HOW TO DESIGN AND ANALYZE ONLINE A/B TESTS WITHIN DECENTRALIZED ORGANIZATIONS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF STATISTICS AND THE COMMITTEE ON GRADUATE STUDIES OF STANFORD UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

David Walsh
August 2019
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Ramesh Johari, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Bradley Efron

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Art Owen

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.
Abstract

From e-commerce to digital media to social networks, essentially any company that does business online is using A/B tests – randomized experiments on its customers – in order to optimize its service. A standard approach to designing and analyzing these experiments, based on classical statistical theory, is ubiquitous in industrial practice. Namely, a sample size for each test should be set in advance, and the data collected should be analyzed in isolation through p-values and confidence intervals that are computed based on a 2-sample t-test of means.

This thesis investigates four ways that this default approach proves insufficient, due to the decentralized manner in which A/B tests at these companies are run:

1. Experimenters adjust the sample size for their experiments dynamically depending on data observed. This inflates the error rates associated with any inferences.

2. Experimenters adjust sample sizes dynamically to respond to treatment effects that evolve over time. The inferences derived may not offer accurate information about the average effect over any suitable timeframe.

3. Experimenters analyze multiple A/B tests in parallel. The t-test inference measures do not offer sufficiently strict error guarantees for this use-case.

4. Experimenters do not prioritize data collection that will enable other employees at the company to design productive A/B tests later on.

For each of these four, we empower experimenters to continue their current behavior, while we offer these companies novel methodology, which is simple to implement and recovers inferential reliability. For the first three, we offer alternative methods to the t-test for computing p-values and confidence intervals, which are respectively robust to dynamic sample sizes, provide inference on average treatment effects when effects change, and control family-wise and false discovery error rates across multiple experiments. For the fourth, we force each experimenter to internalize the value of future experiments through a mechanism that regulates how much past experimental data she can access at any given time.
Much of the work presented was conducted in collaboration with Microsoft and the commercial A/B testing platform, Optimizely. Some of our methodology is now in deployment at Optimizely, and we present empirical data from that deployment to supplement our theoretical results.
## Contents

**Abstract**

1 Introduction

1.1 What do A/B tests look like?

1.1.1 Notation

1.1.2 The default statistical analysis

1.2 Contributions

1.3 Related work

1.3.1 A/B testing practice

1.3.2 Experimental design

1.3.3 Automation

1.3.4 Other statistical models

1.4 Acknowledgements

2 Always valid inference

2.0.1 Contributions

2.0.2 Related work

2.1 Preliminaries

2.1.1 Testing a simple null hypothesis

2.1.2 Extension to A/B tests

2.2 Always valid inference

2.3 Optimal p-value processes

2.3.1 The user model

2.3.2 The mSPRT

2.3.3 Efficiency in the $\alpha \to 0$ limit

2.3.4 Run-length optimization

2.3.5 Improvement over fixed-horizon

2.4 Extension to A/B tests
4.3 Always valid p-values ........................................ 47
  4.3.1 Bonferroni and BH-G .................................. 47
  4.3.2 When does BH-I control FDP? ......................... 47
4.4 Always valid confidence intervals ......................... 49
4.5 Conclusion .................................................. 50

5 Incentivizing better experiment designs ...................... 51
  5.0.1 Contributions ........................................ 53
  5.0.2 Related work .......................................... 54
5.1 Preliminaries ................................................ 54
  5.1.1 Properties of each A/B test .......................... 54
  5.1.2 Using the data ........................................ 56
  5.1.3 Summary ............................................... 58
  5.1.4 Limitations ............................................ 58
5.2 The experiment design game .................................. 59
  5.2.1 The objectives of the agents ......................... 60
  5.2.2 Boundedly rational Markov perfect equilibrium .... 61
  5.2.3 The objective for the company ....................... 63
5.3 Exploration in equilibrium .................................. 65
  5.3.1 BR-MPE ............................................... 65
  5.3.2 Characterizing the benchmark ....................... 66
  5.3.3 Under-exploration .................................... 67
5.4 The Minimum Learning Mechanism ......................... 67
  5.4.1 No private information ............................... 67
  5.4.2 Performance for general $\rho$ ...................... 68
5.5 Discussion ................................................... 70
  5.5.1 The price of anarchy ................................ 70
  5.5.2 Robustness to private information ................ 71
  5.5.3 Generalization to other corporate structures ...... 71

6 Future work ....................................................... 73
  6.1 A/B testing over time ..................................... 74
  6.2 Broader incentive issues .................................. 75
  6.3 Beyond parametric inference ............................. 76

Bibliography .......................................................... 78

A Optimality of the mSPRT for one-variation experiments .... 83
B. Optimality of the mSPRT for two-stream normal data

C. Type I error control for the mSCPRT

D. Sequential FDR control

E. FDR control

F. Constructing the experiment design game

G. Performance of the Minimum Learning Mechanism
Chapter 1

Introduction

A/B tests are the randomized experiments that internet companies run on their customers so they can measure the effectiveness of any potential improvements to their service. Indeed A/B testing represents the most popular and reliable mechanism for product optimization. It is essential that a level of statistical rigor is applied to the design and analysis of these experiments to ensure that the insights derived are accurate.

In this opening chapter, we summarize the core statistical components of a typical A/B test, as well as summarizing the default approach to how they are run. In so doing, we identify some crucial ways that this default has proven insufficient in industrial practice. Novel methodology, which can help internet companies overcome such challenges, is the target of this thesis.

1.1 What do A/B tests look like?

Generally when an employee decides to run a new A/B test, it is because she has developed a new idea and she wants to know whether it improves on the current state of the service. At its simplest, setting up the A/B test requires her to specify two things:

- A treatment and a control. Her new idea must be coded up as a “treatment” experience that can be presented to customers when they use the service. The pre-existing version of the service is defined to be the “control” experience.

- A metric. After subjects in the A/B test are exposed to either the treatment or the control, their interaction with the service will be monitored. A single numerical measurement is recorded on each subject (typically more positive measurements are considered to be indicative of a more positive customer experience). The treatment effect is defined as the mean response under the treatment minus the mean response under the control.
The subjects in the experiment are simply whichever customers use the service while the A/B test is running. Each successive customer that arrives is independently randomized to receive either the treatment or the control experience, and the relevant measurement is taken on her. When the A/B test is over, inference is sought on the treatment effect, which will enable the experimenter to optimize some corresponding decisions. For instance, the most common decision is whether to deploy the treatment permanently to all future customers or to revert to the control experience.

Example 1. A designer at an e-commerce company is responsible for making the webpage for a given product more visually attractive. The goal is that, of those customers who arrive at the page, a higher proportion will add the product to their cart. The treatment is her new design of the page, whereas the control is the pre-existing design. Measurements are binary: one if the customer adds the product to her cart and zero otherwise. The treatment effect is the corresponding difference in means: the probability of adding to cart under the treatment minus that same probability under the control.

Example 2. A machine learning engineer at an e-commerce company develops a new algorithm for recommending products to customers. The goal is to increase the proportion of customers who click on a recommendation, out of those who are shown a recommended product. The treatment is the new algorithm; the control is the existing algorithm. Again measurements are binary, indicating whether the exposed customer clicked on the recommendation, and the effect is the corresponding difference in click probabilities.

Example 3. A content editor at an entertainment website suggests a new layout for the content that she considers to be more engaging. The goal is to keep customers on the website for longer. The treatment is her new layout, while the control is the existing layout. The measurements are the duration of time from when a customer arrives at the site until she leaves. The treatment effect is the difference in the mean time on site.

1.1.1 Notation

As the experiment runs, two sequences of measurements are obtained: measurements on those customers who are randomized to receive the treatment and measurements on those who receive the control. We denote these sequences by $Y^t_1, Y^t_2, \ldots$ and $Y^c_1, Y^c_2, \ldots$. We consider two models for the marginal distribution of each measurement. For binary-valued responses, we have:

\[ Y^t_k \sim \text{Bernoulli}(\mu^t), \quad Y^c_k \sim \text{Bernoulli}(\mu^c). \]  

(1.1)

For continuous data, we restrict to a Gaussian model with common variance:

\[ Y^t_k \sim \mathcal{N}(\mu^t, \sigma^2), \quad Y^c_k \sim \mathcal{N}(\mu^c, \sigma^2) \]  

(1.2)
In either case, the metric is the difference in means, $\theta = \mu^t - \mu^c$.

**Assumption 1.** In the case of continuous data, the common variance $\sigma^2$ is assumed known.

**Assumption 2.** Subjects in the experiment are assigned alternately to receive the treatment or the control, so that whenever the experiment terminates the same number of measurements are available for each of the two experiences. This is an approximation to the real setup, where each treatment assignment is an independent coin flip. The approximation places roughly the correct number of subjects in each group, provided the sample size is moderately large.

We define the filtration $(\mathcal{F}_n)_{n=1}^\infty$ generated by the pairs $W_n = (Y^t_n, Y^c_n)$. Sometimes in A/B testing practice, the sample size for an experiment is fixed in advance of collecting any data, whereas sometimes it is set dynamically mid-experiment depending on the data collected so far. In the first case, the fixed-horizon sample size $n$ is the number of measurements made on each experience. On completion of the experiment, the data gathered is represented by $\mathcal{F}_n$. In general, $T$ denotes the number of measurements made on each experience when the test is terminated. Setting the sample size dynamically means that $T$ is a stopping time with respect to $(\mathcal{F}_n)_{n=1}^\infty$. The data obtained is represented by the stopped $\sigma$-algebra, $\mathcal{F}_T$.

### 1.1.2 The default statistical analysis

The standard in industrial practice is to analyze the data collected in an A/B test using p-values and confidence intervals based on a two sample t-test of means. t-tests have been studied extensively in the classical statistical literature. In short, they view the sample difference of means $\hat{\theta}_n = \bar{Y}^t_n - \bar{Y}^c_n$ as a point estimate of $\theta$, where $\bar{Y}^t_n = \frac{1}{n} \sum_{k=1}^{n} Y^t_k$ and similarly for $Y^c_n$. Then they reject the null hypothesis that $\theta$ takes some given arbitrary value $\theta_0$ when $\hat{\theta}$ is sufficiently far from $\theta_0$.

This approach relies on two assumptions about the data generation process and how experiments are designed:

**Assumption 3.** Given values for the mean parameters $\mu^t$ and $\mu^c$, all measurements $Y^t_1, Y^t_2, \ldots$ and $Y^c_1, Y^c_2, \ldots$ are independent. The rationale here is that each customer is interchangeable and does not impact the behavior of any other.

**Assumption 4.** A fixed-horizon sample size $n$ is chosen in advance of the experiment.

Under these assumptions, we have that for binary data,

$$T_{\text{stat}}(\theta) := \sqrt{\frac{n}{Y^t_n(1 - Y^t_n) + Y^c_n(1 - Y^c_n)}} (\hat{\theta}_n - \theta) \approx N(0, 1),$$

while for normal data,

$$T_{\text{stat}}(\theta) := \frac{n}{2\sigma^2} (\hat{\theta}_n - \theta) \sim N(0, 1).$$
CHAPTER 1. INTRODUCTION

Figure 1.1: Results dashboard from the A/B testing platform Optimizely. Two parallel experiments are described: the treatment “[x] only” and the treatment “later only” are both compared against a control. Optimizely’s “Statistical significance” represents \((1 - \text{p-value}) \times 100\%\). The “Difference interval” is the confidence interval for the difference in success probabilities.

For a given significance level \(\alpha\), the t-test rejects the null \(H^0_0: \theta = \theta_0\) when \(|T_{\text{stat}}(\theta_0)| > z_{1-\alpha/2}\), where this threshold is the \((1 - \alpha/2)\) quantile of the standard normal distribution. This bounds the probability of a false positive at \(\alpha\) in the case that that null is true. Given \(\theta_0\) and \(n\) and any fixed alternative \(\theta'\), this test has approximately maximal power under \(\theta'\) among tests that achieve this Type I error rate bound. For a comprehensive description of the optimality properties of the t-test, see e.g. [48].

The t-test is mapped to a p-value for testing the null that \(\theta = 0\) by taking \(p_n\) to be the smallest \(\alpha\) at which that null is rejected. For a fixed significance level, a \((1 - \alpha)-\text{level confidence interval for } \theta\) is obtained as the set of \(\theta_0\) for which \(H^0_0\) is not rejected. Internet companies typically communicate these quantities via a dashboard such as the one shown in Figure 1.1. The p-value helps her decide whether the treatment improves on the control in terms of her chosen metric. The confidence interval gives her a range for the improvement, which can be valuable if this A/B test is just one piece in a larger cost-benefit analysis.

1.2 Contributions

The default analysis proves insufficient in practice, because it fails to address how A/B testers design experiments at these companies. In particular, their experiment design choices at any given time often depend on previously observed data. Because the default approach is oblivious to this endogeneity, it can produce unreliable inferences or it can lead to inefficient data collection.

This thesis offers novel methodology that addresses four common challenges that arise:

1. **Continuous monitoring of A/B tests** (Chapter 2). In violation of Assumption [\#, it is commonplace for experimenters to track p-values and confidence intervals on the dashboard.
as the test is running, and to stop the test dynamically based on what is observed. This issue is widely referred to in industry as the “continuous monitoring problem”, because the practice invalidates the inferential guarantees of these measures. On completion of the experiment, the resulting p-values and confidence intervals do not bound Type I error, leading to erroneous inferences and consequently poor decisions.

The reason that practitioners continuously monitor is that long experiments are costly, both in terms of the experimenter’s time (she must wait for the data before she can make a decision) and because the subjects in the experiment cannot be used for other A/B tests. Further, they find it difficult to pre-specify a fixed-horizon sample size, which prioritizes shorter experiments while ensuring that true effects can be detected with high probability, because this requires prior understanding of the magnitude of the effect sought.

We term our solution always valid p-values and confidence intervals, as they allow the experimenter to continuously monitor while still guaranteeing Type I error control. This lets her stop dynamically so as to trade off power against run-time. We prove that the always valid measures we derive produce an approximately optimal trade-off for a wide variety of potential experimenters.

2. Continuous monitoring under temporal noise (Chapter 3). Contrary to Assumption 3, customers of the service cannot always be considered interchangeable. Rather they may have latent characteristics that impact their responses. If similar customers tend to arrive at similar times, this can induce autocorrelation in the sequences of measurements taken.

Customer heterogeneity defines another reason that A/B testers continuously monitor in practice. Since the population of customers sampled can evolve over the course of a test, experimenters are concerned that a large effect seen on the types of customers who are sampled first will not persist once the broader population are exposed to the treatment. We derive a stronger form of always valid p-values and confidence intervals, which allow the user to stop at any time. When she does so, valid frequentist inference is obtained on the average treatment effect on the population that has been sampled up to that time.

3. Simultaneous inference across multiple A/B tests (Chapter 4). Often practitioners will run several A/B tests in parallel, because understanding of multiple treatment effects is necessary to make decisions. Standard procedures - the Bonferroni and the Benjamini-Hochberg corrections - can be applied to the t-test p-values and confidence intervals to obtain measures that control the family-wise error rate and the false discovery / false coverage rate respectively.

However, this approach presents two issues in A/B testing practice. The first is that it does not address the continuous monitoring problem. The second issue is that, while the BH-corrected confidence intervals control the false coverage rate among those tests where the null
hypothesis $\theta = 0$ is rejected, it is does not generally bound the FCR among those intervals that the practitioner will actually want to look at. In addition to those intervals, most A/B testers have some set of “guardrail” experiments that are always referenced, because understanding those treatment effects is critical to any decision-making.

We explain what assumptions are needed on the stopping time to say that Bonferroni and BH commute with always validity, so that simply applying these corrections to an arbitrary set of always valid p-values controls Type I error in the presence of continuous monitoring. Moreover, we present a modification of the BH-corrected confidence intervals together with a restriction on the stopping time, such that the false coverage rate on the intervals referenced is bounded, while being robust to which tests fall into the experimenter’s personal guardrail set.

4. Under-exploration when designing experiments (Chapter 5). The default analysis is concerned with obtaining inference on the treatment effect in a single A/B test, given a fixed data set. It does not motivate the experimenter to design her A/B test in a way that prioritizes collecting data that will be valuable for designing subsequent experiments.

We focus on two forms of under-exploration: low sample sizes and only running tests on the types of treatments that have been shown to work well in the past. We study the problem within a Bayesian context – rather than p-values and confidence intervals, we suppose that each experiment is analyzed through a posterior distribution for the treatment effect. This allows us to cast designing experiments as a sequential game between the employees that run A/B tests at a company. At each time, each employee wants a low sample size and a safe treatment type, in order to maximize a myopic reward associated with the inference derived in that single A/B test. However, she would prefer that other employees run long tests on poorly understood treatment types, as this provides information on typical effect sizes, which will be important for everyone when designing future experiments.

We propose that the company manages these externalities between employees by regulating how much data from past A/B tests each employee can access when she comes to design her next experiment. We offer a mechanism – the Minimum Learning Mechanism – for revealing experimental data to the employees, which perfectly aligns their incentives with that of the company, at least in a setting where they have no private information. In reality, employees tend to draw on their own knowledge when deciding how tests should be designed. We show that the MLM is robust to small levels of private information.
1.3 Related work

1.3.1 A/B testing practice

An extensive literature on A/B testing is emerging – particularly from industrial groups like Microsoft, Google, Facebook and others [2, 31, 40, 53]. The methods introduced in this thesis are inspired heavily by their work. Beyond the statistical focus of this thesis, these papers also provide insight on such practical topics as:

- Choosing improvement metrics [21, 22, 31].
- Diagnosing errors in the randomization or data collection [11].
- The computational infrastructure for running experiments [71].
- Fostering a corporate culture that prioritizes designing good experiments [75].

1.3.2 Experimental design

Of course, business optimization is hardly the first context in which practitioners have designed and analyzed experiments – randomized control trials have been considered the gold standard in many medical and social science applications for decades. For a review of the broad field of experiment design, we refer the reader to such textbooks as [11, 55].

The topics of this thesis connect with a number of specific experimental design challenges that have been addressed elsewhere:

- Choosing fixed-horizon sample sizes. This is one of the most fundamental problems in experimental design and much has been written on how it should be approached within a variety of domains [39, 63]. This thesis does not look to provide additional guidance here, focusing instead on facilitating dynamic sample sizes or on helping companies manage misaligned incentives when experimenters choose fixed-horizon sample sizes.

- Rival treatments. In Chapter 5, we consider how experimenters search through the space of available treatments, which is a central question in any experimental field. Again we do not look to provide additional guidance to the experimenters, focusing instead on how the company should manage their incentives.

- Sampling bias. Chapter 3 derives inference measures that handle the temporal noise that arises when the population of subjects entering the experiment evolves over time. While our goal is simply to account for any sampling bias in the analysis, an extensive survey design literature attacks this problem with protocols for how subjects should be recruited into the experiment – see [28] for a review.
On the other hand, some experimental design challenges are quite special to the A/B testing context. These relate particularly to the sequential nature of A/B testing:

- The observations within a single experiment arrive sequentially. As such this thesis draws extensively on prior work on sequential hypothesis testing \[46, 68\] and online FDR control \[76\].

- Related experiments are often run in sequence, which makes careful sequential experimental design important so as to ensure efficient long-term optimization \[11\].

### 1.3.3 Automation

In addition, the availability of real-time data means that A/B testing should be viewed within the broader field of online optimization. Where A/B testing focuses on estimating model parameters or testing null hypotheses, so as to provide decision makers with useful information, alternative approaches look to automate the entire decision-making process:

- **PAC-learning \[25, 36\] and pure exploration bandits \[13, 12\].** Rather than testing the null hypothesis of no qualitative difference between the treatment and control, these approaches just look to pick the better option with high probability when an effect exists.

- **Reinforcement learning and regret minimization bandits \[3, 12\].** The goal of an A/B test is to use observations collected during the experiment to derive inferences which will be used to optimize some decisions once the test is complete. Reinforcement learning techniques, however, use each customer observed both to explore and to exploit that learning. The decision on termination of an A/B test is often whether to offer the treatment to all customers thereafter or to revert to the control. For that use-case, rather than employing a single end time, regret minimization bandits dynamically allocate customers to either the treatment or the control based on the data collected so far.

- **Bayesian optimization \[49\].** Rather than running A/B tests on every treatment, these methods enable a dynamic search through the space of treatments available. Naturally such an approach is most useful when the space of possible treatments is both large and clearly definable up-front; e.g. when the treatments under investigation are the possible values for a set of hyper-parameters in a machine learning model \[61, 69\].

Automation is gaining industrial popularity in areas such as targeted advertising and viral marketing \[15\]. For product optimization, however, where the goals are less clearly defined and where the trade-offs are more subtle, greater human input in the decision-making process remains important.
1.3.4 Other statistical models

This thesis assumes that the data collected in an experiment is either binary or normally distributed. Further we focus on deriving inference from the sample difference of means between the treatment and control groups. In so doing, we tackle most use-cases of A/B testing. However, work has been done to enable more efficient analyses in specific contexts. Techniques include:

- *More complex parametric models for the responses* [18, 43, 26]. The goal here is larger power – this can be particularly important for binary data where success probabilities are very small or continuous data with heavy right tails.

- *Non-parametric p-values and confidence intervals* [3, 59, 1]. These can provide more exact Type I error control.

- *Alternative test statistics* [57, 74]. By focusing inference on statistics like medians or winsorized means, experimenters can introduce robustness without the need for additional assumptions. A downside is that results can be harder to interpret.

1.4 Acknowledgements

The research on always valid inference and on running multiple A/B tests in parallel, which are presented in Chapters 2 and 4 of this thesis respectively, are joint work with Ramesh Johari, Leo Pekelis and Pete Koomen. Much of this material can be found in the following papers: [35, 34]. Chapter 3 on adapting always valid inference measures to handle temporal noise is also joint work with Leo Pekelis. Further, all of this work was done in collaboration with the commercial A/B testing platform, Optimizely. I thank them both for their technical input – with particular thanks to Darwish Gani, Jimmy Jin, Conrad Lee, Ajith Mascarenhas, Fergal Reid and Hao Xia – and for their experimental data which we use to supplement the theory presented in this thesis.

Chapter 5 on incentivizing experimenters to design good A/B tests is joint work with Ramesh Johari, and it was conducted in collaboration with the EXP group, which manages A/B testing at Microsoft. I would especially like to thank Alex Deng, Leandra King, Ronny Kohavi, Jonathan Litz and Jiannan Lu at EXP for their technical contributions.

Finally I want to thank my dissertation committee – Ramesh Johari, Brad Efron, Art Owen, Tze Lai and Yonatan Gur – everyone in the statistics department at Stanford, and many others throughout the university for all of their feedback and advice.

---

1Department of Management Science and Engineering, Stanford University
2OpenDoor, Inc.
3Optimizely, Inc.
Chapter 2

Always valid inference

The industrial standard for analyzing A/B tests - frequentist p-values and confidence intervals derived from a 2-sample t-test of means - requires the experimenter to set a fixed-horizon sample size before any data is collected. In practice it is commonplace for the experimenter to continuously monitor these inference measures as data arrives (using a dashboard such as the one in Figure 2.1), and to adapt the sample size to what she observes. Very high false positive probabilities are obtained, well in excess of the nominal $\alpha$ that the user can tolerate. In fact, even with 10,000 samples (a sample size which is quite common in online A/B testing), the Type I error can easily increase fivefold. In this chapter, we present novel inference measures - always valid p-values and confidence intervals - that recover Type I error control when the experimenter sets her sample size dynamically. These measures have been in deployment at the commercial A/B testing platform Optimizely since 2015, where they have been used to analyze around a million A/B tests.

The reason that continuous monitoring is so pervasive is that the opportunity cost of longer experiments is large: an experimenter wants to detect true effects as quickly as possible, or else give up if it appears that no effect will be detected soon so that she can test something else. Most users further lack good prior understanding of their cost elasticity as well as the effect size they seek, frustrating attempts to optimize runtime in advance. For such users, the ability to trade off maximum detection with minimum run-time dynamically represents a crucial benefit of the availability of real-time data in modern A/B testing environments.

The purpose of our always valid inference measures is to maintain the simple interface of p-values and confidence intervals, but in an environment where users continuously monitor experiments, and where their priorities regarding run-time and detection are not known in advance. First, they control Type I error under any data-dependent rule the user might choose to stop the experiment (i.e. any stopping time for the data). Second, in a sense we make precise below, we show that if the user stops when our p-value drops below her personal $\alpha$, the resulting rule approximately obtains her optimal trade-off for run-time and detection, even with no advance knowledge of her preferences.
Figure 2.1: Dashboard from the A/B testing platform, Optimizely. Displayed is the “Chance to Beat Baseline” (CtBB) over the course of the experiment, which equals one minus the t-test p-value based on the data collected up to that time. The experimenter is supposed to set her fixed-horizon sample size in advance and read off the CtBB at that time. This display is taken from a real A/B test where the null hypothesis of no effect is known to be true. If the experiment is stopped at one of the times circled, as a continuously monitoring experimenter is likely to do, a false positive is made at the 5% significance level.

2.0.1 Contributions

1. Type I error control for “one-variation” experiments. While our ultimate goal is inference on the difference in means between a treatment and control experience, most of this chapter assumes a single stream of data drawn from a one-parameter exponential family to develop our main ideas. We consider sequences of p-values and confidence intervals for the parameter, which are defined in terms of the incoming data, and we ask which such processes control Type I error at any data-dependent stopping time the user might choose. Such processes are termed as always valid p-values and confidence intervals. We show how always valid measures can be constructed using sequential hypothesis tests \[71, 68, 46, 67\]. In particular, we identify a duality between always valid p-values and those sequential tests that do not accept $H_0$ in finite time, known as sequential tests of power one \[64\]. Under this duality, an infinitely patient employee who only cares about power should wait until the p-value falls below a significance level $\alpha$. Stopping at that time produces the same inferences as just running the sequential test, which means that all non-zero treatment effects are detected eventually.

2. Efficiently trading off power and run-time: The mixture sequential probability ratio test. Having controlled Type I error, we then ask: how can we efficiently trade off power and run-time? The challenge is that different employees prioritize power against sample size differently. We
aim to find always valid p-value processes that lead to a near-efficient tradeoff for any user.

It is evident that without restricting the class of user models we consider, no meaningful result is possible; the space of potential user preferences is vast (indeed some are unreasonable, e.g., preferring long experiments that do not detect effects). Instead, we focus on users that generally want short tests and good detection, modeled as follows: the user stops at either the first time the p-value process crosses $\alpha$, or at a fixed maximum failure time $M$, whichever comes first. A larger $M$ indicates greater patience, and a corresponding preference for detection of smaller effects. While this class does not capture all possible objective functions, it does allow us to capture what we consider to be the most interesting axis of user heterogeneity: how much they care about power versus how much they care about run-time.

