiently small. Then we note that the cost for obtaining another sample is at least $1 + (1-\varepsilon)T_k^* \geq 1 + (1-\varepsilon)T^* = T^* + 1/2$. However, if the experimenter rejects the current alternative now, the cost is T_k^*. Thus, if we can show that there exists a K such that for all $k > K$, $T_k^* < T^* + 1/2$, then x is a lower bound on r_n^k for all $k > K$. But above we have shown that $T_k^* \to T^*$, hence such K exists. This implies that r_n^k converges as $k \to \infty$. □