Our main contribution is to demonstrate that always valid p-values derived from a particular class of sequential tests known as mixture sequential probability ratio tests (mSPRTs) achieve an efficient tradeoff between power and run-time for such users, to first order in $M$ as $M \to \infty$. The regime where $M$ is large corresponds to the data-rich regime; we consider this to be the interesting regime of user optimality for A/B testing platforms, where thousands of observations in an experiment will be commonplace. In this sense, our result demonstrates first order optimality in $M$ of a particular class of always valid p-values: those derived from mSPRTs.

The mSPRT involves stopping the first time a mixture of the likelihood ratio of alternative to null crosses a threshold, where the mixture is over potential values of the alternative. Among mSPRTs, power approaches one when $M \to \infty$; therefore we choose among this class by optimizing run-time. In particular, we find a particular choice of mSPRT minimizes run-time simultaneously for all sufficiently large $M$. The run-time minimizing choice of mSPRT has an appealingly simple form: e.g., when there is a Gaussian prior on effect sizes, the optimal mSPRT approximately matches the variance of the mixing distribution to the variance of the prior distribution of effect sizes.

3. Extension to A/B tests and deployment at Optimizely. We adapt the mSPRT from a single stream of data to a test on the difference in means between a treated sequence of observations and a control sequence. Where the single stream amounts to testing a simple null hypothesis that the parameter takes a single value, the challenge here is to test the one-dimensional null associated with a given value for the difference in means. We assume that measurements are either normally distributed or binary-valued. In the case of normal data, inference on the nuisance parameter (the location within this one-dimensional null) can be “modularized” from inference on the difference, allowing exact Type I error control. For binary data, approximate control is achieved for $\alpha$ small.

We conclude with a comparison of our methodology against classical fixed-horizon testing that
incorporates both theory and empirical data from Optimizely. We find that our approach can deliver equivalent power to fixed-horizon testing with a sublinear sample size; i.e. compared with the default t-testing approach, the experimenter obtains the same accuracy with fewer samples.

2.0.2 Related work

While our development of always valid inference measures is motivated by the precise demands of A/B testing, the desire to test hypotheses using data that arrives sequentially over time is far from new. Rather, sequential analysis is a mature field with roots dating back to [1], and sequential tests are widely used in areas such as pharmaceutical clinical trials. For its history, methodology and theory, we direct the reader to the encyclopedic resource of [3]. The implementation of existing sequential tests in online experimentation has gained recent popularity [54], especially in connection to multi-armed bandits [66], but the limitation here is that they function as black boxes. The type I error tolerance $\alpha$ and the desired trade-off between power and run-length are required as pre-specified inputs, and inference is attainable only at a single corresponding stopping time. On the other hand, by using sequential tests as building blocks to construct always valid p-values and confidence intervals, this paper obtains a real-time interface that can better handle users with heterogenous preferences.

More broadly, we note that there is substantial work in pure exploration bandits and PAC-learning that has a similar goal to this paper: to provide reliable inferences based on online data as quickly as possible. Under these methods, the allocation of treatments to incoming traffic is modified dynamically, in order to provide a decision at some minimal stopping time on which treatment is the best [13, 25]. In fact, these algorithms may produce answers noticeably faster than sequential hypothesis tests with comparable error calibration. However, they have a similar limitation to traditional sequential tests, in that trade-offs between accuracy and speed must be pre-determined; thus they do not provide an interface that addresses user heterogeneity.

2.1 Preliminaries

Most of this chapter focuses on one-variation experiments, before we extend the methodology to A/B tests with a treatment and a control. For these one-variation experiments, we assume a single stream of data $Y_1, Y_2, \ldots \sim F_\theta$ drawn from a single parameter exponential family with natural parameter $\theta \in \Theta \subset \mathbb{R}$: $f_\theta(y) = dF_\theta(y) \propto \exp(\theta y - \psi(\theta))$. Everywhere, $(\mathcal{F}_n)_{n=1}^\infty$ will denote the filtration generated by $(Y_n)_{n=1}^\infty$ and $P_\theta$ will denote the measure (on any space) induced under the parameter $\theta$. We focus on testing the simple null hypothesis that the parameter takes a given value, $H_0 : \theta = \theta_0$.

Assumption 5. These technical restrictions on the exponential family will prove useful:
1. For all $\theta$, $\psi''(\theta) > 0$.

2. There is some $p \geq 4$ where $Y_1$ has finite $p^{th}$ moment under any $\theta$.

2.1.1 Testing a simple null hypothesis

**Decision rules.** In general, a decision rule is a pair $(T, \delta)$, where $T$ is a (possibly infinite) stopping time for $(F_n)_{n=1}^{\infty}$ that denotes the terminal sample size. $\delta$ is a binary-valued, $(F_T)$-measurable random variable, where $\delta = 1$ indicates that $H_0$ is rejected; we require that $\delta = 0$ must hold a.s. if $T = \infty$. Note that we allow the possibility that the decision rule can be data-dependent, while we can recover the set of fixed-horizon decision rules by setting $T = n$ deterministically.

**Type I error.** Type I error is the probability of erroneous rejection under the null, i.e., $P_{\theta_0}(\delta = 1)$. We assume that the user wants to bound Type I error at level $\alpha \in (0, 1)$.

**Sequential tests.** Given $\alpha$, we typically consider a family of decision rules parameterized by $\alpha$. Formally, a sequential test is a family of decision rules $(T(\alpha), \delta(\alpha))$ indexed by $\alpha \in (0, 1)$ such that:

1. The decision rules are nested: $T(\alpha)$ is a.s. nonincreasing in $\alpha$, and $\delta(\alpha)$ is a.s. nondecreasing in $\alpha$.

2. For each $\alpha$, the Type I error is bounded by $\alpha$: $P_{\theta_0}(\delta = 0) \leq \alpha$.

Note that the fixed-horizon t-tests are special cases where the stopping time takes a single deterministic value for all values of $\alpha$. While they are not the unique tests that have that deterministic stopping time, among any such tests they have uniformly maximum power for testing the simple null against any alternative (see Subsection 1.1.2).

**Fixed horizon p-values and confidence intervals.** The standard fixed-horizon inference measures are derived from the t-test. For a given $\theta_0$, the p-value is the least $\alpha$ such that $H_0 : \theta = \theta_0$ is rejected, while the $(1 - \alpha)$-level confidence interval is the range of values $\theta_0$ for which the test at level $\alpha$ fails to reject.

Thresholding the p-value at level $\alpha$ is equivalent to implementing the t-test. Thus the Type I error control of the t-test may be rewritten as an equivalent Type I error bound on the p-value:

$$\forall s \in [0, 1], \ P_0(p_n \leq s) \leq s. \quad (2.1)$$

Similarly, the confidence intervals satisfy the following coverage bound:

$$\forall \theta \in \Theta, \ P_0(\theta \in \text{CI}_n) \geq 1 - \alpha. \quad (2.2)$$

In general, we refer to any sequence of $[0,1]$-valued, $(F_n)$-measurable random variables $p_n$ indexed by all possible fixed-horizons, which satisfies (2.1), as a fixed horizon p-value process. A $(1 - \alpha)$-level fixed horizon confidence interval is any $(F_n)$-measurable random subset of $\Theta$ that satisfies (2.2).
2.1.2 Extension to A/B tests

After developing always valid inference measures for a single stream of data, we will adapt these for inference on the difference in means between a treatment and a control. In that case, we assume two streams: \( Y_{t1}^1, Y_{t2}^1, \ldots \) and \( Y_{c1}^1, Y_{c2}^1, \ldots \) and further that the observations arrive as pairs \( W_n = (Y_{tn}^t, Y_{cn}^c) \) (see Assumption 1). \( (\mathcal{F}_n)_{n=1}^\infty \) is now the filtration generated by the pairs \( (W_n)_{n=1}^\infty \). An arbitrary stopping time with respect to this filtration, \( T \) denotes the terminal sample size for each group.

Each stream of data is IID, and we consider two options for the distribution of each observation:

- either \( Y_{tk}^t \sim \text{Bernoulli}(\mu_t), Y_{ck}^c \sim \text{Bernoulli}(\mu_c) \),
- or \( Y_{tk}^t \sim \mathcal{N}(\mu_t, \sigma^2), Y_{ck}^c \sim \mathcal{N}(\mu_c, \sigma^2) \)

with \( \sigma^2 \) known. The target of inference is the difference in means, \( \theta = \mu_t - \mu_c \). A fixed-horizon p-value for testing the null that \( \theta = 0 \) is any adapted sequence \( p_n \), which satisfies the Type I error bound (2.1) uniformly over all common values \( \mu_t = \mu_c \). A fixed-horizon confidence interval for \( \theta \) is any adapted sequence \( \text{CI}_n \) that satisfies (2.2) uniformly over any \( \mu_t \) and \( \mu_c \) whose difference is \( \theta \).

2.2 Always valid inference

Our goal is to let the user stop the test whenever she wants, in order to trade off power with run-time as she sees fit; the p-value obtained should control Type I error. Our first contribution is the definition of always valid p-values as those processes that achieve this control in the case of one-variation experiments.

**Definition 1.** A sequence of fixed horizon p-values \( (p_n) \) is an always valid p-value process if given any (possibly infinite) stopping time \( T \) with respect to \( (\mathcal{F}_n) \), there holds:

\[
\forall s \in [0, 1], \; \mathbb{P}_{\theta_0}(p_T \leq s) \leq s. \tag{2.3}
\]

The following theorem demonstrates that always valid p-values are in a natural correspondence with sequential tests.

**Theorem 1.** 1. Let \( (T(\alpha), \delta(\alpha)) \) be a sequential test. Then

\[
p_n = \inf \{ \alpha : T(\alpha) \leq n, \delta(\alpha) = 1 \}
\]

defines an always valid p-value process.

2. For any always valid p-value process \( (p_n)_{n=1}^\infty \), a sequential test \( (\tilde{T}(\alpha), \tilde{\delta}(\alpha)) \) is obtained from
(\(p_n\))\(\infty_{n=1}\) as follows:

\[
\hat{T}(\alpha) = \inf\{n : p_n \leq \alpha\}; \quad (2.4)
\]

\[
\tilde{\delta}(\alpha) = 1\{\hat{T}(\alpha) < \infty\}. \quad (2.5)
\]

If (\(p_n\))\(\infty_{n=1}\) was derived as in part (1) and \(T = \infty\) whenever \(\delta = 0\), then \((\hat{T}(\alpha), \tilde{\delta}(\alpha)) = (T(\alpha), \delta(\alpha))\).

**Proof.** For the first result, nestedness implies the following identity for any \(s \in [0, 1], n \geq 1, \varepsilon > 0\):

\[
\{p_n \leq s\} \subset \{T(s + \varepsilon) \leq n, \delta(s + \varepsilon) = 1\} \subset \{\delta(s + \varepsilon) = 1\}. 
\]

Therefore:

\[
P_{\theta_0}(p_T \leq s) \leq P_{\theta_0}(\cup_n \{p_n \leq s\}) \leq P_{\theta_0}(\delta(s + \varepsilon) = 1) \leq s + \varepsilon.
\]

The result follows on letting \(\varepsilon \to 0\). For the converse, it is immediate from the definition that the tests are nested and \(\delta(\alpha) = 0\) whenever \(T(\alpha) = \infty\). For any \(\varepsilon > 0\)

\[
P_{\theta_0}(\delta(\alpha) = 1) = P_{\theta_0}(T(\alpha) < \infty) \leq P_{\theta_0}(p_T(\alpha) \leq \alpha + \varepsilon) \leq \alpha + \varepsilon
\]

where the last inequality follows from the definition of always validity. Again the result follows on letting \(\varepsilon \to 0\). 

The p-value defined in part (1) of the theorem is not the unique always valid p-value associated with that family of sequential tests (i.e., for which part (2) holds). However, among such always valid p-values it is a.s. minimal at every \(n\), resulting from the fact that it is a.s. monotonically non-increasing in \(n\). Thus we have a one-to-one correspondence between monotone always valid p-value processes and families of sequential tests that do not give up for failure; i.e., where \(\delta = 0\) implies \(T = \infty\). These processes can be seen as the natural representation of those sequential tests in a streaming p-value format.

**The new user interaction model.** The time \(\hat{T}(\alpha)\) represents the natural stopping time of a hypothetical user who incurs no opportunity cost from longer experiments. By thresholding the p-value at \(\alpha\) at that time, she recovers the underlying sequential test and is able to reject \(H_0\) whenever \(\delta = 0\). Of course, a real user cannot wait forever, so she must stop the test and threshold the p-value at some potentially earlier, a.s. finite stopping time. In so doing, she sacrifices some detection power. This trade-off for the user between power and average run-time is a central concern of our proposed design, and is studied in more detail in Section 2.3.

**Confidence intervals.** Always valid CIs are defined analogously and may be constructed from always valid p-values just as in the fixed-horizon context. Proposition 4 follows immediately from the definitions.
Definition 2. A sequence of fixed-horizon \((1 - \alpha)\)-level confidence intervals \((\text{CI}_n)\) is an always valid \(1 - \alpha\)-level confidence interval process if, given any stopping time \(T\) with respect to \((\mathcal{F}_n)\), there holds:

\[
\forall \theta \in \Theta, \ p_\theta(\theta \in \text{CI}_T) \geq 1 - \alpha.
\]  

(2.6)

Proposition 1. Suppose that, for each \(\tilde{\theta} \in \Theta\), \((p^\tilde{\theta}_n)\) is an always valid p-value process for the test of \(\theta = \tilde{\theta}\). Then

\[
\text{CI}_n = \{\theta : p^\theta_n > \alpha\}
\]

is an always valid \((1 - \alpha)\)-level CI process.

2.3 Optimal p-value processes

As noted in the preceding section, a user who continuously monitors her experiments wants to make a dynamic tradeoff between two objectives: detecting true effects with maximum probability and minimizing the typical length of experiments. A significant challenge for the platform is that always valid p-values must be designed without prior knowledge of the user’s preferences. We are led, therefore, to consider the following design problem: how should always valid p-values be designed to lead users to a near-optimal tradeoff between power and run-length, without access to the user’s preferences in advance?

In this section we continue to focus on one-variation experiments, wherein we introduce a natural model of user behavior that encodes a tradeoff between power and run-length. Then we consider always valid p-value processes derived from a particular family of sequential tests, the mixture sequential probability ratio tests (mSPRT), proving that any mSPRT provides near-optimal decision making for all of these users simultaneously. Specifically, for any such user, it is not possible to improve detection uniformly at every effect size in the limit as \(\alpha \to 0\). Moreover, we optimize among mSPRTs for average-case performance when prior knowledge on the effect size sought is available, and we establish an improvement over fixed-horizon testing in that case. The proofs for this section are given in Appendix A.

2.3.1 The user model

Of course, any specific user’s preferences will be highly nuanced. In our analysis, for technical simplicity we consider the following user model: we assume users can be characterized by a parameter \(M\) representing the maximum number of observations that the user is willing to wait, together with their Type I error tolerance \(\alpha\). Given an always valid p-value process, we consider a simple model of user behavior: users stop at either the first time the p-value drops below \(\alpha\) (in which case they reject the null), or at time \(M\) (in which case they do not reject the null unless the p-value at time \(M\) is also below \(\alpha\)), whichever occurs first.
We refer to such a user as a \((M, \alpha)\) user. In the remainder of the section, our goal will be to make near-optimal tradeoffs between power and run-length for \((M, \alpha)\) users in the limit where \(\alpha \to 0\), without prior knowledge of their preferences. We note the following technical remark: if the always valid p-values shown to a user are derived from a sequential test \((T(\alpha), \delta(\alpha))\), then the \((M, \alpha)\) user’s decision rule is given by \(T(M, \alpha) \triangleq \min\{T(\alpha), M\}\), and \(\delta(M, \alpha) = 1\) if and only if \(T(\alpha) \leq M\).

### 2.3.2 The mSPRT

The always valid p-values we employ are derived from a particular family of sequential tests: the mixture sequential probability ratio test (mSPRT) \([65]\). This family is parameterized by a mixing distribution \(H\) over \(\Theta\), which we restrict to have everywhere continuous and positive derivative. Given an observed sample average \(s_n\) up to time \(n\), the likelihood ratio of \(\theta\) against \(\theta_0\) is \((f_\theta(s_n)/f_{\theta_0}(s_n))^n\). Thus we define the mixture likelihood ratio with respect to \(H\) as

\[
\Lambda_n^H(s_n) = \int_{\Theta} \left( \frac{f_\theta(s_n)}{f_{\theta_0}(s_n)} \right)^n dH(\theta).
\]

The mSPRT is then defined by:

\[
T^H(\alpha) = \inf\{n : \Lambda_n^H(s_n) \geq \alpha^{-1}\}; \quad \text{and} \quad \delta^H(\alpha) = 1 \{T^H(\alpha) < \infty\},
\]

where \(S_n = \sum_{i=1}^n Y_i\). The choice of threshold \(\alpha^{-1}\) on the likelihood ratio ensures Type I error is controlled at level \(\alpha\), via standard martingale techniques \([68]\). Intuitively, \(\Lambda_n^H(S_n)\) represents the evidence against \(H_0\) in favor of a mixture of alternative hypotheses, based on the first \(n\) observations. The test rejects \(H_0\) if the accumulated evidence ever becomes large enough.

Our first motivation for considering mSPRTs is that they are tests of power one \([64]\): for all \(\alpha\) and \(\theta \neq \theta_0\), there holds:

\[
P_\theta(T(\alpha) < \infty, \delta(\alpha) = 1) = 1.
\]

In other words, for the hypothetical user who can wait forever, any mSPRT delivers power one for any alternative. We also note here that in a Bayesian setting where the user’s cost function takes a suitable linear form and the stopping time and \(\alpha\) are fixed optimally, mSPRTs are known to be optimal in terms of Bayes risk in a suitable limit \([46]\).

For later reference, we note the following result due to \([62]\): for any mixing distribution \(H\), as \(\alpha \to 0\),

\[
T^H(\alpha)/\log(1/\alpha) \to I(\theta, \theta_0)^{-1} := \{(\theta - \theta_0)\psi'(|\theta|) - (\psi(|\theta|) - \psi(|\theta_0|))\}^{-1}
\]

holds in probability and in \(L^2\), where \(\psi(\theta)\) is the log-partition function for the family \(F_\theta\). This

---

\(^1\)This limit maps to the setting where \(\alpha \to 0\), where we also focus attention in our results.
result characterizes the run-length of the mSPRT in the small $\alpha$ limit, and plays a key role in our subsequent study of efficiency.

### 2.3.3 Efficiency in the $\alpha \to 0$ limit

We now formalize our study of the power and run-length tradeoff for an $(M, \alpha)$ user. To proceed, suppose that the way this user makes decisions is to use a rule $(T^*, \delta^*)$ that controls Type I error at level $\alpha$ (i.e., $\mathbb{P}_{\theta_0}(\delta^* = 1) \leq \alpha$), and satisfies the constraint that $T^* \leq M$ a.s. We refer to such decision rules as \textit{feasible} decision rules for an $(M, \alpha)$ user.

We first map the two objectives of the user to formal quantities of interest. First, an $(M, \alpha)$ user will want to choose her decision rule to maximize the \textit{power profile} $\nu(\theta) = \mathbb{P}_{\theta_\neq \theta_0}(\delta^* = 1)$ over $\theta \neq \theta_0$.

Second, she will want to minimize the \textit{relative run-length profile} of relative run-lengths measured against the maximum available to her, $\rho(\theta) = \mathbb{E}_{\theta}(T^*)/M$.

**Perfect efficiency** would entail $\rho(\theta) = 0$ and $\nu(\theta) = 1$ for all $\theta \neq \theta_0$. Of course, perfect efficiency is generally unattainable. In this section we study the best achievable performance such a user can hope for, in the limit where $\alpha \to 0$.

Our analysis depends on the characterization of run-length of the mSPRT in (2.10). The consequence is that if we produce always valid p-values using the mSPRT, then in the $\alpha \to 0$ the study of efficiency divides into three cases depending on the relative values of $M$ and $\log(1/\alpha)$.

**“Aggressive” users: $M \gg \log(1/\alpha)$**. Users in this regime are aggressive; $\alpha$ is large relative to the maximum run-length they have chosen. In this regime, any mSPRT asymptotically recovers perfect efficiency in the limit where $\alpha$ small. Intuitively, because the user is willing to wait a substantially longer time than $\log(1/\alpha)$, a sublinear fraction of their maximum run-length is used by the mSPRT; since the mSPRT is a test of power 1, this means the user receives power near 100% in return for a near-zero run-length profile. The proof of the following result follows immediately from (2.10).

**Proposition 2.** Given any mixture $H$, let $\rho(\theta)$ and $\nu(\theta)$ be the relative run-length and power profiles, respectively, associated with $(T^H(M, \alpha), \delta^H(M, \alpha))$. If $\alpha \to 0$ and $M \to \infty$ such that $M/\log(1/\alpha) \to \infty$, we have $\rho(\theta) \to 0$ and $\nu(\theta) \to 1$ at each $\theta \neq \theta_0$.

**“Conservative” users: $M \ll \log(1/\alpha)$**. Users in this regime are conservative; $\alpha$ is small relative to the maximum run-length they have chosen. In this case, experimentation is not productive: the user is unwilling to wait long enough to detect any effects. Thus any mSPRT trivially performs as well as any feasible decision rule for the user.

**Proposition 3.** For each $(M, \alpha)$, let $(T^*(M, \alpha), \delta^*(M, \alpha))$ be any test such that $T^* \leq M$ a.s. and $\nu^*(\theta_0) \leq \alpha$, where $\nu^*$ is the associated power profile. Then if $\alpha \to 0, M \to \infty$ such that $M/\log(\alpha^{-1}) \to 0$, we have $\nu^*(\theta) \to 0$ for each $\theta$. 
“Goldilocks” users: $M \sim \log(1/\alpha)$. This is the interesting case, where experimentation is worthwhile but statistical analysis is non-trivial. To proceed we require an additional definition. For a family of sequential tests, we define the worst-case efficiency over $\theta \neq \theta_0$ for a user with parameters $(M, \alpha)$ as the relative efficiency of the truncated test they obtain by minimizing the relative run-length everywhere, compared with any other test that offers at least as good power everywhere. This is formalized as follows.

**Definition 3.** Given a sequential test $(T(\alpha), \delta(\alpha))$, let $\rho(\theta; \alpha, M)$ and $\nu(\theta; \alpha, M)$ be the relative run-length and power profiles associated with $(T(M, \alpha), \delta(M, \alpha))$. The relative efficiency of this test at $(M, \alpha)$ is

$$\phi(M, \alpha) = \inf_{(T^*, \delta^*)} \inf_{\theta \neq \theta_0} \frac{\rho^*(\theta)}{\rho(\theta; \alpha, M)}$$

where the infimum is taken over tests $(T^*, \delta^*)$, whose run-length and power profiles are denoted by $\rho^*$ and $\nu^*$, for which the following conditions all hold:

1. $T^* \leq M$ a.s.
2. $\mathbb{P}_{\theta_0}(\delta^* = 1) \leq \alpha$.
3. For all $\theta \neq \theta_0$, we have $\nu^*(\theta) \geq \nu(\theta; \alpha, M)$.

Our main result demonstrates that in the regime where $M \sim \log(1/\alpha)$, any mSPRT has relative efficiency approaching unity when $\alpha \to 0$.

**Theorem 2.** Suppose $\psi''$ is absolutely continuous and there is an open interval around $\theta_0$ where $\psi''(\theta) < \infty$. Given any $H$, let $\phi(M, \alpha)$ be the efficiency of the mSPRT $(T^H(\alpha), \delta^H(\alpha))$. If $\alpha \to 0, M \to \infty$ such that $M = O(\log(\alpha^{-1}))$, we have $\phi(M, \alpha) \to 1$.

### 2.3.4 Run-length optimization

Now we go one step further and optimize our choice of mSPRT. As noted in the preceding section, it is users with $M \sim \log(1/\alpha)$ where there is a meaningful tradeoff between run-length and power, and for these users any mSPRT is asymptotically efficient as $\alpha \to 0$. That is, no mSPRT can offer a uniform improvement at every $\theta$ in that limit. However, the choice of $H$ governs a tradeoff between performance at different effect sizes. We now consider a Bayesian setting where we have a prior $\theta \sim G$ under $H_1$. The following theorem gives the mixing distribution $H$ that minimizes the relative run-length on average over this prior for a “goldilocks” user with parameters $(M, \alpha)$.

**Theorem 3.** Suppose $G$ is absolutely continuous with respect to Lebesgue measure, and let $H_\gamma$ be a parametric family, $\gamma \in \Gamma$, with density $h_\gamma$ positive and continuous on $\Theta$. Let $\rho_\gamma(\theta)$ be the profile of relative run-lengths associated with $(T^{H_\gamma}(M, \alpha), \delta^{H_\gamma}(M, \alpha))$. Then up to $o(1)$ terms as
\[ \alpha \to 0, M = O(\log(1/\alpha)) \], the average relative run-length \( \rho_r(M, \alpha) = \mathbb{E}_{\theta \sim G}\{\rho_r(\theta)\} \) is minimized by any \( \gamma^* \) such that:
\[
\gamma^* \in \arg \min_{\gamma \in \Gamma} -\mathbb{E}_{\theta \sim G}\left\{1_{A(M, \alpha)} I(\theta, \theta_0)^{-1} \log h_\gamma(\theta)\right\}
\] (2.11)
where \( A(M, \alpha) = \{\theta : I(\theta, \theta_0) \geq \log(1/\alpha)/M\} \).

If \( h_\gamma(\theta) = q(\gamma_1, \gamma_2)e^{\gamma_1 \theta - \gamma_2 \psi(\theta)} \), a conjugate prior for \( f_0 \), then \( \gamma^* \) solves:
\[
\frac{\partial q(\gamma_1, \gamma_2)/\partial \gamma_1}{\partial q(\gamma_1, \gamma_2)/\partial \gamma_2} = \frac{\mathbb{E}_{\theta \sim G}1_A \theta I(\theta, \theta_0)^{-1}}{\mathbb{E}_{\theta \sim G}1_A \psi(\theta) I(\theta, \theta_0)^{-1}}.
\]

Optimizing for \( h_\gamma \) does not impact first order terms of \( \mathbb{E}_{\theta \sim G}\{\rho_r(\theta)\} \) in this limiting regime.

Heuristically, the mSPRT rejects \( H_0 \) when there is sufficient evidence in favor of any \( H_1 : \theta \neq \theta_0 \), weighted by the distribution \( H \) over alternatives. Thus we might expect optimal sampling efficiency when \( H \) is \textit{matched} to the prior. If we specialize to normal data, \( f_0(x) = \phi(x - \theta) \), a centered normal prior, \( G(\theta) = \Phi(\frac{\theta}{\tau}) \), and consider normal mixtures, \( h_\gamma(\theta) = \frac{1}{\gamma} \phi(\frac{\theta}{\tau}) \), this intuition is mostly accurate. In that case, the optimal choice of mixing variance becomes:
\[
\gamma^2 = \tau^2 \frac{\Phi(-b)}{\phi(b) - \Phi(-b)}
\] (2.12)
for \( b = \left(\frac{\log\alpha^{-1}}{M\tau^2}\right)^{1/2} \). In fact, this is equal to the prior variance multiplied by a factor correcting for anticipated truncation that depends on \( M \). Sampling efficiency is improved by weighting towards larger effects when few samples are available and smaller effects where there is ample data.

Nominally, this truncation correction presents a challenge: it appears the optimal mixture distribution depends crucially on \( M \). However, as we now demonstrate both empirically and through simulation, the choice of mixing distribution is quite robust with respect to variation in \( M \). We use a sample of 40,000 past A/B tests run through Optimizely to construct a prior for \( \theta \) under \( H_1 \) for an arbitrary experiment run on that platform (more detail on how these experiments have been sampled will be given in Subsection \[\text{XX.XX}\]). We find that, given \( \alpha \) and \( M \), the choice of \( H \) given by (2.12) provides a substantial reduction in the average run time over highly suboptimal choices of \( \gamma^2 \) (optimizing \( H \) is worthwhile). However, a single choice of \( \gamma^2 \) is sufficient to achieve near-optimal performance over the range of parameter values seen in practice (see Table \[\text{XX.XX}\]).

### 2.3.5 Improvement over fixed-horizon

This Bayesian formulation also allows us to compare mSPRT p-values with fixed-horizon testing. In general, the fixed-horizon sample size must be chosen in reference to the effect sizes where detection is needed; now we can suppose that it is calibrated to have good average power over the prior \( G \).

For convenience, we focus on the normal case \( G = N(0, \tau^2), H = N(0, \gamma^2) \). We consider two rival tests for an arbitrary goldilocks user: the mSPRT truncated at \( M \) and the fixed-horizon test
### Table 2.1: Simulation for optimal matching of mixing distribution over various choices of Type I error thresholds, $\alpha$, and maximum sample size $n$. We define $r = \gamma^2/\tau^2$. The Predicted Ratio shows the $\gamma^2$ estimate from (2.12). The next 6 columns are the proportion of times the fixed $r$ values were empirically found to have lowest average expected run time, $E_G[T]$. This was done by sampling 200 experiment sample paths with effect sizes drawn from $G \sim N(0, 1)$, and running an mSPRT with various $r$s. The $r$ value with lowest average run time over sample paths was counted to have lowest expected run time for that replication. We perform $B = 25$ replications of each parameter combination. The most winning $r$ values coincide well with the estimate in (2.12). The final column shows the average percent difference between the winning $r$’s estimate of $E_G[T]$ and the runner-up. A factor of 10 misspecification around the optimal $r$ value results in a less than 10% increase in average runtime. Comparatively, there is over a factor of 2 difference between the winning and worst $r$ estimates.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$n$</th>
<th>Predicted Ratio</th>
<th>% average runtime for $r =$</th>
<th>Misspec. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>10,000</td>
<td>0.047</td>
<td>0 0 36 64 0 0</td>
<td>2.2</td>
</tr>
<tr>
<td>0.010</td>
<td>10,000</td>
<td>0.039</td>
<td>0 0 40 56 4 0</td>
<td>2.5</td>
</tr>
<tr>
<td>0.100</td>
<td>10,000</td>
<td>0.027</td>
<td>0 0 52 40 8 0</td>
<td>4.6</td>
</tr>
<tr>
<td>0.001</td>
<td>100,000</td>
<td>0.015</td>
<td>0 24 76 0 0 0</td>
<td>4.3</td>
</tr>
<tr>
<td>0.010</td>
<td>100,000</td>
<td>0.012</td>
<td>0 48 56 0 0 0</td>
<td>4.6</td>
</tr>
<tr>
<td>0.100</td>
<td>100,000</td>
<td>0.009</td>
<td>0 48 48 4 0 0</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Proposition 4. Let $\rho(M, \alpha)$ be the average relative run-length of the truncated mSPRT and let $\rho_f(M, \alpha)$ be the relative run-length of the calibrated fixed-horizon test (i.e. the fixed sample size divided by $M$). If $\alpha \to 0, M \to \infty$, such that $M = O(\log(\alpha^{-1}))$, then $\rho(M, \alpha)/\rho_f(M, \alpha) \to 0$.

### 2.4 Extension to A/B tests

Finally we are ready to turn our mSPRT inference measures for one-variation experiments into always valid p-values and confidence intervals for A/B tests. Since January 2015, these two-stream analogues have been implemented at Optimizely. Combining data from this deployment with numerical simulations, we establish that the improvement over fixed-horizon testing, which was described in the previous section for single-stream normal data, carries over to typical A/B tests.

For A/B tests, observations arrive as pairs $W_n = (Y_t^n, Y_c^n)$. Responses within each group are assumed to be either binary-valued or normally distributed with known variance, and the target of inference is the difference in means, $\theta = \mu_t - \mu_c$. In this section, we describe how the mSPRT may be adapted to produce always valid p-values for testing the null that $\theta = 0$. Always valid confidence intervals for $\theta$ may be obtained similarly.
2.4.1 Normal responses

For normal data, we view \((W_n)_{n=1}^{\infty}\) as a single stream of IID data from a bivariate distribution parameterized by the pair \((\theta, \mu)\), where \(\mu = (\mu_t + \mu_c)/2\). It is straightforward to show that, after fixing \(\mu = \mu^*\) arbitrarily, this distribution corresponds to the one-parameter exponential family \(f_{\theta}(w) \propto \phi\left(\frac{y-x-\theta}{\sqrt{\mu_c^*(1-\mu_c^*)+\mu_t^*(1-\mu_t^*)}}\right)\), where \(w = (x, y)\). Hence we may implement the mSPRT based on \(f_{\theta}\), i.e. we threshold the mixture likelihood ratio given in (2.7) with \(s_n = \frac{1}{n} \sum_{i=1}^{n} w_i\). For any \(\mu^*\), this mSPRT controls Type I error for testing the simple null \(H_0: \theta = 0, \mu = \mu^*\) against \(H_1: \theta \neq 0, \mu = \mu^*\), and so the p-value derived from this mSPRT is always valid for testing the composite null hypothesis.

In Appendix B, we show that it satisfies natural analogues of the single stream optimality results described in Section 2.3.

2.4.2 Binary responses

Unfortunately for binary data, the distribution of \(W_n\) does not reduce to a one-parameter exponential family. Nonetheless we can set \(\mu\) to be the average of the two success probabilities as before, and we denote the density of \(W_1\) by \(f_{\theta, \mu}(w)\). Then, for any \(\theta\) and \(\mu^*\), in the limit as \(n \to \infty\), the likelihood ratio against the pair \((\theta_0, \mu^*)\) in favor of \((\theta, \mu^*)\) approaches \(\tilde{f}_{\theta}(s_n)/\tilde{f}_{\theta_0}(s_n)\), where (with \(w = (x, y)\)):

\[
\tilde{f}_{\theta}(w) = \phi\left(\frac{y-x-\theta}{\sqrt{\mu_c^*(1-\mu_c^*)+\mu_t^*(1-\mu_t^*)}}\right), \quad \mu_c^* = \mu^* - \theta/2, \quad \mu_t^* = \mu^* + \theta/2.
\]

We compute the mSPRT p-values based on this density using the sample means in each stream as direct plug-in estimates for \(\mu_c^*\) and \(\mu_t^*\). If \(\alpha\) is moderate, the mSPRT terminates with high probability before this asymptotic distribution becomes accurate, so Type I error is not controlled. However, for \(\alpha\) small, simulation shows that these p-values are approximately always valid.

2.4.3 Real-world improvement

Proposition 4 established that for a single stream of normal data and a normal prior, the mSPRT has lower average relative run-length than the fixed-horizon test with the same average power. Now we consider two-stream binary-valued data. 10 000 experiments were randomly sampled from those binary experiments run on a large-scale commercial A/B testing platform in early 2015. Customers of the platform at that time could purchase subscriptions at one of four tiers: Bronze, Silver, Gold or Platinum. Within each tier, the observed data was consistent with a normal distribution of true effect sizes \(\theta/\sqrt{\mu_c^*(1-\mu_c^*)+\mu_t^*(1-\mu_t^*)}\) across experiments (additional variation between tiers existed because customers in higher tiers tended to be further along in optimizing their website and so were typically chasing smaller effects). Separately for each tier, we fitted a centered normal prior \(G\) with unknown variance for the effect under the alternative hypothesis by applying appropriate shrinkage via James-Stein estimation to the distribution of observed effect sizes [32].
We consider a range of values for the power $\beta$, and taking $\alpha = 0.1$ we identify the value of $M$ where the $(M, \alpha)$-user obtains average power $\beta$ over the fitted prior $G$ using the mSPRT p-values. This is compared against the fixed-horizon test that achieves the same average power $\beta$. The red curve in Figure 2.2 shows that with high probability, the mSPRT achieves a shorter run-length than fixed-horizon, provided $\beta$ is reasonably large. Performance is not optimal for a very conservative user, whose parameters $(M, \alpha)$ make $\beta$ small.

To be additionally generous to fixed-horizon testing, we consider the possibility that the user has additional side information about the true effect size beyond the prior used to construct the mSPRT p-values, and this allows her to better optimize the fixed-horizon test. The black curves in Figure 2.2 show that she must be able to estimate the effect very accurately if she is to achieve shorter run-lengths with fixed-horizon testing (a relative error in this estimate below 50% is rarely achievable in practice).

## 2.5 Conclusion

In this chapter, we have derived always valid p-values and confidence intervals that leverage sequentially arriving data within a single A/B test. These allow a user with fairly unconstrained preferences to continuously monitor her experiments, obtaining inferences at any time that optimally trade-off power and run-time for her needs.

### 2.5.1 Other data models

For one-variation experiments, we can handle any exponential family model for the measurements made on each customer. However, for typical A/B tests, we are critically reliant on the assumption...
of normal or binary data. The rotational symmetry of the bivariate Gaussian distribution gives us exact Type I error control for inference on the difference of two normal means, because estimation of this difference can be modularized from estimation of the nuisance parameter corresponding to their average. Approximate Type I error control is then feasible for binary data because the sum of independent Bernoulli trials very closely resembles the sum of independent Gaussians. However, in some contexts, A/B testing data can be heavily right-tailed [26]. Various methods for modeling this skewed data are used in practice, and construction of always valid inference measures under these models, or perhaps even non-parametric models, would be a useful extension to this work.

2.5.2 Bandit allocation policies

This chapter has focused on the case where customers are assigned to the treatment or the control alternately. In fact if each customer is randomized between the two experiences by an independent coin flip, as is standard in industrial practice, the two sequences $\left( Y_n^t \right)_{n=1}^{\infty}$ and $\left( Y_n^c \right)_{n=1}^{\infty}$ remain IID with the same binomial or normal distribution; the only difference is that when the test is terminated, there will be an unequal number of observations between the two groups. More generally any bandit allocation policy, where the treatment assignment for the $(n+1)^{th}$ customer depends on the measurements made on the first $n$ customers, will preserve this IID structure. It is straightforward to show that the same p-values and confidence intervals will control Type I error under arbitrary continuous monitoring and any such data collection scheme.

However, with bandit policies, there are many important optimality questions that are beyond the scope of this thesis. For instance, which allocation policy enables the mSPRT to achieve significance most quickly? How are the power and relative run-time profiles of always valid measures impacted if the allocation is optimized to another objective such as regret minimization?

2.5.3 Temporal noise

Finally we note that by allowing the user to stop her A/B tests at any time, we heighten our dependence on the assumption that measurements on successive customers are independent. In particular, seasonal effects can cause a large number of customers early in the experiment to take some common unusual action. If the employee stops the test at that time, our p-values may detect a treatment effect that will not persist over longer timescales. Adapting our inference measures to handle that risk is the focus of the next chapter of this thesis.
Chapter 3

Always valid inference under temporal noise

Both the traditional t-testing approach and the always valid inference measures derived in the previous chapter assume that every customer in an A/B test is interchangeable, and so their responses under either the treatment or the control can be modeled as independent, identically distributed random variables. However, from the experimental data in many real-life A/B tests, it is clear when this assumption is violated (see Figure 3.1). Rather the observations exhibit autocorrelation, because customers have latent characteristics that impact their behavior and similar customers tend to enter the experiment at similar times. The result is that the treatment has some average effect on the population of customers that arrives first; it has a different average effect on the customers who arrive later; and the always valid inference measures computed as in Chapter 2 do not control Type I error for inference on either.

For this chapter, our goal is frequentist inference on the average treatment effect for the duration of the experiment, defined as the mean effect on the population of customers that uses the service from the beginning of the test until whenever the experimenter chooses to terminate it. We obtain an extension to the always valid p-values and confidence intervals derived in Chapter 2 which achieve this goal in the presence of autocorrelation. The experimenter can continuously monitor these measures and stop dynamically, while controlling Type I error, so as to manage power, run-time and the population on which the ATE is defined.

3.0.1 Motivation

The existence of temporal noise, beyond what can be explained by the IID model, is well-known in A/B testing practice. Sources of autocorrelation include:
Figure 3.1: Dashboard from the A/B testing platform, Optimizely. For three different treatments, the sample difference in means between the treatment and control over the course of an experiment is displayed. A seasonal effect with a period of about one week seems to have affected all three treatments.

- **Power users** [58]. There is a size-biasing effect that the customers who use the service most frequently are the first to be sampled in any A/B test. If these power users share important characteristics that impact their responses, sampling many power users in quick succession can be regarded as autocorrelation. The consequence is a true average treatment effect that evolves over time – initially the ATE overweights power users, before it approaches a long-run ATE that weights all customers equally.

- **Weekday vs weekend traffic** [56]. Two distinct groups of customers use the service during these two time periods.

- **Special occasions** [38]. E-commerce sites can receive a surge of traffic leading up to the holidays from individuals who rarely shop there normally.

Even if the same customers use the service at all times, the same effect can emerge if any given individual behaves differently at different times. For instance:

- **Novelty effects** [22]. A customer has a particular reaction to a new treatment the first time she experiences it, but has a different reaction thereafter. Similarity in how different customers react initially, as compared with their later behavior, can be regarded as a type of autocorrelation. In practice, initial reactions tend be more extreme (whether positive or negative), which
produces an initial ATE that is large in magnitude but a long-run ATE that is closer to zero.

- **Weekday vs weekend usage.** The same customer has different reasons for using the service depending on the day of the week.

Given the diverse contexts in which temporal noise can arise, we prioritize a solution which relies on few assumptions about its structure and which gives maximum control to the experimenter. We note that ad hoc fixes to some of the above examples have been implemented in industry. For instance, customer heterogeneity can modeled directly in some cases (e.g. weekday vs weekend traffic can be addressed through a single fixed effect). Regarding the oversampling of power users, some companies require the experimenter to simply wait a length of time, after which the ATE is assumed to have reached its long-run value \(42\). By focusing on always valid inference within a flexible data generation model, however, we look to handle many use-cases simultaneously and to give the experimenter the freedom to choose any timeframe over which she desires inference.

### 3.0.2 Contributions

1. **Modeling autocorrelation** (Section 3.1). We model the mean response under either the treatment or the control as two random processes. Conditional on these means, the responses of all customers are considered independent. We explain why viewing customers as belonging to latent types motivates a model of this form. Data from Optimizely shows that this additional flexibility describes real-world A/B tests better than the classic IID model.

2. **Fixed-horizon inference** (Section 3.2). Within our enhanced model, we find that the t-test p-values and confidence intervals control Type I error for inference on the ATE at that horizon. If the experimenter knows in advance the timeframe over which she wants inference, and she does not want to adapt the end time of the test to manage power dynamically, then this classic approach is sufficient.

3. **Redefining always validity** (Section 3.3). The average treatment effect over any time period is now a random variable, which depends on the evolution of the mean response processes. That makes the goal of frequentist inference ambiguous. We now consider a family of null hypotheses: for any finite set of times, the average treatment effect is conditioned to be exactly zero at each such time. For each of these nulls, a Type I error is defined as the event that the experimenter stops at a time in that set and concludes that there is a non-zero effect. An always valid p-value must bound the probability of a Type I error uniformly over all null hypotheses. Always valid confidence intervals are defined similarly.

4. **The mixture Sequential Conditional Probability Ratio Test** (Sections 3.4 and 3.5). Our always valid measures are computed from an extension of the two-stream mSPRT that we term the mSCPRT. This is not a sequential test in the sense of Chapter 1 in that it does not reject
CHAPTER 3. ALWAYS VALID INFEERENCE UNDER TEMPORAL NOISE

a fixed null hypothesis at some data-dependent time, after which the null remains rejected forever. Rather with each successive pair of customers observed (one receiving the treatment and the other receiving the control), it outputs a binary decision that indicates whether the experimenter should reject the claim that the treatment effect is exactly zero, if she should choose to terminate the test immediately. Implementing the mSCPRT requires knowledge of nuisance parameters defined within our new model; for these we offer plug-in estimates.

The mSCPRT p-values satisfy an important property: for any significance level $\alpha$, the first time that the p-value drops below $\alpha$ coincides with the corresponding time for the mSPRT p-value. In other words, if the experimenter does not care about managing the population of customers on which inference is derived and she simply stops as soon as significance is achieved, then she obtains the same optimal trade-off between power and run-time as with the mSPRT.

3.1 Time dependent responses

In this first section, we develop a new model for the data in an A/B test which allows autocorrelation across successive customers. In short, responses within either the treatment or the control are viewed as conditionally independent given their means, while the mean response is modeled as the sum of a stationary Gaussian process plus a random offset. Depending on the correlation function of this GP, the mean responses will exhibit some combination of the following three types of autocorrelation:

- **High-frequency noise.** On scales much shorter than what the experimenter might reasonably choose for the duration of her A/B test, the mean response fluctuates around its local average.

- **Experiment-duration noise.** The mean response exhibits some trend over the duration of the experiment.

- **Low-frequency noise.** This noise has no impact on the data collected during the experiment. However, if the experiment were run for much longer, the model asserts that the mean response would start to shift.

We will impose a restriction on the correlation function, which rules out the possibility of experiment-duration noise but allows arbitrary high-frequency noise. We show using data from Optimizely that our new model better describes real experiments.

This chapter focuses exclusively on A/B tests that compare a treatment and a control (we do not begin with a single stream of data as in the previous chapter). As in the rest of the thesis, the incoming data over the course of an experiment can be represented by two sequences $(Y^t_n)_{n=1}^\infty$ and $(Y^c_n)_{n=1}^\infty$, corresponding to successive measurements made on the sequence of customers assigned to the treatment or control groups respectively. Marginally we assume that each observation has either
a Bernoulli or a normal distribution:

\[ Y^t_n \sim \text{Bernoulli}(\mu^t), \quad Y^c_n \sim \text{Bernoulli}(\mu^c), \]

or

\[ Y^t_n \sim N(\mu^t, \tilde{\sigma}^2), \quad Y^c_n \sim N(\mu^c, \tilde{\sigma}^2). \]

The experimenter can choose any stopping time \( T \) for the filtration \((F_n)_{n=1}^{\infty}\) generated by pairs \((T_n^t, T_n^c)\) as the terminal sample size for each group.

### 3.1.1 Understanding the IID model

The classic assumption of independence across observations is founded on the principle that all customers are interchangeable. However, the IID model can still be reasonable even if customers belong to distinct subpopulations, provided the frequency at which different types of customers are sampled does not vary over time.

To be precise, we suppose that the \( n \)th customer assigned to either the treatment or the control has a random latent type \( Z^t_n \) or \( Z^c_n \) \( \in \mathcal{Z} \). Thereafter she produces a random response, whose distribution depends on her type as well as which experience she has received. For binary data, we have

\[ Y^t_n|Z^t_n \sim \text{Bernoulli}(\mu^t(Z^t_n)), \quad Y^c_n|Z^c_n \sim \text{Bernoulli}(\mu^c(Z^c_n)) \]

for some conditional mean response functions \( \mu^t(z) \) and \( \mu^c(z) \). If there is no time dependence in how different types are sampled, we can say that each \( Z^t_n \) or \( Z^c_n \) is drawn independently from some distribution \( \mathcal{P}_Z \) over \( \mathcal{Z} \). On marginalizing over types, we obtain the IID model (3.1) with \( \mu^t = \int \mu^t(z) d\mathcal{P}_Z(z) \) and similarly for \( \mu^c \).

For continuous data, we say that the conditional distribution of the response is normal with a mean that depends on the type and the experience, and with some common variance \( \sigma^2 \):

\[ Y^t_n|Z^t_n \sim N(\mu^t(Z^t_n), \sigma^2), \quad Y^c_n|Z^c_n \sim N(\mu^c(Z^c_n), \sigma^2) \]

Let \( \tau^2 = \tilde{\sigma}^2 - \sigma^2 \). Then if types are drawn IID from a mixing distribution \( \mathcal{P}_Z \) which makes \( \mu^t(Z) \sim N(\mu^t, \tau^2) \) and \( \mu^c(Z) \sim N(\mu^c, \tau^2) \), we recover (3.2) as the marginal distribution of the observations.

### 3.1.2 Introducing correlation

On the other hand, when similar customers tend to arrive together we can think of the types \((Z^t_n)_{n=1}^{\infty}\) and \((Z^c_n)_{n=1}^{\infty}\) as two autocorrelated sequences. These sequences define mean response processes \( \mu^t = (\mu^t_n)_{n=1}^{\infty} = (\mu^t(Z_n))_{n=1}^{\infty} \) and \( \mu^c = (\mu^c_n)_{n=1}^{\infty} = (\mu^c(Z_n))_{n=1}^{\infty} \), which we model as follows:
• \( \mu^t \) and \( \mu^c \) are independent and identically distributed.

• To generate \( \mu^t \), a random offset \( \mu^t \) is first sampled from an improper \( N(0, \infty) \) distribution. Formally we will say that \( \mu^t \sim N(0, W) \) before taking the limit as \( W \to \infty \) when computing the distributions of any observable quantities.

• \( \mu^t \) is then distributed as the sum of two processes. The first is \( \mu^t_1 \); i.e. the process that takes the value \( \mu^t \) at every \( n \). The second is a mean zero Gaussian Process with an unknown stationary covariance function \( K \). The former addresses low-frequency noise, while in general the latter could model arbitrary high-frequency and experiment-duration noise. Below we will impose a restriction on \( K \), however, which removes the possibility of experiment-duration noise.

Conditional on these mean response processes, we treat all observations as independent like before. For binary data:

\[
Y^t_n | \mu^t_n \sim \text{Bernoulli}(\mu^t_n), \quad Y^c_n | \mu^c_n \sim \text{Bernoulli}(\mu^c_n). \tag{3.3}
\]

For continuous data:

\[
Y^t_n | \mu^t_n \sim N(\mu^t_n, \sigma^2), \quad Y^c_n | \mu^c_n \sim N(\mu^c_n, \sigma^2). \tag{3.4}
\]

The conditional variance \( \sigma^2 \) is assumed known. In either case, if the A/B test runs until \( n \) customers are sampled within each group, the mean responses over that period are given by \( \bar{\mu}^t_n = \frac{1}{n} \sum_{i=1}^{n} \mu^t_i \) and \( \bar{\mu}^c_n = \frac{1}{n} \sum_{i=1}^{n} \mu^c_i \). The target of inference is the average treatment effect: \( \bar{\theta}_n = \bar{\mu}^t_n - \bar{\mu}^c_n \).

The following assumption removes the possibility of experiment-duration noise, by splitting the mean response processes over the course of the experiment into some number \( J \) IID blocks – this makes the probability of a consistent trend forming over the duration of the experiment small. Nonetheless, arbitrary high-frequency noise within these shorter time blocks is still permitted. Some restriction of this kind is necessary to obtain always valid inference measures, because the structure of any experiment-duration noise cannot be learned in a single experiment.

**Assumption 6.** The covariance function \( K \) is such that the following approximation holds for a single known constant \( J \) and every sample size \( n \):

• For \( j = 0, \ldots, J \), let \( n_j = \lfloor jn/J \rfloor \).

• For \( j = 1, \ldots, J \), let

\[
\bar{\mu}^t_{n_{j-1}, n_j} = \frac{1}{n_j - n_{j-1}} \sum_{i=n_{j-1}+1}^{n_j} \mu^t_i
\]

be the average response under the treatment during the time block \( (n_{j-1}, n_j] \).
• Then the vector of differences between the block averages and the overall mean \((\bar{\mu}_{t_0,1} - \bar{\mu}_{t_n}, \ldots, \bar{\mu}_{t_{J-1},J} - \bar{\mu}_{t_n})\) is distributed as \((Z_1 - \bar{Z}, \ldots, Z_J - \bar{Z})\), where \(Z_j \sim \text{iid } N(0, s_n^2)\) for some (unknown) \(s_n^2\) and \(\bar{Z} = \frac{1}{J} \sum_{j=1}^{J} Z_j\).

3.1.3 Some technical considerations

Before we proceed, three properties of this new model warrant some discussion:

1. The mean responses are not given a well-defined law; rather they include an offset with an improper flat distribution. When we come to define frequentist inference for average treatment effects within this model, this will not be an issue, because as soon as the average effect at any time is fixed, the conditional distributions of the mean response processes are well-defined. That is, if the offset is taken to be \(N(0, W)\), the conditionals we consider will have well-defined limits in distribution as \(W \to \infty\).

2. Formally, for binary data the mean response processes should be constrained to lie in the unit interval. However, below we find that this has little impact on the resulting inference measures, at least for the experiments seen at Optimizely.

3. For continuous data, there are many pairs \(\sigma^2\) and \(K\) that produce the same law for the observed data. For instance, the IID model corresponds to the case where \(K\) makes each \(\mu_t^t \sim N(0, \tau^2)\) independently, and then conditional on \(\mu_t^t\), \(Y_t^t\) has residual variance \(\sigma^2\). However, the decomposition of the marginal variance \(\hat{\sigma}^2\) into \(\tau^2\) and \(\sigma^2\) is not identifiable. We will assume that \(K\) and \(\sigma^2\) are chosen to make \(\hat{\sigma}^2\) maximal (so in the IID case we recover the standard formulation: \(\tau^2 = 0\)). It is within that parameterization that \(\sigma^2\) is assumed known.

3.1.4 Temporal noise exists in practice

Our new model enhances the IID model by allowing arbitrary volatility on timescales at or below \(n/J\), while reducing to the IID version when the block variances \(s_n^2 = 0\). Using empirical data from Optimizely, we can ask whether this more flexible model better describes the A/B tests that most practitioners run.

We have randomly sampled 3000 experiments with binary metrics from those run during April 2015. For this analysis, we take \(J = 100\) and impose the further modeling assumption that the mean responses \(\mu_t^t\) and \(\mu_c^t\) are actually piecewise constant within each time block – that is, we look for volatility on the scale of \(n/100\), disregarding the possibility of higher frequency volatility. The values \(\bar{\mu}_t^t, \bar{\mu}_c^t\) and \(s_n^2\) are estimated jointly by maximum likelihood. A Wilkes test rejects the null hypothesis that \(s_n^2 = 0\) for all these experiments with very high confidence. For most experiments, the estimate for \(s_n^2\) was around 0.01 (see Figure 3.2). This is of a comparable magnitude to the conditional variance of the sample mean response on each time block, \(\mu_t^t(1 - \mu_t^t) \times 100/n\), confirming that the over-dispersion is non-trivial.
Figure 3.2: Histogram (on a log scale) of maximum likelihood estimates for the block variances in 3000 experiments. These experiments had binary-valued data and were run on Optimizely in April 2015.

3.2 Fixed-horizon inference

The primary goal of this chapter is to extend the always valid measures of Chapter 2 so that the experimenter can obtain frequentist inference on the average treatment effect \( \bar{\theta}_n = \bar{\mu}_t - \bar{\mu}_c \) at any data-dependent sample size. However, if she only requires inference on the ATE at a single fixed-horizon, we find that the t-test p-values and confidence intervals computed under the assumption of no time dependence suffice.

To define fixed-horizon frequentist inference, we condition on the mean responses \( \bar{\mu}_t \) and \( \bar{\mu}_c \) at the horizon. A Type I error is a rejection of the null hypothesis that \( \bar{\theta}_n = \bar{\mu}_t - \bar{\mu}_c \) takes a given value on the event that the hypothesis is true. We seek to bound the conditional probability of a Type I error rate is well-defined, even though our model contains an improper flat distribution for the offset.

**Proposition 5.** Write \( \mu^t = Z + \mu^t 1 \), where \( Z \) is a mean zero Gaussian process with covariance function \( K \) and \( \mu^t \sim N(0,W) \), and let \( n \) be any fixed sample size. The conditional distribution of \( \mu^t - \bar{\mu}_n^t 1 \) given \( \bar{\mu}_n^t \) has a well-defined limit in distribution as \( W \to \infty \). Specifically, it approaches the distribution of \( Z - \bar{Z}_n 1 \), where \( \bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i \).
Proof. Let \( \{n_1, \ldots, n_r\} \subset \mathbb{N} \) be an arbitrary finite set of times. It is suffices to show that

\[
(\mu_{n_1}^t - \bar{\mu}_n^t, \ldots, \mu_{n_r}^t - \bar{\mu}_n^t) \mid \bar{\mu}_n = (Z_{n_1} - \bar{Z}_n, \ldots, Z_{n_r} - \bar{Z}_n).
\]

Let

\[
\Sigma_{11} = \text{Var}\{(Z_{n_1} - \bar{Z}_n, \ldots, Z_{n_r} - \bar{Z}_n)\}
\]
\[
\Sigma_{12} = \text{Cov}\{(Z_{n_1} - Z_n, \ldots, Z_{n_r} - Z_n), Z_n\}
\]
\[
\Sigma_{22} = \text{Var}(\bar{Z}_n)
\]

Then we have that

\[
(Z_{n_1} - \bar{Z}_n, \ldots, Z_{n_r} - \bar{Z}_n) \sim N(0, \Sigma_{11}),
\]

while for any \( W \),

\[
(\mu_{n_1}^t - \bar{\mu}_n^t, \ldots, \mu_{n_r}^t - \bar{\mu}_n^t) = (Z_{n_1} - \bar{Z}_n, \ldots, Z_{n_r} - \bar{Z}_n, \bar{Z}_n + \mu^t)
\]
\[
\sim N\left(0, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} + W \end{bmatrix}\right)
\]

Thus

\[
(\mu_{n_1}^t - \bar{\mu}_n^t, \ldots, \mu_{n_r}^t - \bar{\mu}_n^t) \mid \bar{\mu}_n \sim N\left(\frac{\bar{\mu}_n^t}{\Sigma_{22} + W}, \Sigma_{11} - \left(\frac{1}{\Sigma_{22} + W}\right)\Sigma_{12}\Sigma_{12}^T\right)
\]
\[
\implies N(0, \Sigma_{11}).
\]

\[\square\]

**Proposition 6.** Suppose normal data and fix any horizon \( n \). Let \( p_n \) and \( \text{CI}_n \) be the t-test p-value and confidence interval based on \( \bar{Y}_n^t \) and \( \bar{Y}_n^c \). For any covariance function \( K \) and any \( 0 \leq \alpha \leq 1 \), we have the Type I error bound:

\[
\mathbb{P}_K\left(p_n \leq \alpha \mid \bar{\mu}_n^t, \bar{\mu}_n^c\right) = \alpha
\]

almost surely on the event that \( \bar{\mu}_n^t = \bar{\mu}_n^c \). Similarly,

\[
\mathbb{P}_K\left(\bar{\theta}_n \in \text{CI}_n \mid \bar{\mu}_n^t, \bar{\mu}_n^c\right) = 1 - \alpha \quad a.s.
\]

**Proof.** In short, the result holds because

\[
T_{\text{stat}} = \sqrt{\frac{n}{2\sigma^2}}(\bar{Y}_n^t - \bar{Y}_n^c - \bar{\theta}_n) = \sqrt{\frac{n}{2\sigma^2}} \times \frac{1}{n} \sum_{i=1}^n [(Y_{i}^t - \bar{\mu}_n^t) - (Y_i^c - \bar{\mu}_n^c)]
\]
is a function only of the measurement error; i.e. it is independent of $\mu^t$ and $\mu^c$ and its distribution does not depend on $K$. Since the t-test controls Type I error in the absence of time-varying conversion rates, we have that

$$P_K(|T_{stat}| > z_{1-\alpha/2} | \bar{\mu}_n^t, \bar{\mu}_n^c) = P_0(|T_{stat}| > z_{1-\alpha/2}) = \alpha.$$

Finally we note that $\{|T_{stat}| > z_{1-\alpha/2}\} = \{\bar{\theta}_n \notin CI_n\}$. Further this event coincides with $\{p_n \leq \alpha\}$ on the event that $\bar{\theta}_n = 0$. \hfill \Box

The corresponding results for binary data are messier in that the Type I error guarantees of the t-test are only asymptotic as the sample size grows large, even in the case of static testing environments. Up to that approximation, however, we find that the t-test inference measures are robust to time dependent mean responses. We omit the details here.

### 3.3 Always valid inference

To define the Type I error rate for fixed-horizon inference, we conditioned on the average mean responses at the horizon, with the rate taken to be the conditional probability of rejecting the hypothesis that the ATE at the horizon takes a given value. When the experimenter selects a data-dependent end time, we define frequentist inference by conditioning on the ATE taking some given common value at an arbitrary finite set of horizons. Proposition 7 confirms that the conditional law of the mean response processes is well-defined (the proof is similar to that of Proposition 5 and is included in Appendix C). A Type I error is the event that the experimenter stops somewhere in this set and makes an incorrect conclusion about the common value.

**Proposition 7.** Write $\mu^t = Z + \mu^t 1$, where $Z$ is a mean zero Gaussian process with covariance function $K$ and $\mu^t \sim N(0, W)$, and let $\{n_1, \ldots, n_s\} \subset \mathbb{N}$ be any finite non-empty set of sample sizes. The conditional distribution $\mu^t$ given $(\bar{\mu}_n^t, \bar{\mu}_n^c)_{n \in \mathbb{N}}$ has a well-defined limit in distribution as $W \to \infty$.

Further let $\mu = \frac{1}{2}(\mu^t + \mu^c)$ and $\theta = \mu^t - \mu^c$, and let $\theta$ be arbitrary. Condition on $(\bar{\mu}_n^t, \bar{\mu}_n^c)_{n \in \mathbb{N}}$, restrict to the event where $\bar{\theta}_n = \cdots = \bar{\theta}_n = \theta$ and pass to the limit as $W \to \infty$. Then $\mu$ and $\theta$ are independent, and the latter is distributed as

$$[\sqrt{2}(Z' - \bar{Z}'_{n_s}) + \theta 1] \{\bar{Z}'_{n_1} = \cdots = \bar{Z}'_{n_s}\}$$

where $Z'$ is another mean zero Gaussian Process with covariance function $K$.

**Definition 4.** A p-value process $(p_n)_{n=1}^\infty$ is always valid if for any finite $\mathcal{N} \subset \mathbb{N}$, $0 \leq \alpha \leq 1$, and any stopping time $T$ for the observed data that is constrained to take values in $\mathcal{N}$, we have the following bound:

$$P_K \left( p_T \leq s \left| (\bar{\mu}_n, \bar{\mu}_n)_{n \in \mathcal{N}} \right. \right) \leq \alpha$$ (3.5)
A confidence interval process \((CI_n)_{n=1}^{\infty}\) is always valid if for all \(N, \theta,\) and \(T\) taking values in \(\mathcal{N}\),

\[
P_K\left(\theta \in CI_T \mid (\hat{\mu}^t_n, \hat{\mu}^c_n)_{n \in \mathcal{N}}\right) \geq 1 - \alpha \tag{3.6}
\]

holds a.s. on the event that \(\bar{\theta}_n = \theta\) for all \(n \in \mathcal{N}\).

Note that this reduces to the existing definition of always valid inference measures (Definition 1) when \(K = 0\). In that case, the conditional probability measure considered in (3.5) and (3.6) (restricted to the relevant event) reduces to the randomness in the sequence of measurements under the time-invariant parameter \(\theta\).

### 3.4 Constructing always valid measures

Now we derive always valid inference measures by extending the two-stream mSPRT to handle the additional noise in the mean responses. For normal data, we obtain measures that control Type I error exactly but require some knowledge of the covariance function \(K\). We then explain how plug-in estimates can be used to compute the measures approximately. For binary data, our construction relies further on a normal approximation just as in Chapter 2, which ensures approximate Type I error control provided the chosen sample size is not too small. The proofs of the results in this section are given in Appendix C.

#### 3.4.1 A single null hypothesis

To be always valid, the inference measures must control Type I error uniformly over any set of times where the ATE is conditioned to take a given value. Before attempting that goal though, we first identify how the mSPRT can be adapted to control Type I error for a single null set.

Recall from Section 2.4 that for normal data in the absence of any time dependence, the sequence \((W_n)_{n=1}^{\infty} = (Y^t_n, Y^c_n)_{n=1}^{\infty}\) can be regarded as a single stream of IID data, where each \(W_n\) has an independent bivariate Gaussian distribution whose mean is parameterized in terms of the pair \((\theta, \mu)\) with \(\mu = (\mu^t + \mu^c)/2\). After a suitable rotation, the covariance matrix of this bivariate Gaussian is diagonal, and the first co-ordinate of the mean vector depends only on \(\theta\) while the second co-ordinate depends only on \(\mu\). It follows that if we fix any arbitrary value for \(\mu\), the bivariate density reduces to a univariate normal likelihood in terms of \(\theta\), \(f_\theta(w) \propto \phi\left(\frac{w - \bar{X} \cdot \theta}{\sigma \sqrt{2}}\right)\) where \(w = (x, y)\). This gives rise to a two-stream mSPRT that controls the Type I error rate uniformly over the composite null hypothesis \(H_0 : \theta = \theta_0\). Given a choice of mixing distribution \(H\), this test rejects the null as soon as the mixture likelihood ratio

\[
L_n = \int \left[\frac{f_\theta\left(\sum_{n=1}^{n}W_n\right)}{f_\theta_0\left(\sum_{n=1}^{n}W_n\right)}\right]^n dH(\theta) = \frac{\int f_\theta\left(\sqrt{\frac{2}{\sigma^2}}(\bar{Y}^t_n - \bar{Y}^c_n - \theta)\right) dH(\theta)}{\phi\left(\sqrt{\frac{2}{\sigma^2}}(\bar{Y}^t_n - \bar{Y}^c_n - \theta_0)\right)} \tag{3.7}
\]
that all times in \( N \) (i.e. \( \mathcal{N} = \mathbb{N} \)). In that case, each \( W_n \) has an independent bivariate Gaussian distribution whose mean depends on the pair \( \theta_n = \mu^t_n - \mu^c_n \) and \( \mu_n = (\mu^t_n + \mu^c_n)/2 \). In particular, if we fix an arbitrary sequence of values for \( \mu_1, \mu_2, \ldots, \), the density of each \( W_n \) reduces to \( f_{\theta_n}(w) \propto \phi \left( \frac{y-x-\theta_n}{\sigma\sqrt{2}} \right) \). It follows that the conditional likelihood ratio in favor of a mixture of alternative hypotheses, under which \( \hat{\theta}_n \) takes some common value \( \theta \) for all \( n \), corresponds to the same LR as in (3.77). Thus the same mSPRT provides the desired control Type I error bound uniformly over any sequence \( (\mu_n)_{n=1}^{\infty} \).

Now suppose we condition on \( \bar{\mu}^t_m \) and \( \bar{\mu}^c_m \) for each \( m \) in some set of times \( \mathcal{N} = \{m_1, \ldots, m_s\} \) and we restrict to the event where \( \bar{\theta}_{m_1} = \cdots = \bar{\theta}_{m_s} = \theta \). By Proposition 3, the joint conditional density of \( W_1, \ldots, W_n \) is given by

\[
\begin{align*}
    f^n_{\theta, \mathcal{N}}(w_1, \ldots, w_n) \propto \mathbb{E}_K \left\{ \prod_{i=1}^{n} \phi \left( \frac{y_i - \bar{x}_i - \theta_i}{\sigma\sqrt{2}} \right) \mid (\bar{\mu}^t_m, \bar{\mu}^c_m)_{m=1}^{i} \right\} \\
    = \mathbb{E}_K \left\{ \phi \left( \sqrt{\frac{n}{2\sigma^2}} (\bar{y} - \bar{x} - \bar{\theta}) \right) \mid (\bar{\mu}^t_m, \bar{\mu}^c_m)_{m=1}^{n} \right\} \\
    = \mathbb{E}_K \left\{ \phi \left( \sqrt{\frac{n}{2\sigma^2}} \left( \bar{y} - \bar{x} - \theta - \sqrt{2} (\bar{Z}_n - \bar{Z}_{n_s}) \right) \right) \mid \bar{Z}_{n_1} = \cdots = \bar{Z}_{n_s} \right\}
\end{align*}
\]

where \( \mathbf{Z} \) is a mean zero GP with covariance function \( K \). Letting

\[
V_{n, \mathcal{N}}(K) = 2 \mathbb{V}_{\mathcal{N}} \left\{ (Z_n - Z_{n_s}) \mid \bar{Z}_{n_1} = \cdots = \bar{Z}_{n_s} \right\}, \tag{3.8}
\]

this reduces to

\[
\begin{align*}
    f^n_{\theta, \mathcal{N}} = \phi \left( \frac{\bar{y} - \bar{x} - \theta}{\sqrt{2\sigma^2/n + V_{n, \mathcal{N}}}} \right) \tag{3.9}
\end{align*}
\]

To test the null hypothesis that the ATE at all times in \( \mathcal{N} = \theta_0 \), we define the mixture conditional likelihood ratio against that null as

\[
L_n(\mathcal{N}) = \frac{\int f^n_{\theta, \mathcal{N}}(Y^t_n - Y^c_n) d\mathcal{H}(\theta)}{\int f^n_{\theta_0, \mathcal{N}}(Y^t_n - Y^c_n)} \tag{3.10}
\]

\textbf{Proposition 8.} For normal data and any finite \( \mathcal{N} \subset \mathbb{N} \), the mSPRT based on \( L_n(\mathcal{N}) \) controls Type I error. Specifically, for any stopping time \( T \) and \( 0 < \alpha < 1 \),

\[
\mathbb{P}_K \left( L_T(\mathcal{N}) \geq 1/\alpha \mid (\bar{\mu}_m, \bar{\mu}^c_m)_{m \in \mathcal{N}} \right) \leq \alpha \tag{3.11}
\]
holds a.s. on the event that $\bar{\theta}_m = \theta_0$ for all $m \in N$.

### 3.4.2 Uniform Type I error control

Next we extend this mSPRT to give uniform control over any set of times where the ATE is conditioned to take some value, on the event that the experimenter chooses to stop at one of those times. The idea is that, if the user should stop at some sample size $m$, the hardest null to control for would be the null that conditions on that time only; i.e. $N = \{m\}$, because that is the null that leaves the most residual variance in the sample differences $(\bar{Y}_n^c - \bar{Y}_n^t)_{n \leq m}$. If the mSPRT for testing that singleton null would reject at this sample size, then we can be sure that the mSPRT for testing any other relevant null would also.

**Definition 5.** For normal data, the mixture Conditional Probability Ratio Test is a sequence of binary decisions for $m = 1, 2, \ldots$ that indicate a rejection if the user chooses to stop at that sample size. Let

$$\Lambda_m = \max_{n \leq m} L_n(\{m\}).$$

A rejection is offered at $m$ if $\Lambda_m \geq 1/\alpha$; i.e. if the mSPRT based on $L_n(\{m\})$ truncated at $m$ produces a rejection.

**Theorem 4.** For normal data, consider any finite $N \subset \mathbb{N}$ and any stopping time $T$ for the observed data that is constrained to lie in $N$.

$$\Pr_{K} \left( \Lambda_m \geq 1/\alpha \left| (\bar{\mu}_m^l, \bar{\mu}_m^c)_{m \in N} \right. \right) \leq \alpha$$

(3.12)

holds a.s. on the event that $\bar{\theta}_m = \theta_0$ for all $m \in N$.

As in Chapter 3, always valid p-values are derived from the mSCPRT by taking $\theta_0 = 0$ and identifying the least significance level that produces a rejection at each sample size – specifically, $p_n = 1/\Lambda_m$. Confidence intervals are obtained as the set of $\theta_0$ where the mSCPRT does not reject at the current sample size.

### 3.4.3 Binary data

For binary observations, we approximate the product of binomial likelihoods across the treatment and control by a bivariate Gaussian to obtain the conditional likelihood ratio for a single null hypothesis,

$$L_n(N) = \frac{\int f^n_{\bar{Y}_n^c, \bar{Y}_n^t} dH(\theta)}{f^n_{\theta_0, N}(\bar{Y}_n^c, \bar{Y}_n^t)}$$
with
\[ f_{\theta,N}^n(\bar{x}, \bar{y}) = \phi \left( \frac{\bar{y} - \bar{x} - \theta}{\sqrt{\frac{1}{n} \{ \bar{x}(1 - \bar{x}) + \bar{y}(1 - \bar{y}) \} + V_{n,N}}} \right). \] (3.13)

From here the mSCPRT, and consequently the p-values and confidence intervals, are defined exactly as for the normal case. Approximate Type I error control is obtained, provided \( T \) is constrained to be moderately large.

### 3.4.4 Plugging in for the covariance function

Hereafter we will abuse notation and write simply \( V_{n,m} \) to denote the variances \( V_n, \{ m \} \), when conditioning on a singleton null time \( N = \{ m \} \). Computing the mSCPRT p-values and confidence intervals requires knowledge of the covariance function \( K \), in so far as the conditional likelihood ratios depend on these variances
\[ V_{n,m}(K) = 2 \text{Var}_K(\bar{Z}_n - \bar{Z}_m). \]

Fortunately we can leverage Assumption 6 to derive plug-in estimates.

By assumption, at each sample size \( m \), the vector of differences between the block averages and the overall mean \( (\bar{\mu}_{0,1} - \bar{\mu}_m, \ldots, \bar{\mu}_{J-1,J} - \bar{\mu}_m) \) is distributed as \((\zeta_1 - \bar{\zeta}, \ldots, \zeta_J - \bar{\zeta})\), where \( \zeta_j \overset{iid}{\sim} N(0, s^2_{m}) \).

For normal data, it follows that the sample block averages \((\bar{Y}_{0,1} - \bar{Y}_m, \ldots, \bar{Y}_{J-1,J} - \bar{Y}_n)\) are distributed as \((\zeta'_1 - \bar{\zeta}', \ldots, \zeta'_J - \bar{\zeta}')\), where \( \zeta'_j \overset{iid}{\sim} N(0, s^2_{m} + J\sigma^2/n) \). Pooling the sample block averages from the treatment and control groups at each time can be used to obtain an estimate for \( s^2_{m} \). For binary data, we can use a normal approximation to estimate \( s^2_{m} \) similarly.

Now suppose \( Z \) is a mean zero GP with covariance function \( K \), and fix a sample size \( m \). Then the block averages \((\bar{Z}_{0,1} - \bar{Z}_m, \ldots, \bar{Z}_{J-1,J} - \bar{Z}_m)\) are similarly distributed as \((\zeta_1 - \bar{\zeta}, \ldots, \zeta_J - \bar{\zeta})\), where \( \zeta_j \overset{iid}{\sim} N(0, s^2_{m}) \). Up to a discretization error, for \( n \leq m \) we have
\[
\bar{Z}_n - \bar{Z}_m = \frac{m}{Jn} \sum_{j=1}^{Jn/m} (\bar{Z}_{j-1,j} - \bar{Z}_m)
= \frac{m}{Jn} \sum_{j=1}^{Jn/m} (\zeta_j - \bar{\zeta}) = \frac{1}{J} \left( \frac{m}{n} - 1 \right) \sum_{j=1}^{Jn/m} \zeta_j - \frac{1}{J} \sum_{j=Jn/m+1}^{J} \zeta_j \]

Thus we find \( V_{n,m} = 2s^2_{m}(m/n - 1)/J \). Plugging in the estimate for \( s^2_{m} \) provides an estimate for \( V_{n,m} \) at each \( n \leq m \), which then lets us compute a value for \( \Lambda_m \).
3.5 Evolution over time

The mSCPRT inherits the optimality benefits of the mSPRT, if the experimenter is keen to stop her test as soon as any effect is detected, while it also lets the experimenter keep waiting for valid inference on longer timescales.

Note that for any $m$, $V_{m,m} = 0$ (and in fact the plug-in scheme will necessarily preserve this property even if the estimate of $s^2_m$ is poor). Hence $L_m(\{m\})$ coincides with the mixture likelihood ratio given in (3.7) that assumes the IID model. It follows that the first time that the mSCPRT offers a rejection, $T_0$, is simply the rejection time of the static two-stream mSPRT. If the experimenter is not interested in setting her sample size to manage the population sampled, then she obtains the optimal trade-off between power and run-time discussed in Chapter 2. On the other hand, suppose the experimenter chooses to let her A/B test keep running past $T_0$. Then $V_{T_0,m}$ will increase with $m$ until $L_{T_0}(\{m\})$ potentially falls below $1/\alpha$. The mSCPRT will stop offering a rejection unless subsequent data indicates that a non-zero ATE persists.

This ability to adapt to a changing ATE is seen in Figure 3.3, where we have simulated a novelty effect. We fixed the sequence of mean responses so that the initial effect is positive, but after $n = 1000$ the ATE approaches zero. We then sampled binary data under these mean responses, and we plotted the sample difference $\bar{Y}_n^t - \bar{Y}_n^c$ over time, as well as the confidence intervals for the mSCPRT and the mSPRT. Both the mSCPRT and the mSPRT detect the initial effect, with both intervals strictly positive at $n = 1000$. However, as the effect disappears and so the sample difference approaches zero, the mSPRT cannot adapt — this confidence interval actually just vanishes! The mSCPRT interval tracks the falling sample difference as one would hope.
3.6 Conclusion

Through the inference measures derived in this chapter, we have enabled the user to set sample sizes dynamically so as to manage both the power of her A/B tests and the timeframe over which average effects are measured. Specifically we have obtained frequentist measures with approximate Type I error control, which sacrifice nothing for a user who cares only about power as compared with the always valid measures computed in Chapter 2. To do this, we have not had to impose much structure on which customers use the service at different times and how that induces autocorrelation in their responses. We assume only that this autocorrelation is stationary and that experiment-duration noise can be ignored. In future work, one could alter this assumption and look for always valid measures under alternative time series models that allow different types of temporal noise.

3.6.1 Optimality

Moving beyond users who only care about power, we could ask whether the mSCPRT measures are optimal for an experimenter who has some competing priorities in terms of power and the timeframe of inference. For instance, suppose there is some range of times where she deems average effects interesting, and she stops as early as possible within that range in order to obtain a rejection. Do these measures trade off power and expected sample size optimally, as compared with other measures that control the Type I error rate uniformly over that range? This is an interesting direction for future work.

3.6.2 Broader decision-making

We have focused on inference for the average effect while remaining agnostic about how the experimenter should use this inference in any subsequent decisions. The strength of this approach is that it makes our inference measures widely applicable, but with additional structure on her action space, there are likely to be more efficient mappings straight from the observed data to decisions. Valuable future work would be to compare natural decision rules based on the mSCPRT p-values with alternatives such as restless bandit algorithms [73].
Chapter 4

Inference across many A/B tests

So far we have focused on frequentist inference for the parameter in a single A/B test. In practice, though, it is common for an experimenter to run many tests in parallel, because there are several related questions that need to be answered. If she relies on accurate inference in all of these tests in order to optimize any subsequent decisions, she may require stronger frequentist guarantees than simply a separate bound on the Type I error probability in each experiment. Our goal in this chapter is robust simultaneous inference across a collection of A/B tests.

4.0.1 Setup

There are diverse contexts in which practitioners may run more than one A/B test, but we focus on the following common scenario:

- A fixed set of experiments $s = 1, \ldots, m$ are run in parallel. The experimenter starts and stops all of her tests at the same time.
- The incoming data in each test is assumed IID, so that the target of inference is a time-invariant parameter $\theta^s \in \Theta^s$.
- No further assumption is made on the distribution of the observations, except that always valid p-values and confidence intervals for each parameter can be obtained (see Definitions 11 and 12).
- The data streams, and hence the inference measures, may be independent or correlated across tests.

Examples here include:

- Comparing the same treatment and control on $m$ different metrics.
• Comparing \( m \) different treatments against a single control.

Then we focus on the two most common types of multiple hypothesis testing controls sought. For p-values, these are

1. Bounding the \textit{Family-Wise Error Rate}: the probability of obtaining any false positives among the \( m \) tests.

2. Bounding the \textit{False Discovery Rate}: the expected proportion of any significant p-values that are false positives.

For confidence intervals, the FWER is defined similarly as the probability that any of the \( m \) intervals fail to cover their true parameter. The analogy of FDR is the \textit{False Coverage Rate}: the expected proportion that fail to cover the parameter, out of some subset of intervals that the experimenter chooses to look at.

### 4.0.2 Demands of A/B testing practice

Standard procedures take fixed-horizon p-values and confidence intervals as input and output corrected versions that control either the FWER or the FDR / FCR; these are the Bonferroni [23] and Benjamini-Hochberg [7, 8, 9] procedures respectively. In the latter case, there are two versions of the procedure: one that assumes the \( m \) data streams are independent and another that allows arbitrary correlation across the data streams.

However, these methods are insufficient for use in A/B testing practice. First they do not allow for continuous monitoring. Second, in the case of FCR control, BH requires that the rule by which the experimenter chooses which intervals to look at be pre-specified – most typically it is assumed that she only looks at those tests where a significant p-value is obtained. In A/B testing, an experimenter will often have a set of “guardrail” tests where she will necessarily reference the confidence intervals, because her decision-making relies on understanding the range of possible values for those parameters. The corrected confidence intervals must be robust to which tests might lie in her personal guardrail set.

**Example 4.** Each test compares the same treatment and control on a different metric. Some of these are guardrail metrics – the experimenter does not expect a positive treatment effect on any of these, but it is important that none of these effects are too large and negative. Thus she will be sure to check the lower end of each of the guardrail confidence intervals so that she can appraise the worst-case impact of deploying the treatment.

### 4.0.3 Contributions

1. \textit{FCR control for an unknown guardrail set}. We derive a variant of the BH procedure under independence, which takes as input fixed-horizon confidence intervals and outputs corrected
intervals that control the FCR approximately, when the experimenter chooses to look at all significant tests plus some guardrail set of size $j \ll M$.

2. *Commutation with always valid p-values.* We find that both Bonferroni and the version of the BH procedure that allows for arbitrary correlation between the p-values commute with always validity. That is, if these corrections are applied to an arbitrary collection of always valid p-values at each time, then the FWER or FDR respectively is controlled at any data-dependent sample size. For the version of BH which assumes independence, we derive a sufficient condition on the stopping time at which FDR control can be obtained for any collection of independent always valid p-value processes.

3. *Sequential FCR control.* We obtain a sufficient condition on the stopping time, such that given any collection of always valid confidence intervals as input, our new corrected intervals control the FCR in the presence of continuous monitoring.

### 4.0.4 Related work

There is growing interest in achieving multiple hypothesis testing controls in sequential contexts. For the most part, work in this area considers a different form of streaming data to the one described in this paper: [27] and [33] provide methods to control the family-wise error rate or the false discovery rate when experiments are performed sequentially, but within each experiment the data is accessed only once. However, [76] combines these approaches with sequential hypothesis testing to enable FDR control in the A/B testing regime, where the data within each experiment arrives over time as well. The major difference between that paper and our own work is that we look to preserve the simple user interface of p-values and confidence intervals.

### 4.1 Preliminaries

We begin by reviewing the setup, the multiple hypothesis controls we seek, and the standard procedures we leverage: the Bonferroni and Benjamini-Hochberg procedures.

#### 4.1.1 Multiple A/B tests

**Tests.** $s = 1, \ldots, m$.

**Parameters.** $\theta^s \in \Theta^s$ for $s = 1, \ldots, M$.

**Data.** For each experiment, data arrives at times $n = 1, 2, \ldots$. We allow both one-variation experiments where the data forms a single IID stream, and regular A/B tests where each data point is a pair – one observation on the treatment and one on the control – and each of the two streams are IID. We will consider both the case that the data are independent across tests and the case that they are correlated.
Inference measures. At each $n_i$, for each experiment, the data available is condensed into a p-value $p_{n_i}$ for testing the hypothesis that $\theta^s = 0$ and a $(1 - \alpha)$-level confidence interval $\text{Cl}_{n_i}(1 - \alpha)$ for $\theta^s$. We will consider both fixed-horizon and always valid inference measures. In either case, we will assume that the confidence intervals are derived from the p-values, so that zero is in the interval if and only if the p-value is not significant at level $\alpha$.

Dynamic sample sizes. For always valid measures, we will let the user stop at any stopping time $T$ for the filtration generated by all of the incoming data.

4.1.2 Target error bounds

When using the p-values to test the null hypotheses that each $\theta^s = 0$, we seek to bound the following cross-experiment error rates:

1. The Family-Wise Error Rate:

$$\text{FWER} = \max_{\theta} \mathbb{P}_{\theta}(\delta^s = 1 \text{ for at least one } s \text{ s.t. } \theta^s = 0).$$

(4.1)

Here $\delta^s$ is a binary variable that indicates a rejection in test $s$.

2. The False Discovery Rate:

$$\text{FDR} = \max_{\theta} \mathbb{E}_{\theta}\left\{ \frac{\#\{1 \leq r \leq m : \theta^s = 0, \delta^s = 1\}}{\#\{1 \leq s \leq m : \delta^s = 1\} \wedge 1} \right\}.$$

(4.2)

For confidence intervals, we seek to bound the following non-coverage rates:

1. The Family-Wise Error Rate:

$$\text{FWER} = \max_{\theta} \mathbb{P}_{\theta}(\theta^s \notin \text{Cl}^s \text{ for at least one } s).$$

(4.3)

2. The False Coverage Rate:

$$\text{FCR} = \max_{\theta} \mathbb{E}_{\theta}\left\{ \frac{\#\{1 \leq r \leq m : \theta^s \notin \text{Cl}^s, \mathcal{L}^s = 1\}}{\#\{1 \leq s \leq m : \mathcal{L}^s = 1\} \wedge 1} \right\}.$$

(4.4)

$L^s$ is a binary variable indicating that the user chooses to look at the interval on test $s$.

4.1.3 Standard procedures

For testing the null hypotheses that $\theta^s = 0$, the FWER is controlled using the Bonferroni correction. This takes fixed-horizon p-values as input and rejects hypotheses $(1), \ldots, (R)$ where $R$ is maximal such that $p^{(R)} \leq \alpha/m$, and $p^{(1)}, \ldots, p^{(m)}$ are the p-values arranged in increasing order. For FDR, the standard procedure is Benjamini-Hochberg. Given fixed-horizon p-values, two versions of
BH are used depending on whether the data are known to be independent across experiments. If independence holds (BH-I), the procedure rejects hypotheses (1), \ldots, (R) where R is maximal such that \( p^{(R)} \leq \alpha R / m \); in general (BH-G), the procedure chooses the maximal R such that:

\[
p^{(R)} \leq \frac{\alpha R}{m \sum_{r=1}^{m} 1/r}.
\]

Both of these procedures can be restated as mappings from the fixed-horizon p-values to a vector of fixed-horizon q-values: \( q^* = p^* m \) for Bonferroni, \( q^{(s)} = \max_{r=1}^{s} p^{(r)} m / s \) for BH-I and similarly for BH-G. These maintain the simple user interface of p-values. The user thresholds each q-value at her personal tolerance \( \alpha \) and obtains a set of rejections that bound the relevant error rate at \( \alpha \).

Similarly, Bonferroni and BH-I both provide mappings from \((1 - \alpha)\)-level fixed-horizon confidence intervals to corrected intervals, which bound the corresponding error rate at \( \alpha \). For the FWER, the corrected intervals are given by \( \text{Cl}^*(1 - \alpha / m) \). For BH-I, the intervals are \( \text{Cl}^*(1 - \alpha R / m) \), where R is suitably defined in terms of the rule by which the user chooses which intervals to reference \([9]\). If the user only references intervals in those tests where the p-value is significant, this \( R \) is simply the number of such significant tests, \( R^{BH} \).

### 4.2 Fixed-horizon: FCR control

There exists extensive theory for how Bonferroni and BH can be used in the fixed-horizon context. However, there remains one roadblock to applying these procedures in fixed-horizon A/B testing: we need FCR control for an unknown selection rule. Here we adapt the BH-I procedure to produce corrected, fixed-horizon confidence intervals, which approximately bound the FCR for an experimenter who looks at all significant A/B tests, together with a fixed but unknown set of guardrail experiments \( J \) with \( j = |J| \ll m \).

We achieve this by extending the approach of \([9]\), where FCR control is obtained by replacing fixed-horizon intervals \( \text{Cl}^*(1 - \alpha) \) by \( \text{Cl}^*(1 - \alpha R / m) \). If the user is known to look only at significant tests, \([9]\) takes \( R = R^{BH} \). We find that approximate FCR control is possible by using a different \( R \) for each test, where \( R \) is a lower bound on the number of experiments the user could be looking at, given that she is referencing that particular one. Namely for significant tests, \( R = R^{BH} \), whereas for insignificant tests, \( R = R^{BH} + 1 \). The proof of the following theorem is given in Appendix E.

**Theorem 5.** Given fixed-horizon p-values \( p^* \) and corresponding intervals \( \text{Cl}^*(1 - \alpha) \), let \( S^{BH} \) be the rejection set under BH-I with \( R^{BH} = |S^{BH}| \). Define the corrected confidence intervals:

\[
\tilde{\text{Cl}}^*(s) = \begin{cases} 
\text{Cl}^*(1 - R^{BH} \alpha / m) & s \in S^{BH}; \\
\text{Cl}^*(1 - (R^{BH} + 1) \alpha / m) & s \notin S^{BH}.
\end{cases}
\] (4.5)
These ensure that, for any set of guardrails $J \subset \{1, \ldots, m\}$, the FCR is at most $\alpha(1 + j/m)$.

Note that there is some slack in this result when $j$ is large. Indeed, if $j = m$, the user necessarily looks at every test, so the FCR is bounded at $\alpha$.

### 4.3 Always valid p-values

Now we consider applying Bonferroni and BH to an arbitrary collection of always valid p-values, and we ask when the resulting q-values control FWER or FDR in the presence of continuous monitoring. We find that Bonferroni or BH-G “commute” with always validity: they achieve the desired error bounds at any data dependent sample size. For applying BH-I to an arbitrary set of independent always valid p-values, we find a sufficient condition on the stopping time such that FDR is controlled.

#### 4.3.1 Bonferroni and BH-G

**Proposition 9.** Let $p^*_n$ be any collection of always valid p-values, all of which are adapted to a filtration $(\mathcal{F}_n)_{n=1}^{\infty}$. Let $T$ be any stopping time with respect to $(\mathcal{F}_n)_{n=1}^{\infty}$. The Bonferroni correction applied to the collection $p^*_T$ controls the FWER.

**Proof.** It follows immediately from the definition of always valid p-values that the stopped processes $p^*_T$ define a collection of fixed-horizon p-values. The result follows because Bonferroni controls FWER for any collection of fixed-horizon p-values as input.

**Proposition 10.** Let $p^*_n$ be any collection of always valid p-values, all of which are adapted to a filtration $(\mathcal{F}_n)_{n=1}^{\infty}$. Let $T$ be any stopping time with respect to $(\mathcal{F}_n)_{n=1}^{\infty}$. BH-G applied to the collection $p^*_T$ controls the FDR.

Although the idea here is the same, we cannot use quite the same proof as for Proposition 9 because Theorem 1.3 in [8], which establishes that BH-G controls FDR under arbitrary correlation, requires that the fixed-horizon p-values be strictly uniformly distributed on the unit interval (rather than super-uniform as $p^*_T$ may be). A proof of Proposition 10 is given in Appendix D.

#### 4.3.2 When does BH-I control FDR?

In general, BH-I does not commute with always validity for independent p-value processes, because stopping times that depend on every experiment can introduce correlation in the stopped p-values; i.e., given always valid p-values $p^*_n$, $p^*_T$ may not define a collection of independent fixed-horizon p-values. Nonetheless Theorem 1 gives a sufficient condition on the stopping time such that FDR control is achieved. The proof can be found in Appendix D.
Definition 6. Given independent always valid p-values \( p_n \), let \( S_n^{BH} \) be the rejections when BH-I is applied to these at level \( \alpha \) and let \( R_n^{BH} = |S_n^{BH}| \). Define:

\[
T_r = \inf \{ n : R_n^{BH} = r \}; \\
T_r^+ = \inf \{ n : R_n^{BH} > r \}; \\
T_r^s = \inf \{ n : p_n^s \leq \frac{\alpha r}{m} \}.
\]

Now, if \( p_n^{(1),-s}, p_n^{(2),-s}, \ldots \) are the p-values for the experiments other than \( s \) placed in ascending order, consider a modified BH procedure that rejects hypotheses \( (1), \ldots, (k) \) where \( k \) is maximal such that \( p_n^{(k),-s} \leq \alpha(k + 1)/m \), in parallel to the fixed-horizon approach in [2]. Define the rejection set \( (S_n^{BH})_{0}^{-s} \) as those obtained under the original BH-I procedure if \( p_n^s = 0 \). Let \( (R_n^{BH})_{0}^{-s} = |(S_n^{BH})_{0}^{-s}| \) and define:

\[
(T_r)_0^{-s} = \inf \{ n : (R_n^{BH})_{0}^{-s} = r \} \\
(T_r^+)_0^{-s} = \inf \{ n : (R_n^{BH})_{0}^{-s} > r \}.
\]

Theorem 6. Given a stopping time \( T \), let \( M_0 \) be the number of truly null hypotheses and let \( I \) be the set of null hypotheses \( s \) such that:

\[
\sum_{r=1}^{M} \mathbb{P} \left( (T_{r-1})_0^{-s} \leq T < (T_{r-1})_0^{+s} \mid T_r^s \leq T, T < \infty \right) > 1 \tag{4.6}
\]

Then the rejection set \( S_T^{BH} \) has FDR at most

\[
\alpha \left( \frac{M_0}{M} + \frac{|I|}{M} \sum_{k=2}^{M} \frac{1}{k} \right).
\]

In particular, if we permit only stopping times where \( I \) is empty, BH-I controls FDR.

The set \( I \) is empty for many natural choices of stopping time.

Example 5. Suppose the user stops at the first time that some fixed number \( x \leq m \) hypotheses are rejected; i.e. \( T = \inf_n \{ n : R_n = x \} \). In that case,

\[
\mathbb{P} \left( (T_{r-1})_0^{-s} \leq T_x < (T_{r-1})_0^{+s} \mid T_r^s \leq T_x, T_x < \infty \right) = \mathbb{P} \left( (T_{r-1})_0^{-s} \leq T_x < (T_{r-1})_0^{+s} \mid T_x < \infty \right)
\]

for each A/B test \( s \). This probability is 1 if \( r = x \) and 0 otherwise, so \( I \) is empty and FDR is controlled exactly.
4.4 Always valid confidence intervals

It is straightforward to show that Bonferroni and BH-G commute with always valid confidence intervals, just as they do with p-values – we omit the details here. This just leaves the question: when does applying the modified BH-I procedure we derived in Section 4.2 to an arbitrary collection of independent always valid confidence intervals control FCR?

Theorem 7 gives a sufficient condition on the stopping time, at which the FCR is controlled up to the \((1 + j/m)\) inflation factor that was incurred in the fixed-horizon case. The proof of this result combines the results of Theorem 6 with the same approach used to prove Theorem 5 – it is given in Appendix E. The major difference between controlling FDR and FCR is that for confidence intervals we are not interested in the times at which rejections are made, but rather the times at which true values of parameters are dropped from intervals.

**Definition 7.** If \(p^k_{n, \theta_0}\) is the p-value for testing \(H_0: \theta^s = \theta_0\), let

\[
T^s,\theta_0 = \inf\{n : p^s,\theta_0 \leq \frac{\alpha r}{m}\}
\]

\[
(T^s)_0^{-s,J} = \inf\{n : |(S^BH)_0^{-s} \cup J| = r\}
\]

\[
(T^s)_r^{-s,J} = \inf\{n : |(S^BH)_0^{-s} \cup J| > r\}.
\]

The last two stopping times denote the first times at least \(r\) and more than \(r\) experiments other than \(s\) are selected.

Now consider an additional modified BH procedure that rejects hypotheses \((1), \ldots, (k)\) where \(k\) is maximal such that

\[
p^{(k)-s}_{n, \theta_0} \leq \frac{k}{m}.
\]

These are the rejections obtained under the original BH-I procedure if \(p^s_{n, \theta_0} = 1\). We define stopping times associated with this procedure \((T^s)_1^{-i,J}\) and \((T^s)_1^{+i,J}\) analogous to the two stopping times above.

**Theorem 7.** Given independent always valid p-values \(p^s_{n, \theta_0}\) and corresponding CIs \(\text{Cl}^s_{n}\), we define new confidence intervals

\[
\tilde{\text{Cl}}^s_{n} = \begin{cases} 
\text{Cl}^s_{n}(1 - R^{BH}_n\alpha/m) & s \in S^{BH}_n \\
\text{Cl}^s_{n}(1 - (R^{BH}_n + 1)\alpha/m) & s \notin S^{BH}_n
\end{cases}
\]

Let \(J\) be a subset of the A/B tests run and let \(T\) be a stopping time such that the following conditions hold for every \(s\), where \(\theta^s\) is the true parameter value for that hypothesis:

\[
\sum_{r=1}^{m} \mathbb{P}((T^s)_0^{-s,J} < T < (T^s)_r^{-s,J} | T^s,\theta^s \leq T < \infty) \leq 1
\]

\[
\sum_{r=1}^{m} \mathbb{P}((T^s)_1^{-s,J} < T < (T^s)_r^{+s,J} | T^s,\theta^s \leq T < \infty) \leq 1
\]
Then the stopped intervals $\bar{C}_T^\alpha$ have FCR at most $\alpha(1 + j/m)$.

### 4.5 Conclusion

We have shown that always valid p-values and confidence intervals may be combined with two commonly used multiple testing procedures to obtain sequential FDR control and approximate sequential FCR control. Only weak conditions on the dynamic sample size that the user may choose and the rule by which she may select which confidence intervals to look at are needed. Further, the theory in this chapter is quite general, in that it requires no direct assumptions on the types of hypotheses tested or on the distributions of the incoming data streams.

As in Chapter 2, the greatest drawback to our work is that we have little to say about power. Future work might ask: if optimal always valid p-values and confidence intervals are provided as input, do the procedures discussed here obtain an optimal trade off between run-time and suitable cross-experiment notions of power?

Finally we note that we have only addressed one context in which practitioners want joint inference across multiple A/B tests. In addition to running tests in parallel, it is common for experiments to be started sequentially – indeed there is often endogeneity in which hypotheses are tested later, depending on what the user discovers in her initial experiments. There could be a wide range of valuable future work in establishing sequential FDR and FCR control in such scenarios.
Chapter 5

Incentivizing better experiment designs

In the last three chapters, we have grappled with how the observations within each A/B test arrive sequentially. Now we turn our attention to another sequential aspect of A/B testing practice: experiments themselves are run one after another. At a company that runs many tests, it is important that each experiment is designed in a way that not only facilitates useful inference on the parameters measured in that test, but also ensures that the company can keep running productive experiments in the future. In this chapter we ask: how can such a company motivate its employees to design A/B tests in a way that aids subsequent experiments, even those that the employee does not run herself?

We consider a company with $N$ employees, each of whom runs her own sequence of experiments over time. We focus on two experiment design choices that the employees must make in each test, which impact future experimentation across the company:

1. **Fixed-horizon sample sizes.** Optimizing the sample size for inference in a single experiment is a trade-off between power and run-time – one that must be tailored to the magnitude of typical effects sought (see Chapter 4). Choosing a larger sample size than otherwise necessary benefits later experiments, though, because it leads to a more accurate estimate of the effect, which in turn helps later experimenters understand what effect sizes are typical.

2. **Treatment types.** Some types of treatments offer larger effect sizes than others. Experiments with larger effects are preferable, because the effect can be detected with fewer samples and detecting it is generally more lucrative. In the short-term it is optimal to test a treatment whose type is known to produce large effects, but subsequent experimentation benefits when an employee runs an A/B test on a type that is poorly understood.
We study this problem within a Bayesian setup. At each time, the uncertainty among the employees about the magnitude of typical effects associated with each treatment type is summarized as a common prior. Each A/B test is not analyzed through the frequentist interface of p-values and confidence intervals, but rather this prior is combined with the experimental data to produce a posterior for the treatment effect in that test. This provides a clean formulation for us to study how the employees update their beliefs about typical effect sizes over time, and how they trade off immediate inference with the need to learn. Indeed we consider the formal model we develop for how employees reason about experiment design to be an important contribution of our work.

We let the company motivate its employees by regulating how much experimental data the employees get to see. Note that if an employee embarks on greater exploration (she selects a larger sample size or a riskier type) and all the resulting data is released to all $N$ employees, this exploration produces an $N$-fold benefit for the company versus what the employee feels personally. We present a novel mechanism, which we term the Minimum Learning Mechanism, that determines what data the employees can access at any time depending on their past experiment design choices. This mechanism is designed so that any additional exploration unlocks an $N$-fold quantity of experimental data for her personally. This aligns the incentives of each employee with that of the company as a whole.

**Example 6.** Consider a content aggregation service, whose business model centers on making content go viral. Employees at this service source content externally, before publishing it on either the company’s website or on its mobile application. The company makes money when visitors to either the site or the app click on the published content. More lucrative than a single click, though, is when a visitor shares that content to her social media, as this can ultimately attract many more visitors.

Before any new content is published, an A/B test is run that compares the existing version of the site or app (as appropriate) with an updated version with the new content displayed. For each version, the proportion of visitors who share any available content to their social media is measured. The treatment effect is the increase in share rate, and the company will only decide that the content should be published if there is evidence that this effect is positive.

For any new content she sources, there are two decisions for an employee to make:

1. Where to launch the content. She can run an A/B test measuring the appeal of the new content on the website or she can run an A/B test about its appeal on the app.

2. A fixed-horizon sample size for that experiment.

The employee will naturally choose the type of experiment (web or mobile) that has tended to produce larger effects in the past, although we note that she may also be guided by private information about where this particular content is most likely to go viral. The company as a whole benefits if the employee can be motivated to choose the type where typical effect sizes are less well understood, as well as to choose a larger sample size than is needed to detect an effect in this single experiment.
5.0.1 Contributions

We model designing experiments at each time period as a sequential game between the employees. Accurate inferences in these tests produce rewards for the experimenter, and each employee looks to maximize her expected discounted sum of rewards across all time. This casts the goal for the company as a mechanism design problem; namely how best to regulate the flow of information within this game, in order to maximize the average welfare across all the employees.

We focus on a novel solution concept that we term stationary, symmetric, \textit{boundedly rational Markov perfect equilibria}. Stationary, symmetric Markov perfect equilibria are commonly studied phenomena, where the action of each agent at each time is a fixed function of some suitably defined “current state” for that agent \cite{X, Y}. In this case, the state is taken to be a representation of the current belief that all employees have about typical effect sizes, together with a private signal that she personally receives about the treatment effect in her next experiment. The bounded rationality assumption we impose is that employees only use the experimental data revealed when updating their beliefs. We evaluate the performance of any mechanism by comparing the welfare at the most favorable stationary, symmetric BR-MPE against the benchmark welfare that could be obtained if the company could reveal all data to the employees, could ensure that all experiments run are high-impact, and could dictate how employees choose experiment types and sample sizes so as to achieve an optimal exploration-exploitation trade-off.

Within this framework, we present two contributions:

1. \textit{The Minimum Learning Mechanism}. We derive a mechanism that perfectly aligns incentives in the absence of private information among employees. At the end of each time period, the MLM only reveals data from experiments on the riskiest type of treatment investigated in that period. Further, in each of these experiments, only a number of observations equal to the smallest sample size chosen in that period are revealed. In the absence of private information, the benchmark for the company is defined to be the maximum payoff achievable when all employees are required to play a common strategy of the company’s choosing. We prove that, under the MLM, the most favorable BR-MPE recovers the benchmark welfare exactly.

2. \textit{Robustness to private information}. In reality, experiment design choices at these companies are delegated to individual experimenters, because the experimenters can tailor their choices to some private understanding of which tests will be most lucrative. We focus on the case where there is only one type of treatment, so that employees only need to select sample sizes. To define a benchmark, we consider the hypothetical scenario where all experiments are high-impact, and then we allow the company to dictate a common strategy under which the employees map the current belief at each time to a sample size. We obtain a bound on the welfare loss under the MLM against this benchmark in terms of the strength of the employees’ private information. The loss is found to approach zero smoothly if the private signals are
only weakly informative.

5.0.2 Related work

Although less common than p-values and confidence intervals, Bayesian approaches to A/B testing have started to gain industrial popularity [60, 10]. For our purposes, we suppose simply that the employees begin with a common prior about typical effect sizes, which they update using the empirical data revealed to them. However, in the absence of an initial prior, [17] offers an empirical Bayes approach that fits the prior directly to the observed data. While we restrict our attention to fixed-horizon sample sizes in this chapter, [19] also discusses when Bayesian A/B testing is robust to continuous monitoring.

The primary goal of this chapter is to motivate employees at a company to embark on more exploration than they otherwise would. Papers such as [29, 50] have investigated this goal in a range of contexts, usually incentivizing exploration by paying money to the employees. [51] considers a more similar setup to ours, where in the absence of money an internet company recommends actions to the agents, and in so doing reveals information about what the other agents have already learned. More generally, the approach of revealing information to agents in order to alter their behavior is known as Bayesian persuasion – there have been a recent number of papers in this area, inspired initially by [4] and [37].

5.1 Preliminaries

First we set out all the notation that we will need to cast designing experiments as a game between the employees, where the rules are set by the company.

Time. Time is discrete, and indexed by $t = \ldots, -1, 0, 1, \ldots$.

Employees. The firm consists of $N$ employees, indexed by $i = 1, \ldots, N$.

5.1.1 Properties of each A/B test

Experiment types. At each time step, each employee runs exactly one experiment. Experiments can be of one of $M$ types, denoted $x = 1, \ldots, M$. We use $x_{it}$ to denote the type of the experiment run by employee $i$ at time $t$.

Sample sizes. In addition to choosing experiment types, employees choose sample sizes for their experiments. We let $n_{it} \geq 0$ denote the sample size chosen by employee $i$ at time $t$.

Experiment designs. Thus $(x_{it}, n_{it})$ describes the experiment design chosen by $i$ at $t$. We let $E_t = \{(x_{it}, n_{it}), i = 1, \ldots, N\}$ denote the collection of designs chosen at time $t$.

Experiment effects. The experiment chosen by employee $i$ at time $t$ has an associated effect denoted $\theta_{it} \in \mathbb{R}$. This effect depends on the experiment type chosen; in particular, it is drawn from
a distribution $G(\omega_{x,t})$, where $G$ is a known family and $\{\omega_{xt}\}$ are unknown parameters in $\mathbb{R}^d$.

The employees begin the following common prior. The parameters $\omega_{xt}$ are independent across $x$, and for each $x$, $\{\omega_{xt}\}_{t=-\infty}^{\infty}$ is a stationary AR(1) process; in other words, for each $t$ there holds $\omega_{xt} \sim \mathcal{N}(0, I_d)$, and:

$$\omega_{x,t+1} = \sqrt{1 - \varepsilon^2} \omega_{xt} + \varepsilon N(0, I_d).$$

Thus $\{\omega_{xt}\}$ is the unique stationary solution to the preceding Markov system. We refer to $\varepsilon$ as the pace of innovation.

Finally, conditional on $(\omega_{1t}, \ldots, \omega_{Mt})$, the effects $\theta_{it}$ are independent across $i$ and effects at different time steps are independent of each other.

**Data.** The data generated by the experiment at time $t$ chosen by employee $i$ is $Y_{it}^1, \ldots, Y_{it}^n$. These are i.i.d. $\mathbb{R}^k$-valued random variables drawn from some distribution $F_{\theta_{it}}$ with everywhere positive density. Note that the data depends on the experiment type through $\theta_{it}$.

We use $Y_t = \{Y_{1t}^i, \ldots, Y_{nt}^i, i = 1, \ldots, N\}$ to denote all the data at time $t$.

**Decisions.** Inference in each experiment matters in so far as it lets the experimenter optimize some immediate decision. Associated to each experiment that is run, a decision is taken from a finite set $D$; we let $\delta_{it}$ denote the decision taken on the experiment run by employee $i$ at time $t$.

**Discount factor.** Future rewards are discounted by $\beta \in (0, 1)$.

**Rewards.** Each experiment generates a reward. In short, each employee looks to maximize the discounted sum of rewards incurred in her own experiments, while the company looks to maximize the total sum of discounted rewards across all experiments. More detail regarding the optimization problems faced by the employee and the company is given in Section 5.3.

We say that an experiment with sample size $n$, decision $\delta$, and true effect $\theta$ generates a base-level reward $R(n, \delta, \theta)$. Longer experiments are costly, so we assume that for each $\delta$ and $\theta$, $R$ is non-increasing in $n$. Further, we assume a normalization of these base-level rewards such that:

- $\sup_{n, \delta, \theta} R(n, \delta, \theta) \leq 1$.
- $\inf_{n, \delta, \theta} R(n, \delta, \theta) > -\infty$.
- There is at least one $\delta$ such that for all $n, \theta$, $R(n, \delta, \theta) \geq 0$.

**Impact factors.** To model private information by employees, we say that each experiment has an impact factor $L$; the experiment can be either high impact $(L = 1)$ or low impact $(L = 0)$. Informally, high impact experiments deliver higher rewards for the same parameters than low impact experiments. Specifically for each type $x$ that employee $i$ could choose at time $t$, there exists an impact factor $L_{ix,t}$, and so the factor for the type she chooses is $L_{i,x,t,t}$. Then we assume there exists a constant $\rho \geq 0$, such the reward earned by employee $i$ at time $t$ is given by

$$r_{it} = (1 + \rho L_{i,x,t,t}) R(n_{it}, \delta_{it}, \theta_{it}).$$  (5.1)
We make the key informational assumption that at time $t$, all $L_{i,t} = \{L_{i1t}, \ldots, L_{iMt}\}$ are observed by employee $i$ but none of these are seen by the other agents or by the company. We assume that the impact factors are IID across all $i, x$ and $t$. Further, for simplicity we model the distribution of each impact factor as $\text{Bernoulli}(1/2)$, although we note that similar results could be proven for any $\text{Bernoulli}(p)$. We denote the full collection of factors observed at $t$ by $L_{\cdot t} = \cup_{i=1}^{N} L_{i,t}$.

**Example 7.** Returning to the content aggregation service of Example 6:

- There are $M = 2$ types of treatment: web or mobile.
- The data points are whether each visitor shared some content, while the effect $\theta$ is the increase in share rate.
- Following the approach of [17], we could take $G$ to be a mixture of a centered normal distribution and a spike at zero. The unknown parameters $\omega$ are then the variance of the normal component and the mixing weight, both suitably transformed so that $\omega \sim N(0, I_2)$ is reasonable.
- On completion of each A/B test, there are two possible decisions: publish the content or abandon it. If the true effect is positive, the base-level reward favors publishing, whereas if the effect is negative it is better to abandon. The base-level reward penalizes larger $n$, as the employee wants to publish the content or give up as soon as possible.
- An impact factor $L = 1$ indicates that the employee knows that the content is ripe to go viral. Specifically the employee knows that each share will lead to $1 + \rho$ as many clicks as compared with lower impact content, resulting in a multiplicative increase in the payoff obtained.

### 5.1.2 Using the data

Now we can discuss how the data collected in each A/B test is utilized; namely how the company determines what data is revealed to the employees, how the employees use this data to update their beliefs about typical effects, and how inferences about the treatment effect in that single experiment are derived.

**The unfiltered data.** At the beginning of period $t$, the company has access to the unfiltered history, $h_t = \cup_{-\infty<\tau<t}(Y_\tau \cup E_\tau)$. We denote by $H_t$, the set of all possible unfiltered histories at time $t$ over any sequences of experiment designs and over any values for the observed data.

**The company’s policy.** At $t = 0$, all employees have access to the initial history, $h_0$. Thereafter the company adopts a policy $\sigma = (\sigma_t)_{t=1}^{\infty}$ for how data will be shared with employees going forwards. Specifically, each $\sigma_t$ is a function from $H_t$ to its power set, which indicates the set of data that is revealed to the employees at the beginning of that period. We impose the constraint that the policy must be invariant under permuting employees: if for some $t$, two histories $h_t$ and $\tilde{h}_t$ can be obtained from one another through some permutation of the employees, we require that $\sigma_t(h_t)$ and $\sigma_t(\tilde{h}_t)$ are related by the same permutation.
Filtered history. Between the data that the company has shared and the experiment designs, which we consider to be public, all of the publicly available data at the beginning of period $t$ is given by the filtered history,

$$f_t = h_0 \cup (\bigcup_{1 \leq \tau \leq t} \sigma_\tau(h_\tau)) \cup (\bigcup_{0 \leq \tau < t} E_\tau).$$

The mappings from unfiltered to filtered histories under the company’s chosen policy define a sequence of functions, $\Sigma_t(\sigma): H_t \rightarrow F_t(\sigma)$, where

$$F_t(\sigma) = \{h_0 \cup (\bigcup_{1 \leq \tau \leq t} \sigma_\tau(h_\tau)) \cup (\bigcup_{0 \leq \tau < t} E_\tau): h_t \in H_t\}.$$

Beliefs. At the beginning of period $t$, the beliefs of the employees over the treatment effect parameters $(\omega_{1t}, \ldots, \omega_{Mt})$ is given by the posterior for this vector after observing the filtered data, $f_t$. This is the measure $\nu_t$ where for any Borel sets $A_1, \ldots, A_M \in \mathbb{R}^d$,

$$\nu_t(A_1, \ldots, A_M) = \mathbb{P}(\omega_{xt} \in A_x \forall x | f_t).$$

For each $t$, the mapping from $f_t$ to $\nu_t$ defines a function

$$N_t(\cdot; \sigma): F_t(\sigma) \rightarrow \mathcal{B}(\mathbb{R}^{d \times M}),$$

where $\mathcal{B}(\mathbb{R}^{d \times M})$ denotes the set of Borel measures on $\mathbb{R}^{d \times M}$.

From the definitions, we see that the composition $N_t \circ \Sigma_t$ gives the mapping from the unfiltered history at each time to the employees’ belief. We note here that for any $\sigma$, $F_0(\sigma) = H_0$, $\Sigma_0(\sigma)$ is the identity and $\nu_0 = N_0(h_0)$.

Local mechanisms. In this chapter we restrict our attention to mechanisms of the following form: $\sigma_t(h_t) = \sigma(\nu_{t-1}, E_{t-1}, Y_{t-1}) \subset Y_{t-1}$. We refer to these as local mechanisms. The mechanism takes as input only the beliefs, experiment designs and data collected at the previous period, and applies some time-invariant mapping $\sigma$ to obtain a subset of the data collected at the previous period, which is then shared with the employees. We refer to $Y^f_{t-1} = \sigma(\nu_{t-1}, E_{t-1}, Y_{t-1})$ as the filtered data for period $(t-1)$ and note that $f_t = h_0 \cup (\bigcup_{0 \leq \tau < t} Y^f_\tau \cup E_\tau)$.

The non-filtering mechanism. Throughout this chapter, it will be useful to refer to one particularly obvious local mechanism. The non-filtering mechanism reveals all new data at each time: $\sigma^0(\nu_{t-1}, E_{t-1}, Y_{t-1}) = Y_{t-1}$. In this case, $F_t(\sigma^0) = H_t$ holds for all $t$, each $\Sigma_t$ is the identity, and $\nu_t = N_t(h_t; \sigma^0)$.

Bayes-optimal decision making. On completion of any A/B test, the company chooses the decision to maximize the base-level reward in expectation over the posterior for the treatment effect, given the filtered history up to that time and all the data collected in that test. For each $i$ and $t$,
δ = δ_{it}(ν_{t}, x_{it}, n_{it}, Y_{1}^{it}, \ldots Y_{n_{it}}^{it}) maximizes

\mathbb{E}[R(n_{it}, δ, θ_{ix_{it}})|f_{t}, x_{it}, n_{it}, Y_{1}^{it}, \ldots Y_{n_{it}}^{it}] = \int \mathbb{E}[R(n_{it}, δ, θ)|ω, x_{it}, n_{it}] d\hat{ν}_{t}(ω),

where \hat{ν}_{t} is the Bayesian update from ν_{t} after observing Y_{1}^{it}, \ldots Y_{n_{it}}^{it}.

Note that:

• δ_{it} can be allowed to depend on all the data collected in this test, not just the subset that will be revealed to employees, because the decision is implemented by the company.

• Maximizing the base-level reward actually maximizes the realized reward r_{it} whether L_{i,x_{it},t} = 0 or 1, because the impact factor only affects this reward multiplicatively.

The fixed-design expected base-level reward. Once employee i has fixed her design at time t, her expected reward from that experiment over both her uncertainty about effect sizes and the randomness in the data collected at time t is given by the product of (1 + ρ L_{ix_{it},t}) and

R_{0}(ν_{t}, x_{it}, n_{it}) := \mathbb{E}[R(n_{it}, δ, θ_{it})|f_{t}, x_{it}, n_{it}] = \int \mathbb{E}[R(n_{it}, δ, θ)|ω, x_{it}, n_{it}] dν_{t}(ω).

We refer to R_{0}(ν, x, n) as the fixed-design expected base-level reward, given those parameters.

5.1.3 Summary

At each time step t, the following sequence of events unfolds.

1. The company has access to the unfiltered data, h_{t}.

2. The company chooses the filtered dataset Y_{t-1}^{f} = σ(ν_{t-1}, E_{t-1}, Y_{t-1}) ⊂ Y_{t-1}. This is revealed to the employees, whose filtered history is now f_{t}.

3. Each agent i (privately) observes L_{i,t}.

4. Each agent i simultaneously chooses x_{it} and n_{it}.

5. The company observes the data Y_{t}.

6. The company takes decisions d_{it} so as to maximize the posterior expected rewards.

5.1.4 Limitations

In general we have prioritized a model where we can analyze the externalities between employees, which is as simple as possible while allowing for the possibility that the employees have some private information. However, we should take this moment to note some limitations to our model:
The model supposes that the pace of experimentation is limited by the employees’ time, not by the quantity of data available. That is, we assume that it takes one unit of time for the employee to identify a new treatment and to set up an A/B test, so that each employee can only run one experiment at each period. On the other hand, while larger sample sizes are considered more costly, there is no global budget constraint on the number of samples available. Such a model is reasonable for the company described in Example 1, if it takes employees time to find new content but the service receives a large number of visitors who can be used in any A/B test. For some companies, however, identifying new treatments is comparatively trivial while data is more expensive.

We model the private information as a multiplicative factor on the reward obtained for any fixed value of the treatment effect. In reality, there are many aspects of the experiment where the employees may have some private information; e.g. they may have additional information about the value of the effect.

We suppose that the employees only update their beliefs about typical effect sizes using the filtered data revealed to them, \( f_t \). In the next section, we will formalize this as a bounded rationality assumption on the employees’ optimization problem. In reality an experimenter may be able to learn about typical effect sizes from other sources, such as the decision \( \delta_t \) that the company takes on completion of her A/B test and the reward obtained, \( r_t \). Nonetheless, this is reasonable if the filtered data is communicated to employees on a clear interface, while reasoning from other data sources would require a sophisticated level of understanding regarding how any decisions or rewards depend on the treatment effects.

### 5.2 The experiment design game

In this section we characterize the game between the employees and what that means for the company. We focus on *boundedly rational Markov perfect equilibria* (BR-MPE), where the experiment designs chosen at each period depend only on the current belief and the employee’s privately observed impact factors for that period. The goal for the company is to select a mechanism for filtering the data which maximizes welfare at the most favorable BR-MPE. We benchmark the performance of any mechanism against the payoﬀ the company would achieve if it could reveal the unﬁltered data, could ensure that all experiments are high-impact, and could force every employee to play a common strategy of its choosing, which maps the history at each time to an experiment design. Proofs of the results in this section are given in Appendix F.
5.2.1 The objectives of the agents

Strategies. At each time, an agent maps the filtered history and her vector of private impact factors to a choice of experiment design. Formally, once the company has fixed its policy $\sigma$, a strategy for agent $i$ is two collections of functions, indexed by $t = 0, 1, \ldots$:

$$\chi_{it} : F_t(\sigma) \times \{0, 1\}^M \to \{1, \ldots, M\}, \quad \eta_{it} : F_t(\sigma) \times \{0, 1\}^M \to \mathbb{N}.$$  

At time $t$, the experiment type chosen by agent $i$ under this strategy is $x_{it} = \chi_{it}(f_t, L_{i:t})$ and the sample size is $n_{it} = \eta_{it}(f_t, L_{i:t})$.

Note that we do not allow mixed strategies; given $f_t$ and $L_{i:t}$, the experiment designs selected are deterministic.

Public strategies. A public strategy is one where the agent ignores her private information. These public strategies will be central in how we define the benchmark for the company.

Formally, a strategy under $\sigma$ is public if there exist functions

$$\tilde{\chi}_{it} : F_t(\sigma) \to \{1, \ldots, M\}, \quad \tilde{\eta}_{it} : F_t(\sigma) \to \mathbb{N},$$

such that for all $t, f_t \in F_t(\sigma)$ and $L \in \mathbb{R}^M$, $\chi_{it}(f_t, L) = \tilde{\chi}_{it}(f_t)$ and $\eta_{it}(f_t, L) = \tilde{\eta}_{it}(f_t)$. Under this public strategy, the chosen experiment type and sample size at time $t$ are $x_{it} = \tilde{\chi}_{it}(f_t)$ and $n_{it} = \tilde{\eta}_{it}(f_t)$ respectively. For public strategies, we will abuse notation and refer to the collections $(\chi_{it})_{t=0}^\infty$ and $(\eta_{it})_{t=0}^\infty$, or the collections $(\tilde{\chi}_{it})_{t=0}^\infty$ and $(\tilde{\eta}_{it})_{t=0}^\infty$, interchangeably as the strategy.

Strategies of other agents. We denote the collection of strategies played by the $N-1$ agents besides $i$ by $(\chi_{-i}, \eta_{-i})$. In the case where all other agents play some common strategy $(\tilde{\chi}_{it}, \tilde{\eta}_{it})$, we will denote that by $\chi_{-i} = \tilde{\chi}^{N-1}, \eta_{-i} = \tilde{\eta}^{N-1}$.

The boundedly rational agent’s optimization problem. Suppose that the platform has chosen policy $\sigma$ and the agents other than $i$ have chosen strategies $(\chi_{-i}, \eta_{-i})$ under $\sigma$. Informally, agent $i$ chooses a strategy to optimize her expected discounted payoff, assuming her beliefs evolve according to the filtered data revealed to her, $f_t$.

Formally, for each initial history $f_0 = h_0$ and each vector of initial impact factors $L_{i:0}$, and for fixed policies $\sigma$ and $(\chi_{-i}, \eta_{-i})$, agent $i$ chooses $(\chi_{it})_{t=0}^\infty$ and $(\eta_{it})_{t=0}^\infty$ to maximize:

$$V_{i0}(f_0, L_{i:0}; (\chi_{it}), (\eta_{it}), \chi_{-i}, \eta_{-i}, \sigma) = (1 - \beta)\mathbb{E}_{(\chi_{it}), (\eta_{it}), \chi_{-i}, \eta_{-i}, \sigma} \left[ \sum_{t=0}^\infty \beta^t r_{it} \left| f_0, L_{i:0} \right. \right] \quad (5.2)$$

By standard theory on dynamic programming, we see that once $(\chi_{it})_{t=0}^\infty$ and $(\eta_{it})_{t=0}^\infty$ are chosen optimally, starting at any time $t$, these functions will maximize the continuation value,

$$V_{it}(f_t, L_{i:t}; (\chi_{it})_{t\geq t}, (\eta_{it})_{t\geq t}, \chi_{-i}, \eta_{-i}, \sigma) = (1 - \beta)\mathbb{E}_{(\chi_{it}), (\eta_{it}), \chi_{-i}, \eta_{-i}, \sigma} \left[ \sum_{\tau=t}^\infty \beta^{\tau-t} r_{i\tau} \left| f_t, L_{i:t} \right. \right] \quad (5.3)$$
5.2.2 Boundedly rational Markov perfect equilibrium

In short, boundedly rational Markov perfect equilibria are equilibria where each agent plays a strategy that only uses the filtered data $f_t$ through the public belief $\nu_t = N_t(f_t)$ that it induces. We restrict attention to stationary, symmetric BR-MPE; here all agents play a common strategy, where the mapping from the belief and the private information to an experiment design does not depend on the time step $t$.

Stationary Markov strategies. A stationary Markov strategy is one where the experiment design chosen at each time only depends on the filtered history $f_t$ as a time-invariant function of the belief distribution $\nu_t$ that it generates. Formally, a strategy under $\sigma$ is stationary and Markov if there exist functions $\chi_i: B(\mathbb{R}^{d\times M}) \times \{0,1\}^M \to \{1,\ldots,M\}$, $\eta_i: B(\mathbb{R}^{d\times M}) \times \{0,1\}^M \to \mathbb{N}$, such that for all $t,f_t \in F_t(\sigma)$ and $L \in \mathbb{R}^M$,

$$\chi_{it}(f_t, L) = \chi_i(N_t(f_t), L), \quad \eta_{it}(f_t, L) = \eta_i(N_t(f_t), L).$$

In this case, the experiment design at time $t$ is given by $x_{it} = \chi_{i}(\nu_{t+1}, L_{it})$, $n_{it} = \eta_{i}(\nu_{t+1}, L_{it})$. For stationary Markov strategies, we will abuse notation and refer to the pair $(\chi_i)_{t=0}^{\infty}$ and $(\eta_i)_{t=0}^{\infty}$ or the pair $(\chi_i, \eta_i)$ equivalently as the strategy.

By first mapping $f_t$ to $\nu_t$, these strategies reduce the filtered history, which grows over time, to a belief distribution that lives in the time-independent set $B(\mathbb{R}^{d\times M})$. It is natural to focus on such strategies because $\nu_t$ captures all of the relevant information in $f_t$ about typical treatment effects at the current time. Moreover, reasoning about such strategies will prove more straight-forward, because the beliefs $(\nu_t)$ evolve as a Markov process with a time-invariant transition kernel, which depends only on the experiment designs chosen at the current time period. This is established in the next proposition. In the next section, the Markov structure will let us characterize the objective of each agent in equilibrium in terms of a time-invariant dynamic program.

Proposition 11. For a given $t \geq 0$, fix the experiment designs of all agents, $E_t$, and a local mechanism $\sigma$. Then there are a family of probability measures $P(\cdot; \nu, \sigma, E_t)$ such that $\nu_{t+1}|f_t \sim P(\cdot; \nu_t, \sigma, E_t)$ a.s.

The proof of this result in Appendix B provides a formal specification of the transition probabilities $P$. Heuristically the kernel $P(\cdot; \nu, \sigma, E_t)$ captures two competing effects. Compared with $\nu_t$, $\nu_{t+1}$ is “flattened” towards a standard multivariate normal distribution, because the AR(1) process adds irreducible noise to each $\omega_{xt,t+1}$. On the other hand, $\nu_{t+1}$ contracts because the filtered data that is obtained under $E_t$ and $\sigma$ provides information about $\omega_{xt}$, which may be extrapolated to $\omega_{xt,t+1}$.

Stationary Markov public strategies. Combining the definitions, we see that a strategy is
stationary, Markov and public if there exist functions
\[ \chi_i : \mathcal{B}(\mathbb{R}^{d \times M}) \to \{1, \ldots, M\}, \quad \eta_i : \mathcal{B}(\mathbb{R}^{d \times M}) \to \mathbb{N}, \]
such that for all \( t, f_t \in F_t(\sigma) \) and \( L \in \mathbb{R}^M, \)
\[ \chi_{it}(f_t, L) = \chi_i(N_t(f_t)), \quad \eta_{it}(f_t, L) = \eta_i(N_t(f_t)). \]

The experiment design at time \( t \) is given by
\[ x_{it} = \chi_i(\nu_t), \quad n_{it} = \eta_i(\nu_t). \]

The stationary value function. This next proposition establishes that if all agents play stationary, Markov strategies, the continuation payoff for any employee starting at any time is a time-invariant function of the public belief and her private information at that time. We refer to this function \( W_i \) as her stationary value function.

**Proposition 12.** Fix an agent \( i \), a stationary, Markov strategy \((\chi_i, \eta_i)\), a collection of stationary, Markov strategies \((\chi_{-i}, \eta_{-i})\), and a local mechanism \( \sigma \). Then there exists a function
\[ W_i(\cdot ; \chi_i, \eta_i, \chi_{-i}, \eta_{-i}, \sigma) : \mathcal{B}(\mathbb{R}^{d \times M}) \times \{0, 1\}^M \to \mathbb{R}, \]
which takes as inputs a belief distribution and a vector of impact factors, such that for all \( t, f_t \) and \( L, \)
\[ V_{it}(f_t, L; \chi_i, \eta_i, \chi_{-i}, \eta_{-i}, \sigma) = W_i(N_t(f_t), L; \chi_i, \eta_i, \chi_{-i}, \eta_{-i}, \sigma). \]

For later reference it will be useful to define the expected payoff for agent \( i \) before she observes her impact factors for the period: for any \( \nu \in \mathcal{B}(\mathbb{R}^{d \times M}), \)
\[ \bar{W}_i(\nu; \chi_i, \eta_i, \chi_{-i}, \eta_{-i}, \sigma) = 2^{-M} \sum_{L \in \{0, 1\}^M} W_i(\nu, L; \chi_i, \eta_i, \chi_{-i}, \eta_{-i}, \sigma). \]

Note that \( 2^{-M} \) appears here because each element of the \( M \)-dimensional vector \( L \) is an independent Bernoulli(1/2) random variable.

**Stationary, symmetric BR-MPE.** Fix the platform’s policy \( \sigma \) and a single stationary, Markov strategy \((\chi, \eta)\) with respect to \( \sigma \). This strategy defines a boundedly rational Markov perfect equilibrium under \( \sigma \) if, for an arbitrary choice of \( i \), \((\chi, \eta)\) maximizes the payoff for agent \( i \) given in (D1), when all other agents play \( \chi_{-i} = \chi^{N-1}, \eta_{-i} = \eta^{N-1} \). In our analysis we will focus on these stationary, symmetric BR-MPE, where all agents play the same strategy and the common strategy is stationary.

Note that at a stationary, symmetric BR-MPE, the stationary value function is the same for all agents. It is given by
\[ W_1(\nu, L; \chi^N, \eta^N, \sigma) := W_1(\nu, L; \chi, \eta, \chi^{N-1}, \eta^{N-1}, \sigma) \]
(of course the choice of agent 1 here is arbitrary). Before observing impact factors, the expected payoff for any agent is

$$W_1(\nu; \chi^N, \eta^N, \sigma) := W_1(\nu; \chi, \eta, \chi^{N-1}, \eta^{N-1}, \sigma).$$

5.2.3 The objective for the company

We have detailed how the agents select their strategies once the company’s policy, $\sigma$, is fixed. Next we explain how the company looks to optimize $\sigma$. Informally, the company wants a policy that maximizes the per-employee discounted reward, assuming that the agents adopt a stationary, symmetric BR-MPE in response.

**The company’s payoff.** Fix an initial history $h_0$, a policy $\sigma$, and a strategy $(\chi_t)_{t=0}^\infty$, $(\eta_t)_{t=0}^\infty$ under that policy. The payoff for the company when all employees play this single strategy is:

$$U_0(h_0; (\chi_t), (\eta_t), \sigma) = \left(\frac{1 - \beta}{N}\right) \mathbb{E}_{(\chi_t)^N, (\eta_t)^N, \sigma} \left[ \sum_{t=0}^\infty \sum_{i} \beta^t r_{i1} \bigg| h_0 \right]$$

$$= (1 - \beta) \mathbb{E}_{(\chi_t)^N, (\eta_t)^N, \sigma} \left[ \sum_{t=0}^\infty \beta^t r_{11} \bigg| h_0 \right].$$

Here we use the notation $(\chi_t)^N, (\eta_t)^N$ to indicate that all $N$ agents play $(\chi_t), (\eta_t)$.

For any $t$ and $h_t \in H_t$, the continuation payoff for the company under $\sigma$ and this common strategy is:

$$U_t(h_t; (\chi_t)_{\tau \geq t}, (\eta_t)_{\tau \geq t}, \sigma) = \left(\frac{1 - \beta}{N}\right) \mathbb{E}_{(\chi_t)^N, (\eta_t)^N, \sigma} \left[ \sum_{\tau=t}^\infty \sum_{i} \beta^{\tau-t} r_{i1} \bigg| h_t \right]$$

$$= (1 - \beta) \mathbb{E}_{(\chi_t)^N, (\eta_t)^N, \sigma} \left[ \sum_{\tau=t}^\infty \beta^{\tau-t} r_{11} \bigg| h_t \right].$$

**The company’s optimization problem.** For each $h_0 \in H_0$, the company solves the following problem:

$$\text{maximize } U_0(h_0; \chi, \eta, \sigma) \quad \text{subject to } (\chi, \eta) \text{ is a stationary, symmetric BR-MPE.}$$

With this definition, if there are multiple stationary, symmetric BR-MPE for a given $\sigma$, we allow the company to choose the most favorable equilibrium.

**Proposition 13.** For any $h_0 \in H_0$, any local mechanism $\sigma$, and any stationary, Markov strategy $(\chi, \eta)$ under $\sigma$,

$$U_0(h_0; \chi, \eta, \sigma) = \bar{W}_1(N_0(h_0); \chi^N, \eta^N, \sigma).$$
Thus the goal for the company is to maximize in equilibrium the stationary value function for each agent, evaluated at the initial belief at \( t = 0 \) before any impact factors are observed, \( \nu_0 = N_0(h_0) \).

**The benchmark payoff.** In this chapter, we do not offer a global solution to the company’s optimization problem. We simply consider a benchmark for what the company could achieve if it could dictate how employees design their experiments, and against this we measure the performance of our proposed mechanism.

For \( \rho = 0 \), the benchmark payoff is defined as the best achievable payoff over all \( \sigma \) and all public strategies under \( \sigma \):

\[
U^*_0(h_0) = \max_{(\chi_t), (\eta_t), \sigma} U_0(h_0; (\chi_t), (\eta_t), \sigma)
\]

For \( \rho > 0 \), we need a higher benchmark that considers how employees might be motivated to tailor designs to the private impact factors they observe. We note that at best the employees could ensure that a high-impact experiment is selected every time, and then optimize the sample sizes accordingly. This is the payoff that the company could achieve by designing experiments centrally, if it were assured that every impact factor were one (rather than one with probability 1/2).

For any strategy \((\chi_t), (\eta_t)\) under \( \sigma \), we define the high-impact payoff:

\[
U^\uparrow_0(h_0; (\chi_t), (\eta_t), \sigma) = (1 - \beta)(1 + \rho) E_{(\chi_t), (\eta_t), \sigma} \left[ \sum_{t=0}^{\infty} \beta^t R(n_1 t, \delta_t, \theta_t) \bigg| h_0 \right].
\]

The benchmark is the maximum high-impact payoff over all \( \sigma \) and all public strategies:

\[
U^*_0(h_0) = \max_{(\chi_t), (\eta_t), \sigma} U^\uparrow_0(h_0; (\chi_t), (\eta_t), \sigma).
\]

While the benchmark considers all possible mechanisms, the following proposition clarifies that the maximum is attained at the non-filtering mechanism, \( \sigma^0 \). That is to be expected because, if the company could control how employees design experiments, it would have no reason to withhold data. We see also that the benchmark payoff depends on the initial history only through the initial belief that it generates.

**Proposition 14.** 1. For all \( h_0 \),

\[
U^*_0(h_0) = \max_{(\chi_t), (\eta_t), \sigma} U^\uparrow_0(h_0; (\chi_t), (\eta_t), \sigma^0),
\]

where the max is taken over all public strategies under \( \sigma^0 \).

2. There is a function \( W^* : \mathcal{B}(\mathbb{R}^{d \times M}) \rightarrow \mathbb{R} \), such that for all \( h_0 \in H_0 \),

\[
U^*_0(h_0) = W^*(N_0(h_0)).
\]
The benchmark continuation payoff. For any \( t \) and \( h_t \in H_t \), we define the continuation high-impact payoff,

\[
U^*_t(h_t; (\chi_t)_{t \geq t}, (\eta_t)_{t \geq t}, \sigma) = (1 - \beta)(1 + \rho) \mathbb{E}_{(\chi_t)_{N_t}, (\eta_t)_{N_t}, \sigma} \left[ \sum_{\tau = t}^{\infty} \beta^{\tau - t} R(n_{1\tau}, \delta_{1\tau}, \theta_{1\tau}) \bigg| h_t \right],
\]

Thus we obtain the benchmark continuation payoff,

\[
U^*_t(h_t) = \max_{(\chi_t)_{t \geq t}, (\eta_t)_{t \geq t}} U^*_t(h_t; (\chi_t), (\eta_t), \sigma^0),
\]

where the max is taken over all public strategies under \( \sigma^0 \).

**Proposition 15.** For all \( t \) and \( h_t \in H_t \),

\[
U^*_t(h_t) = W^*(N_t(h_t; \sigma^0)),
\]

where \( W^* \) is the same function as was defined in Proposition 14.

### 5.3 Exploration in equilibrium

Focusing on BR-MPE makes analysis of the experiment design game tractable, because we can characterize the strategy of each agent as the solution to a time-invariant dynamic program. Further, the benchmark payoff may be realized as the solution to another dynamic program. Comparison of the two DPs lets us highlight how the non-filtering mechanism leads to under-exploration by employees in equilibrium, and it lets us motivate a solution – the Minimum Learning Mechanism – which we will describe in the next section. Proofs of the results in this section are given in Appendix F.

#### 5.3.1 BR-MPE

In equilibrium, each agent trades off the myopic reward from the single experiment at each period against the evolution of the beliefs at the next period, while viewing the designs of all other agents as held fixed. Given a public belief \( \nu \) and impact factors \( L \), if she chooses the design \((x, n)\), her myopic reward is \((1 - \beta)(1 + \rho L_x)R_0(\nu, x, n)\). Her expected payoff from the next period onwards is the common stationary value function before observing impact factors, evaluated at the random next state when she chooses \((x, n)\) but each of her colleagues choose the equilibrium strategy evaluated at her own vector of impact factors. Thus for a common strategy to define a BR-MPE, it must solve the dynamic program stated in the following proposition.

**Proposition 16.** Let \( \sigma \) be a local mechanism and let \((\chi, \eta)\) be a Markov strategy.
1. Suppose that for all \( \nu \in B(\mathbb{R}^{d \times M}) \) and \( L = (L^1, \ldots, L^M) \in \{0, 1\}^M \), the pair \( x = \chi(\nu, L), n = \eta(\nu, L) \) maximizes the following expression:

\[
(1 - \beta)(1 + \rho L^x)R_0(\nu, x, n) + 2^{-M(N-1)} \beta \sum_{L_2, \ldots, L_N \in \{0, 1\}^M} \mathbb{E}_1(\nu' \mid \chi^N, \eta^N, \sigma) \mathbb{P}(\nu' \mid \nu, \sigma, E(x, n, \nu, L_2, \ldots, L_N))
\]

where

\[
E(x, n, \nu, L_2, \ldots, L_N) = \{x, n, \chi(\nu, L_2), \eta(\nu, L_2), \ldots, \chi(\nu, L_N), \eta(\nu, L_N)\}.
\]

Then \((\chi, \eta)\) defines a stationary, symmetric BR-MPE under \( \sigma \).

2. More generally, suppose that there exists a function \( w(\nu, L) \), such that the triple \( x = \chi(\nu, L), n = \eta(\nu, L) \) and \( w(\nu, L) \) solve the dynamic program:

\[
w(\nu, L) = \max_{x, n} \left\{ (1 - \beta)(1 + \rho L^x)R_0(\nu, x, n) + 2^{-M(N-1)} \beta \sum_{L_2, \ldots, L_N \in \{0, 1\}^M} \mathbb{E}_1(\nu' \mid \chi^N, \eta^N, \sigma) \mathbb{P}(\nu' \mid \nu, \sigma, E(x, n, \nu, L_2, \ldots, L_N)) \right\}
\]

where \( w(\nu) = 2^{-M} \sum_{L \in \{0, 1\}} w(\nu, L) \). Then \((\chi, \eta)\) defines a stationary, symmetric BR-MPE under \( \sigma \). The stationary value function is given by \( W_1(\nu, L; \chi^N, \eta^N, \sigma) = w(\nu, L) \).

5.3.2 Characterizing the benchmark

The benchmark payoff for the company can be expressed as the value function of a different dynamic program. In this case, we are interested in identifying a strategy that maximizes the payoff for the company when it is played by all agents. Such an optimal common strategy must balance the myopic reward for each employee against the future value obtained when all employees choose larger sample sizes and riskier types of treatments.

**Proposition 17.** 1. The function \( W^* \) defined in Proposition 14 is the solution to the following dynamic program:

\[
W^*(\nu) = \max_{x, n} \left\{ (1 - \beta)(1 + \rho L^x)R_0(\nu, x, n) + \beta \int W^*(\nu') \mathbb{P}_0(\nu' \mid \nu, \{x, n\}^N) \right\}
\]

2. Let \( \chi^{opt}(\nu), \eta^{opt}(\nu) \) be the public, Markov strategy given by the solution to this dynamic program. For any initial history \( h_0 \), the benchmark payoff is given by

\[
U_0^*(h_0) = U_0^*(h_0; (\chi^{opt})^N, (\eta^{opt})^N, \sigma^0) = W^*(N_0(h_0)).
\]
5.3.3 Under-exploration

To get intuition for why the non-filtering mechanism motivates under-exploration, let’s focus on the case of $\rho = 0$. In that case, the instantaneous reward in the dynamic programs of Propositions 16 and 17 coincide. Where the dynamic programs differ is in the transition probabilities, over which the expected value at the next period is evaluated. In particular, any modification to the choice of experiment design has an $N$-fold impact on the next state in the DP of Proposition 17, as compared with the DP of Proposition 16 with $\sigma = \sigma^0$. Trading off the same instantaneous reward against a much smaller impact on the next state motivates a far more myopic design.

To solve this issue, the company requires a mechanism where deviations from the benchmark strategy by a single agent have an $N$-fold impact on beliefs at the next period.

5.4 The Minimum Learning Mechanism

We now present the Minimum Learning Mechanism. In the absence of private information, we find that the MLM has a BR-MPE that achieves precisely the benchmark payoff. We study the case of private information when $M = 1$; i.e. where employees choose sample sizes but not treatment types. In this regime, we obtain a bound for the welfare loss in terms of $\rho$. Proofs of the results in this section are included in Appendix G.

In short, the MLM only reveals data on experiments where the treatment was of the least risky type chosen in that period; among such experiments, it only reveals a number of observations equal to the minimum sample size chosen in that period.

**Definition 8.** Given the prior $\nu_{t-1}$ and experiment designs $E_{t-1}$ at period $(t - 1)$, we identify the minimum sample size, $n_{MLM}^{t-1} = \min_i n_{i,t-1}$, and the least risky type,

$$x_{MLM}^{t-1} = \arg\max_{x_{i,t-1}, n_{i,t-1}} R_0(\nu_{t-1}, x_{i,t-1}, n_{i,t-1}).$$

The Minimum Learning Mechanism is the following local mechanism:

$$\sigma^{MLM}(\nu_{t-1}, E_{t-1}, Y_{t-1}) = \{y_k^{i,t-1} : x_{i,t-1} = x_{MLM}^{t-1}, k \leq n_{MLM}^{t-1} \}$$

5.4.1 No private information

Suppose that $\rho = 0$, the company implements the MLM, and all agents play the benchmark strategy $(\chi^{opt}, \eta^{opt})$. In that case, at each time $t$, every experiment run is of type $\chi^{opt}(\nu)$ and so data on all $N$ of them is shared with the employees at $t + 1$. Further, all of these experiments have the same sample size, $\eta^{opt}(\nu)$, so every data point from these experiments is shared. The upshot is that this scenario would achieve the benchmark payoff for the company.
Crucially, the MLM is designed so that a deviation by a single agent to a more myopic design at some time $t$ results in a N-fold loss in the data that is shared at $t+1$. A deviation to a less risky type would reduce the number of experiments on which data is shared from $N$ to one; a reduction by one in the sample size would result in one less data point being shared from each of the $N$ experiments. Consequently the benchmark strategy defines a BR-MPE.

**Theorem 8.** Suppose $\rho = 0$ and the company implements the MLM. Then $(\chi^{\text{opt}}, \eta^{\text{opt}})$ defines a stationary, symmetric BR-MPE. Moreover, at this equilibrium, the company achieves the benchmark payoff. That is, for any initial history $h_0$,

$$U^*_0(h_0) = U_0(h_0; \chi^{\text{opt}}, \eta^{\text{opt}}, \sigma^{\text{MLM}}).$$

### 5.4.2 Performance for general $\rho$

The MLM produces some welfare loss in the case of private information, because at each period it forces all employees to embark on the same level of exploration as each other. A larger payoff for the company would be obtained, if agents could explore more in periods when they observe low impact factors and exploit more when they have the chance to run a high impact experiment.

For our analysis, we restrict to the case of $M = 1$. When an agent observes $L = 1$, what sample size would she like all employees to select for that period? To optimize her reward, she must balance a myopic payoff of $(1 - \beta)(1 + \rho)R_0(\nu, n)$ against the expectation of the stationary value function before observing her next impact factor, evaluated at the random next state when all agents select $n$. We refer to this as the high-impact sample size, and it is the sample size that she will select in equilibrium under the MLM. When an agent observes $L = 0$, her preference would be for all employees to select the sample size that trades off her smaller myopic payoff of $(1 - \beta)R_0(\nu, n)$ against the expectation of the stationary value function at the next state. However, this is not the sample size that the MLM motivates her to select. Rather she knows that, in all likelihood, there will be another agent who has observed $L = 1$ for this period and who will be playing the high-impact sample size. This is smaller (i.e. more myopic) than the sample size she would have selected, and as the MLM filters out any data above the lowest chosen sample size, she will reduce her own to match the choice of her colleague.

Note that when all agents play high-impact sample sizes, the expected myopic reward for each agent at each period before $L$ is observed is given by $(1 - \beta)(1 + \rho/2)R_0(\nu, n)$, where $n$ is the high-impact sample size that corresponds to the current belief, $\nu$. Hence the stationary value function before observing the impact factor satisfies the following recursion: the function evaluated at any $\nu$ is equal to this average myopic reward at $\nu$ plus $\beta$ times the expectation of the function evaluated at the next state.

**Definition 9.** Suppose $M = 1$. The high-impact strategy $\eta^*(\nu)$ is the solution to the following
**Proposition 18.** Suppose $M = 1$. Then under the MLM, $\eta^\dagger$ defines a BR-MPE. For this equilibrium, the stationary value function before impact factors are observed is given by

$$\bar{W}_1^\dagger(\nu; (\eta^\dagger)^N, \sigma^{MLM}) = \bar{W}_1^\dagger(\nu).$$

The high-impact sample sizes may not be the unique equilibrium under the MLM. In fact, we conjecture that for small $N$, there are additional equilibria that embark on greater exploration and so produce larger payoffs for the company. For when $N$ is small, there is a non-trivial probability that all agents will observe $L = 0$ at any given period. We would expect this to lead to an equilibrium, where agents select the high-impact sample size when $L = 1$ but they select a slightly larger sample size when $L = 0$, betting on the possibility that all other agents will observe $L = 0$ and this extra data will not be filtered out.

Nonetheless, the high-impact sample sizes define a BR-MPE where the payoff for the company is close to the benchmark. Proposition 13 establishes that the payoff for the company is

$$U_0(h_0; \eta^\dagger, \sigma^{MLM}) = \bar{W}_1^\dagger(N_0(h_0); (\eta^\dagger)^N, \sigma^{MLM}) = \bar{W}_1^\dagger(N_0(h_0)),$$

and so the welfare loss is the gap, $U_0^\ast(h_0) - \bar{W}_1^\dagger(N_0(h_0))$. Arguing directly from the definition of $\eta^\dagger$ and $\bar{W}_1^\dagger$ (see appendix), we obtain the following welfare bound.

**Theorem 9.** For any $h_0 \in H_0$,

$$U_0(h_0; \eta^\dagger, \sigma^{MLM}) \geq U_0^\ast(h_0) - \rho/2.$$

Theorem 4 does not say anything positive about the MLM when employees have a large amount of private information. Rewards are at most one, so the benchmark payoff lies between zero and one, as does the payoff for the company under any common strategy for the agents. Thus if $\rho \geq 2$, the theorem is vacuous. However, this result shows how the loss degrades smoothly to zero when $\rho$ is small. Returning to the example of the content aggregation service that we have considered throughout this chapter, it seems unlikely that the employees would have very strong private information about which content will go viral. If $\rho = 0.1$, perhaps, so that employees identify high-impact content that garners 10% more clicks on average than low-impact content, we can say that the MLM would achieve close to optimal welfare.
5.5 Discussion

In this chapter we have addressed how a company can motivate its employees to design A/B tests in a way that furthers long-term company-wide optimization – for this we offer a simple mechanism for how the company should share data with the experimenters. This Minimum Learning Mechanism aligns incentives perfectly in the absence of private information, and is robust when the employees draw on a small amount of private knowledge to inform their experiment designs.

The MLM can be applied in a broad range of contexts:

- We have imposed no restrictions on the structure of the company, except to say that at any time the treatment effect in an experiment \( \theta_{it} \) can be modeled as an IID draw from a distribution, which is indexed by some parameters \( \omega_{xt} \) that are common to all employees.

- We make only weak restrictions on the distributions of the treatment effects and the observations in each experiment. We just assume that computing the posterior for the effect is tractable, so that a Bayesian analysis is feasible.

- We are agnostic to why the experimenters want inference on these treatment effects. We suppose simply that the goal is to optimize some decision that is taken immediately, the reward associated with any effect under any decision is fully understood, and maximizing the posterior expected reward is computationally feasible.

However, the theory presented in this chapter leaves some open questions.

5.5.1 The price of anarchy

The MLM achieves close to the benchmark welfare when \( \rho \) is small, whereas the naive approach of revealing all experimental data to the employees leads to under-exploration and a lower payoff for the company. Before implementing the MLM, though, it is reasonable to ask how the large the welfare gap is between these two options.

In fact, even if all data is revealed to her, an employee will still embark on some exploration. At worst she will choose a sample size \( n \) that provides a sufficiently good estimate of the treatment effect for her to optimize her immediate decision; i.e. the sample size that maximizes her myopic reward \( R_0(\nu, x, n) \). Similarly, her private information will lead her to keep trying out different treatment types, depending on which type offers a high impact at each period. While the MLM is seen to work for small \( \rho \), this free exploration is most pronounced at larger \( \rho \). In particular, if

\[
\max_{\nu} \left\{ \frac{\max_x \{\max_n R_0(\nu, x, n)\}}{\min_x \{\max_n R_0(\nu, x, n)\}} \right\} \leq 1 + \rho, \quad (5.7)
\]

maximizing the myopic reward alone requires her to choose each type infinitely often almost surely.
CHAPTER 5. INCENTIVIZING BETTER EXPERIMENT DESIGNS

With this free exploration, even the naive approach allows the parameters $\omega_{xt}$ to be learned, albeit more slowly than under the MLM. Whether the slower learning rate has a large effect on welfare depends on how patient the company can afford to be. Suppose that (5.7) holds, $\varepsilon = 0$ so the parameters do not evolve over time, and consider the limit as $\beta \to 1$, where the company’s payoff is the long-run average reward per experiment. Under these conditions, one can show that the welfare loss associated with the non-filtering mechanism approaches zero. However, if the pace of innovation $\varepsilon$ is large or the discount factor is small, it becomes essential for the company to learn any parameters quickly.

Valuable future work would be to evaluate the welfare loss at moderate values of $\varepsilon$ and $\beta$. Further, it would be useful to understand how this loss depends on the distribution of the observations and the parametric family $G(\omega)$ used to model the distribution of treatment effects. We expect that higher dimensional, or otherwise more complex families should produce a larger welfare loss, as they make learning $\omega$ more difficult and thus make exploration more urgent.

5.5.2 Robustness to private information

Theorem 9 establishes that the MLM is robust to private information when the employees only choose sample sizes. It is an open question how robust the MLM might be when the employees choose treatment types. Further, it would be interesting to examine its performance under different structures of private information; e.g., private signals on the treatment effects.

5.5.3 Generalization to other corporate structures

In this chapter, we have viewed all employees as interchangeable, with the distribution of any treatment effect depending only on parameters $\omega_{xt}$ that are common to all employees. In reality, some groups of employees may run many A/B tests on an aspect of the service where typical effects are quite different from those seen elsewhere.

Example 8. Consider the content aggregation service of Examples 6 and 7. To recap, employees source new content and run A/B tests to establish whether including the new content on the company’s service increases the share rate. The distribution of the treatment effect in any test, $G(\omega)$, is taken to be the mixture of a normal distribution and a spike at zero; $\omega$ contains (some transformation of) the mixing weight on the spike and the normal variance.

Now suppose that some employees are responsible for sourcing video content, whereas others are responsible for text content. Fewer videos tend to gain traction with visitors to this service, but those that do can be shared many times. Thus for employees who source videos, each treatment effect is drawn from some $G(\omega)$ where the weight on the spike and the normal variance are both large; for employees who source text, each treatment effect is drawn under some $\omega$ where both are small.
If the parameters $\omega_{xt}$ are different for each group of employees, then it is optimal just to run the MLM separately within each group. However, one could consider more complex models where each group has its own parameters, but there are also some parameters that are common to all employees. It would be interesting to explore whether some extension of the MLM could be used to align incentives in such contexts.
Chapter 6

Future work

In this thesis we have tackled what we consider to be four pressing issues facing A/B testing practitioners:

1. Continuous monitoring.
2. Temporal noise.
3. Multiple inference.

For each of the four problems, we have developed novel approaches that are straight-forward for an internet company to integrate into its A/B testing infrastructure. For the first three, this approach is an alternative way to compute p-values (or q-values) and confidence intervals, which can be communicated to experimenters through the same interface as the traditional t-testing approach. For under-exploration, we present a simple mechanism for the company to decide what data from past A/B tests each experimenter should be allowed to access. A crucial strength of our methodology is that we are highly agnostic to the types of decisions that the A/B tests are being used to optimize, as well as to the individual preferences of each experimenter regarding statistical trade-offs, such as power versus run-time. At large, decentralized organizations where the use-cases for A/B testing are widely heterogeneous, this flexibility is essential.

Nonetheless there is still plenty of work that can be done, both to enhance our solutions to the four problems we have identified and to aid the design and analysis of A/B tests more generally. We highlight some potential avenues for future research here.
6.1 A/B testing over time

All four technical chapters in this thesis grapple with how A/B tests are conducted sequentially. In Chapters 2, 3 and 4, we considered how the data within each experiment arrives sequentially, giving importance to how the experimenter selects her sample size. In Chapter 5, we considered how experiments themselves are run in sequence, so exploration in early tests is important to ensure that later experiments can be designed effectively.

Useful future work would be to unify the theory in these chapters into a complete picture of how a company’s A/B tests develop over time. Complexities here include:

- **Dynamic under-exploration.** In Chapter 5, we addressed how experimenters naturally choose fixed-horizon sample sizes that are too small, in that they do not prioritize learning the magnitude of typical treatment effects associated with different types of treatments, as would aid future experiments. If an experimenter is empowered to continuously monitor, she has yet greater flexibility to stop her tests earlier than would benefit the company. We anticipate that developing a mechanism that aligns incentives in this context could be much more challenging.

- **Long-term multiple testing control.** In Chapter 4, we obtained family-wise error and false discovery rate control when a given set of A/B tests are started and stopped simultaneously. Achieving such control would be more difficult if tests may be stopped individually or if new hypotheses are tested sequentially – perhaps endogenously depending on the inferences obtained in earlier tests. Work here could build on the $\alpha$-investing approaches of [27] and [33], which seek FDR control when experiments arrive in sequence but exogenously.

- **Delayed decisions.** Long-term multiple testing control is less important when each experiment is used only to optimize an immediate decision, but can be essential when some optimization requires aggregation of the inferences across a long sequence of tests. Understanding the structure of the reward function associated with these long-ranging decisions will be important to understanding what type of long-term multiple testing controls should be sought.

- **Shifts in optimization goals.** Throughout this thesis, we have considered the testing environment to be stationary over time: in Chapter 4 we modeled the conversion rates in each experiment as stationary Gaussian processes, while in Chapter 5 we modeled the evolution of the parameters that govern the magnitude of typical treatment effects as AR(1). However, over a very long time horizon, one might want to consider the development of the company from a hacky start-up to a well-optimized service. This will impact the types of treatments investigated, the effect sizes sought, and the rewards associated with any decisions.
6.2 Broader incentive issues

In Chapter 5 we looked at how the incentives of employees may not be aligned with that of the company in regards to prioritizing exploration that will aid the design of later A/B tests. In fact there are other ways that experimenters may not be motivated to act in the company’s long-term best interest:

- **Broader externalities in data collection.** In some situations, employees rely on their colleagues’ data not just to inform future experiment designs but even to optimize their own decisions. That is, while one employee may set up an A/B test and determine what data is collected, others may need to analyze the same experimental output. Mechanisms are needed that encourage an employee to internalize any rewards that depend on her data.

  **Example 9.** An entertainment website provides visitors with free content but makes money when visitors click on ads. One employee is a content editor who is responsible for customer engagement. She designs a new layout for the content and she sets up an A/B test to establish whether it increases the mean duration of time that each visitor spends on the site.

  A second employee is responsible for ad revenue. She wants to use the same experimental data to identify whether the new layout decreases the proportion of visitors who click on any ad. However, for this company, typical effect sizes for ad click through rates are smaller than for mean time on site, so the second employee needs a larger data set to be confident that any effect would be detected. The company should motivate the first employee to choose a larger sample size than she otherwise would.

- **Short versus long-term optimization.** Throughout this thesis we have assumed that experimenters plan for the long-term as much as the company desires. In Chapter 1, we assumed that the company and all employees have the same discount factor \( \beta \) for rewards earned in later experiments. In Chapter 3, we gave the experimenter the freedom to select the timeframe over which average treatment effects are measured through her decision of when to stop the test – implicitly we assumed that the experimenter and the company agree about which timeframe is most important.

  In practice, limited time at the company as well as short-term financial incentives can make employees myopic. The company may need to motivate experimenters to run longer tests, for reasons such as obtaining longer-run ATEs and to better support the design of later experiments. “Short-termism” can make efforts to establish long-term multiple testing controls especially challenging, if experimenters do not care about the higher-level decisions that depend on these strict controls.

- **Misaligned incentives within experiments.** This thesis has assumed that experimenters only
want to deploy treatments that really work. In Chapter 5, we asserted that the reward associated with any combination of a treatment effect and a decision is the same for the employee as for the company. In the frequentist chapters we assumed that the experimenter chooses her significance level $\alpha$ appropriately.

In reality, experimenters can prefer a positive result even if it is false, because making the treatment look successful can lead to some remuneration for the experimenter. Indeed the continuous monitoring problem is sometimes cast as an incentive alignment issue, as it enables the experimenter to surreptitiously obtain false positives at well above the nominal significance level. Our always valid $p$-values provide robustness to this issue, as they make the actual false positive rate transparent. In general, it would be useful to examine how other methods used in A/B testing may be susceptible to abuse, and perhaps to develop mechanisms that address these misaligned incentives.

6.3 Beyond parametric inference

This thesis has focused on the A/B testing approach to business optimization, where the quality of any new idea is summarized as the difference in a mean response per visitor under the treatment versus the control. A new data set is collected, which provides statistical inference on this difference, and the experimenter can map this inference to any decision that she feels will improve the service. The strength of the A/B testing perspective is that it empowers the experimenter to use her own judgement, and it imposes no structure on the types of decisions she is allowed to make. The drawback is that the statistical inferences computed are not directly optimized to any specific decision-making goal.

On the other hand, reinforcement learning approaches to business optimization are gaining popularity. These require the set of possible decisions, as well as the rewards associated with any outcomes, to be pre-specified; then they dynamically optimize these decisions as data arrives. Such algorithms are particularly popular in areas such as targeted advertising and viral marketing, where the goals are clear and no dynamic human intervention is required.

Example 10. A designer at an e-commerce company wants to decide whether a new webpage design should be shown to visitors who view a given product. The goal for the company is to maximize the proportion of visitors who buy this product. The designer could run an A/B test with the new webpage as a treatment and the old as a control, and measure the increase in purchase rate between the two groups. At the end of the test, all subsequent visitors would be assigned to the winning version. Alternatively she could use a bandit algorithm to allocate each visitor to one of the two webpages, which is directly optimized to maximize the proportion who buy.

In the future it will be important for research to bridge the gap between these rival perspectives:
• *Comparison of A/B testing and RL.* While the advantage of A/B testing is its ability to handle a wide range of use-cases, it is reasonable to ask how much is lost when the decision-making goals can be codified into the methodology. Useful future work would be to focus on scenarios such as Example 10 and to compare the regret obtained under the A/B testing and bandit approaches. Such comparisons would be particularly interesting in the context of temporal noise or when inference on many parameters is sought.

• *Deriving statistical inference from RL algorithms.* The range of contexts where RL approaches can be useful is expanding. Nonetheless we expect that even in such domains, there will remain some questions where human input is important. The challenge here will be to combine the two approaches, so that most decision-making can be automated while still offering some interpretable statistical output that employees can use to optimize higher-level decisions.

Deriving interpretable output from machine learning methods has gained recent interest in more adversarial settings, where the focus is on securing fairness and transparency [22]. We expect that it will become equally important in the context of business optimization to ensure that humans and machines can work co-operatively.
Bibliography


2012.


[41] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. Online controlled experiments at large scale. In Proceedings of the 19th ACM SIGKDD international


Appendix A

Optimality of the mSPRT for one-variation experiments

This appendix contains proofs of the results stated in Section 2.3.

Proof. Proof of Theorem 2. To establish asymptotic efficiency, given \((M, \alpha)\), it is sufficient to find some \(\theta_1\), where for every feasible test \((T^*, \delta^*)\) where \(\nu^*(\theta_1) \geq \nu(\theta_1; M, \alpha)\), we have that \(\rho^*(\theta_1) \geq \rho(\theta_1; M, \alpha)(1 + o(1))\).

Since the family \(F_\theta\) can be equivalently viewed as exponential tilts of any \(\theta' \in \Theta\), we assume \(\theta_0 = 0\) wlog and write \(I(\theta) = I(\theta, 0)\). Using the normal approximation from Theorem 2 of [11], we have that for any fixed \(\theta\),

\[
\mathbb{P}_\theta(\delta(M, \alpha) = 0) = \Phi\left\{\log(1/\alpha)^{1/2} B(M, \alpha, \theta)\right\} (1 + o(1))
\]

where \(B = \left(\frac{I(\theta)}{\theta^2 \psi'(\theta)}\right)^{1/2} \left(\frac{M}{\log(1/\alpha)} - I(\theta)^{-1}\right)\). On the other hand, standard results on the log partition function \(\psi\) imply that for fixed \((M, \alpha)\),

\[
\log(1/\alpha)^{1/2} B(\theta) \sim \eta_2 \log(1/\alpha)^{1/2} \left(\frac{M \theta^2}{\log(1/\alpha)} - \eta_3\right)
\]

as \(\theta \to 0\), where \(\eta_2\) and \(\eta_3\) are both positive constants. Combining the two results, it follows that for \(\theta_1 = \sqrt{\frac{\log(1/\alpha)}{M}} (\sqrt{2/\eta_2} + \eta_3)\), we have that eventually

\[
\mathbb{P}_{\theta_1}(\delta(M, \alpha) = 0) \leq \Phi\left(\sqrt{2\log(1/\alpha)}\right) := \beta_1
\]

i.e. the mSPRT has power at least \(1 - \beta_1\) at \(\theta_1\) in the limit. Suppose that \((T^*, \delta^*)\) is another feasible test that achieves this power at \(\theta_1\). Once \(\alpha\) is sufficiently small that \(0 < \alpha + \beta_1 < 1\), we have by
Hoeffding’s inequality, for any $\theta \in (0, \theta_1)$,

$$
E_\theta(T^*) \geq \frac{|\log(\alpha + \beta_1)| - \frac{1}{2} \theta_1^2 \psi''(\theta)|\log(\alpha + \beta_1)|^{1/2}}{\max\{I(\theta), I(\theta, \theta_1)\}}
$$

$$
= I(\theta)^{-1} \log(1/\alpha)(1 + o(1))
$$

By continuity, the result holds at $\theta_1$ also. Comparing the above expression with (2.10) gives the desired inequality on the relative run-times at $\theta_1$.

**Proof.** Proof of Proposition 3. Again wlog we assume $\theta_0 = 0$. We fix $\theta \neq 0$, and for contradiction, we suppose that there is some $\beta < 1$ such that feasible tests with $P_\theta(\delta^*(M, \alpha) = 0) \leq \beta$ exist in this limit. Combining this Type II error bound with the Type I error bound at $\alpha$, the Hoeffding inequality implies the existence of some $\kappa$ such that

$$
E_\theta(T^*) \geq \kappa \log(1/\alpha)(1 + o(1))
$$

In the limit, this expectation exceeds $M$, so $T^*$ must certainly exceed $M$ with positive probability. \qed

We now prove three lemmas that will let us prove Theorem 4 and Proposition 4.

**Lemma 1.** Given $H, \theta \neq \theta_0$, there exists a $\lambda > 0$ such that for any $0 < \varepsilon < 1$

$$
\mathbb{P}_\theta \left( |T^H(\alpha) - \frac{\log(1/\alpha)}{I(\theta, \theta_0)}| > \varepsilon \left( \frac{\log(1/\alpha)}{I(\theta, \theta_0)} \right) \right) = O(\alpha^\lambda).
$$

(A.1)

**Proof.** Proof. Immediate when Lemmas 2 & 3 of [62] are combined with Lemma 6 of [47]. \qed

**Lemma 2.** Let

$$
A = \left\{ \theta : I(\theta, \theta_0) \geq \frac{\log(1/\alpha)}{M} \right\},
$$

$$
M \rho(M, \alpha) = \mathbb{E}_{\theta \sim G} \left\{ 1_A \mathbb{E}_\theta(T^H(\alpha)) \right\} + M \mathbb{P}_{\theta \sim G(\theta)}(\bar{A}) + o(1).
$$

(A.2)

**Proof.** Proof. Let $0 < \varepsilon < 1$. Define two times, $n_1 = (1 - \varepsilon)\log(1/\alpha)/I(\theta, \theta_0)$, $n_2 = (1 + \varepsilon)\log(1/\alpha)/I(\theta, \theta_0)$. For $\theta \in \bar{A}$, we have the following bounds. The final inequality is an application of (2.10) in probability.

$$
M \geq \mathbb{E}_\theta(T(M, \alpha)) \geq (n_1 \wedge M)\mathbb{P}_\theta(T^H(\alpha) > n_1)
$$

$$
\geq (1 - \varepsilon) M \mathbb{P}_\theta(T^H(\alpha) > n_1) \geq (1 - \varepsilon) M + o(1).
$$

Let

$$
B^\varepsilon = \left\{ \theta : I(\theta, \theta_0) \geq (1 + \varepsilon) \left( \frac{\log(1/\alpha)}{M} \right) \right\},
$$
For $\theta \in B^c$, $M \geq n_2$. Thus
\[
\mathbb{E}_\theta(T^H(\alpha)) \geq \mathbb{E}_\theta(T(M, \alpha)) \geq \mathbb{E}_\theta(T(n_2, \alpha)) \geq \int_{T^H(\alpha) \leq n_2} T^H(\alpha) d\mathbb{P}_\theta = \mathbb{E}_\theta(T^H(\alpha)) - \int_{T^H(\alpha) > n_2} T^H(\alpha) d\mathbb{P}_\theta.
\]
By Cauchy-Schwartz, in $L^2$ and Lemma 4,
\[
\int_{T^H(\alpha) > n_2} T^H(\alpha) d\mathbb{P}_\theta \leq \left( \mathbb{E}_\theta(T^H(\alpha)^2) \mathbb{P}_\theta(T^H(\alpha) \geq n_2) \right)^{1/2} = O(\alpha^{-\lambda/2} \log \alpha^{-1/2}) = o(1).
\]
For $\theta \in A \setminus B^c$,
\[
\mathbb{E}_\theta(T^H(\alpha)) \geq \mathbb{E}_\theta(T(M, \alpha)) = \mathbb{E}_\theta(T^H(\alpha)) + \int_{T^H(\alpha) \geq n_2} \{ n_S - T^H(\alpha) \} d\mathbb{P}_\theta \\
\geq \mathbb{E}_\theta(T^H(\alpha)) + \int_{M \leq T < n_2} \{ n_S - T^H(\alpha) \} d\mathbb{P}_\theta - \int_{T^H(\alpha) > n_2} T^H(\alpha) d\mathbb{P}_\theta \\
\geq \mathbb{E}_\theta(T^H(\alpha)) - (n_2 - M) + o(1) \geq \mathbb{E}_\theta(T^H(\alpha)) - \varepsilon M + o(1).
\]
Putting the three cases together, we integrate over $\theta \sim G$ to obtain (4), up to some error linear in $\varepsilon$. To justify this step, it is easy to check that each term is finite. The result now holds on letting $\varepsilon \to 0$. \hfill \Box

**Lemma 3.** Let
\[
C(\alpha) = \int_{0}^{1} \Phi \left( \sqrt{\frac{1}{2} \log(1/\alpha)(x^2 - 1)} \right) dx, \\
C_f(\alpha) = \int_{0}^{1} \Phi \left( \sqrt{\log(1/\alpha)(x - 1)} \right) dx.
\]
For the prior $G = N(0, \tau^2)$, let $\nu(M, \alpha)$ be the average power of the mSPRT. If $M = O(\log(1/\alpha))$,
\[
\nu(M, \alpha) \sim C(\alpha) \frac{2\sqrt{2}}{\tau} \left( \frac{\log(1/\alpha)}{M} \right)^{1/2}
\]
Let $\nu_f(n, \alpha)$ be the average power of the fixed-horizon test with sample size $n$. If $n = \Omega(\log(1/\alpha))$,
\[
\nu_f(n, \alpha) \sim C_f(\alpha) \frac{2\sqrt{2}}{\tau} \left( \frac{\log(1/\alpha)}{n} \right)^{1/2}
\]
**Proof.** Proof. Wlog we suppose $\theta_0 = 0$. We begin with the fixed horizon result. It is simple to show that $z_1 - 2\alpha \sim \sqrt{2 \log(1/\alpha)}$ as $\alpha \to 0$. Hence
\[
\nu_f(\theta) = \Phi \left( |\theta| \sqrt{n} - z_1 - 2\alpha \right) = \Phi \left( \log(1/\alpha)^{1/2} S_f(\theta, n, \alpha) \right)
\]
where \( S_f(\theta, n, \alpha) = |\theta| \left( \frac{\log(1/\alpha)}{n} \right)^{-1/2} - \sqrt{2} \).

Let \( B_f = \{ \theta : A_f(\theta, n, \alpha) \geq \sqrt{2} \} \). We split up the average power as

\[
\nu_f = \int_{B_f} \nu_f(\theta) \frac{1}{\tau} \phi(\theta/\tau) d\theta + \int_{\bar{B}_f} \nu_f(\theta) \frac{1}{\tau} \phi(\theta/\tau) d\theta
\]

denoting the two terms by (i) and (ii) respectively. For \( \theta \in B_f \), the standard tail bound on the Normal CDF, \( \Phi(x) \leq x - \frac{1}{\phi(x)} \) gives

\[
\Phi \left( \frac{\log(1/\alpha)}{n} \right)^{1/2} S_f(\theta, n, \alpha) \leq \frac{4 \pi \log(1/\alpha)}{\alpha} \frac{1}{2} = o(\alpha),
\]

so that (i) = o(\alpha) as well. For term (ii), we note that \( \bar{B}_f \to \{0\} \) so that \( \phi(\theta/\tau) \sim 1 \). This, the change of variable \( x = \left( \frac{2 \log(1/\alpha)}{n} \right)^{-1/2} \theta \) and symmetry of the integrand give

\[
(ii) \sim \frac{2 \sqrt{2}}{\tau} \left( \frac{\log(1/\alpha)}{n} \right)^{1/2} \int_0^2 \Phi \left( \frac{\log(1/\alpha)}{n} \right)^{1/2} (x - 1) \, dx.
\]

The result follows on noting \( \Phi \left( \frac{\log(1/\alpha)}{n} \right)^{1/2} (x - 1) \) = o(1) when \( x > 1 \).

For the mSPRT, we use the normal approximation to the tail probabilities of the mSPRT stopping time from Theorem 2 of \cite{44} which, in the case of standard normal data, gives

\[
P_\theta(T(\alpha) > M) \sim \Phi \left( \frac{\log(1/\alpha)}{n} \right)^{1/2} S(\theta, M, \alpha)
\]

where \( S(\theta, M, \alpha) = \frac{1}{2 \sqrt{2}} \left\{ \theta^2 \left( \frac{\log(1/\alpha)}{M} \right) - 2 \right\} \). The rest of the proof proceeds as for the fixed horizon test, except with \( B = \{ \theta : S \geq \sqrt{2} \} \) and changes in the integrand of (ii) as stated in the proposition.

\[\square\]

**Proof.** Proof of Theorem 3. Combining Lemma 4 with equation (67) of \cite{44}, we find that, up to o(1),

\[
M \rho(M, \alpha) = -2E_{\theta \sim G} 1_A I(\theta, \theta_0)^{-1} \log h(\gamma) + K(G, \alpha)
\]

for some function \( K \). The stated \( \gamma^* \) is the minimizer of this expression.

\[\square\]

**Proof.** Proof of Proposition 4. From Lemma 5, we see that to match the average power of the truncated mSPRT \( (\nu_f = \nu) \), the calibrated fixed-horizon test must have sample size \( O(M) \); i.e. \( \rho_f = O(1) \). Thus it is sufficient to show that \( \rho = o(1) \).

Again we take \( \theta_0 = 0 \) wlog, so for standard normal data \( I(\theta, 0) = \theta^2 / 2 \). We invoke Lemma 4 and
attack the two terms in that result separately. Let $\delta = \left(\frac{2\log(1/\alpha)}{M}\right)^{1/2}$.

$$Pr_{\theta \sim N(0, \tau)}(\tilde{A}) = 2\int_{0}^{\delta} \frac{1}{\tau} \phi(\theta/\tau) d\theta \sim \frac{2\sqrt{2}}{\tau} \left(\frac{\log(1/\alpha)}{M}\right)^{1/2} = o(1)$$

by similar arguments to those in the proof of Lemma 3.

The first term in Lemma 2 is more complicated. By equation (67) of [45] and the discussion there that follows it,

$$E_{\theta}(T(\alpha)) = 2\theta^{-2} \log(1/\alpha) + \theta^{-2} \log \log(1/\alpha) + D_{1}\theta^{-2} \log |\theta|$$

$\qquad + D_{2}\theta^{-2} + D_{3} + D_{4}\theta^{-1} B(\theta/2) + o(1)$

as $\alpha \to 0$, where

$$B(u) = \sum_{k=1}^{\infty} k^{-1/2} \phi(uk^{1/2}) - u\Phi(-uk^{1/2}).$$

It remains to show that each term in (A.3) has $o(M)$ expectation on $1_{A}$. We focus on terms 1, 3 & 6 as the remainder are clearly lower order.

Term 1.

$$E_{\theta \sim N(0, \tau)}(1_{A}\theta^{-2}) = 4 \int_{\delta}^{\infty} \theta^{-2} \frac{1}{\tau} \phi\left(\frac{\theta}{\tau}\right) d\theta = \frac{4}{\tau^2} \left\{ \frac{\tau}{\delta} \phi\left(\frac{\delta}{\tau}\right) - \Phi\left(\frac{\delta}{\tau}\right) \right\}$$

which is bounded over $\alpha \in (0, 1 - \varepsilon)$. It follows that

$$E_{\theta \sim N(0, \tau)}(1_{A}\theta^{-2}) \log(1/\alpha) \sim \frac{2\sqrt{2}}{\tau} M \left(\frac{\log(1/\alpha)}{M}\right)^{1/2} = o(M).$$

Term 3. By calculus,

$$E_{\theta \sim N(0, \tau)}(1_{A}\theta^{-2} \log |\theta|) \propto \int_{\delta}^{\infty} \theta^{-2} \log \theta e^{-\theta^2/2\tau^2}$$

$$= \frac{1}{4} \left[ \frac{1}{\sqrt{2\tau}} \Gamma \left( -\frac{1}{2}, \frac{\delta^2}{2\tau^2} \right) \log \delta \right.$$  

$$+ \delta^{-1}\text{MeijerG} \left( \left\{ \left\{ \frac{3}{2}, \left\{ \frac{3}{2} \right\}, \left\{ 0, \frac{1}{2}, \frac{1}{2} \right\}, \left\{ \right\}, \left\{ \right\} \right\}, \left\{ \frac{\delta^2}{2\tau^2} \right\} \right) \right]$$

The MeijerG term is asymptotically constant as $\delta \to 0$, and

$$\Gamma \left( -\frac{1}{2}, \frac{\delta^2}{2\tau^2} \right) \to 2\sqrt{2}\tau \delta^{-1}. $$
It follows that
\[ E_{\theta \sim N(0, \tau)}(1_A \theta^{-2} \log |\theta|) \propto M \left[ K_1 (M \log(1/\alpha))^{-1/2} + K_2 \left( \frac{\delta \log \delta}{\log(1/\alpha)} \right) \right] \]
where \( K_i \) are both constants depending on \( \tau \). Both terms in the bracketed sum clearly converge to 0 since \( \delta \to 0 \).

Term 6. By standard bounds on the normal CDF, \( B(u) \geq 0 \) and
\[ B(u) \leq u^{-2} \left( \int_1^{\infty} x^{-3/2} \phi(ux^{1/2})dx + \phi(u) \right) \]
\[ = \theta^{-2} (3\phi(\theta) - 2\theta\Phi(-\theta)) \]
Hence,
\[ E_{\theta \sim N(0, \tau)}(1_A \theta^{-1}B(\theta/2)) \leq K_3 E_{\theta \sim N(0, \tau)}(1_A \theta^{-3} \phi(\theta/2)) + K_4 \]
\[ \leq K_4 \delta^{-2} e^{K_5\delta^2} - K_7 \Gamma(0, K_6\delta^2) + K_4, \]
where \( \delta^{-2} = M/\log(\alpha^{-2}) = o(M) \) and \( \Gamma(0, K_6\delta^2) \sim \log K_6/\delta^2 = O(\log \delta) = o(M) \).
Appendix B

Optimality of the mSPRT for two-stream normal data

Given a choice of mixing distribution $H$, the two-stream p-values are derived from the mSPRT which rejects the null if

$$\Lambda^H_n(S_n) = \int_{\Theta} \left( \frac{f_{\theta}(S_n)}{f_{\theta_0}(S_n)} \right)^n dH(\theta)$$

ever exceeds $\alpha$, where $S_n = \frac{1}{n} \sum_{i=1}^{n} W_i$. First we notice that $\Lambda^H_n$ depends on the data only through $(-1,1)^T S_n$ and this is distributed as $N(\theta, 2\sigma^2/n)$, so the power and the run-time of this test do not depend on $\mu$. Let $\nu^H(\theta; M, \alpha)$, $\rho^H(\theta; M, \alpha)$ be the power and average relative run-length of the truncated test. We say that the relative efficiency of this test at $(M, \alpha)$ is

$$\phi^H(M, \alpha) = \inf_{(T^*, \delta^*)} \inf_{\theta \neq \theta_0, \mu} \frac{\rho^*(\theta, \mu)}{\rho^H(\theta; M, \alpha)}$$

where the infimum is taken over all tests with $T^* \leq M$ a.s., $\sup_{\mu} \nu(\theta_0, \mu) \leq \alpha$, and for all $\theta \neq \theta_0$, $\inf_{\mu} \nu^*(\theta, \mu) \geq \nu^H(\theta; M, \alpha)$.

Proposition 19. For any $H$, if $\alpha \to 0, M \to \infty$ such that $M = O(\log(\alpha^{-1}))$, we have $\phi^H(M, \alpha) \to 1$.

Proof. Fix $\mu = \mu^* \text{ arbitrarily. Then any } (T^*, \delta^*) \text{ satisfying the above conditions is also feasible for testing } H_0 : \theta = \theta_0, \mu = \mu^* \text{ against } H_1 : \theta \neq \theta_0, \mu = \mu^*. \text{ The result follows by Theorem }$ \ref{thm:optimality}. 

Now we consider any prior for the pair $(\theta, \mu)$ under $H_1$, such that $\theta \sim N(0, \tau^2)$ marginally. For normal mixtures $H = N(0, \gamma^2)$, let $\rho_\gamma(M, \alpha)$ be the average power and relative run-time over this prior.
**Proposition 20.** To leading order as $\alpha \to 0, M \to \infty, M = O(\log(\alpha^{-1})), \rho, \gamma$ is minimized by

$$\gamma^2 = \tau^2 \frac{\Phi(-b)}{\frac{1}{\sqrt{b}} \phi(b) - \Phi(-b)}$$

where $b = \left( \frac{2\sigma^2 \log \alpha^{-1}}{M\tau^2} \right)^{1/2}$.

**Proof.** Immediate from Theorem 3.

Now we compare the truncated mSPRT to the fixed-horizon t-test based on the difference between the sample means in the two streams, which is calibrated to have the same average power on this prior. Noting that the fixed-horizon sample size does not depend on $\mu$, we see that Proposition 4 carries over to two-stream normal data.
Appendix C

Type I error control for the mSCPRT

Proof. Proof of Proposition 7. As in the proof of Proposition 5, we let \( \{m_1, \ldots, m_r\} \subset \mathbb{N} \) be an arbitrary.

\[
(\bar{\mu}^{t}_{\cdot m_1}, \ldots, \bar{\mu}^{t}_{\cdot m_r}, \bar{\mu}^{t}_{\cdot n_s}, \ldots, \bar{\mu}^{t}_{\cdot n_s-1}, \bar{\mu}^{t}_{\cdot n_s}) = (Z_{m_1}^{t} - \bar{Z}_{n_s}^{t}, \ldots, Z_{m_r}^{t} - \bar{Z}_{n_s}^{t}, \bar{Z}_{n_1}^{t} - Z_{n_s}^{t}, \ldots, \bar{Z}_{n_s-1}^{t} - Z_{n_s}^{t}, \bar{Z}_{n_s}^{t} + \mu^{t})
\]

\sim N\left(0, \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\
\Sigma_{12}^{T} & \Sigma_{22} & \Sigma_{23} \\
\Sigma_{13}^{T} & \Sigma_{23}^{T} & \Sigma_{33} + W
\end{bmatrix}\right)

where \( \Sigma_{11} \in \mathbb{R}^{r \times r}, \Sigma_{22} \in \mathbb{R}^{(s-1) \times (s-1)} \) and \( \Sigma_{33} \in \mathbb{R} \) are suitably defined in terms of \( K \).
Thus
\[
(\mu_{m_1}^t - \bar{\mu}_{n_1}, \ldots, \mu_{m_r}^t - \bar{\mu}_{n_r}) | (\bar{\mu}_{n_1}^t - \bar{\mu}_{n_2}, \ldots, \bar{\mu}_{n_{r-1}} - \bar{\mu}_{n_r}) \]
\[
= N \left( \Sigma_{12} \Sigma_{13} \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} + W \end{bmatrix}^{-1} (\bar{\mu}_{n_1}^t - \bar{\mu}_{n_2}, \ldots, \bar{\mu}_{n_{r-1}} - \bar{\mu}_{n_r}) \right)
\]
\[
\Sigma_{11} - [\Sigma_{12} \Sigma_{13}] \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} + W \end{bmatrix}^{-1} [\Sigma_{12} \Sigma_{13}]^T \]
\[
\rightarrow N \left( [\Sigma_{12} \Sigma_{13}]^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} (\bar{\mu}_{n_1}^t - \bar{\mu}_{n_2}, \ldots, \bar{\mu}_{n_{r-1}} - \bar{\mu}_{n_r}) \right)
\]
\[
\Sigma_{11} - [\Sigma_{12} \Sigma_{13}] \begin{bmatrix} \Sigma_{22} & 0 \\ 0 & 0 \end{bmatrix} [\Sigma_{12} \Sigma_{13}]^T \]
\[
= N \left( \Sigma_{12} \Sigma_{13}^{-1} (\bar{\mu}_{n_1}^t - \bar{\mu}_{n_2}, \ldots, \bar{\mu}_{n_{r-1}} - \bar{\mu}_{n_r}), \Sigma_{11} - \Sigma_{12} \Sigma_{13}^{-1} \Sigma_{12}^T \right)
\]
and so
\[
(\mu_{m_1}^t, \ldots, \mu_{m_r}^t) | (\bar{\mu}_{n_1}^t) \rightarrow N \left( \mu_{m_1}^t, \ldots, \mu_{m_r}^t, \Sigma_{11} - \Sigma_{12} \Sigma_{13}^{-1} \Sigma_{12}^T \right).
\]

For the second part, we note that the processes \( \mu \) and \( \theta \) are marginally independent, so they remain independent after conditioning on \( (\bar{\mu}_{n_1}^t, \bar{\mu}_{n_2}^t) \). \( (\mu_{m_1}^t, \ldots, \mu_{m_r}^t) \) is distributed just as \( (\mu_{m_1}^t, \ldots, \mu_{m_r}^t) | (\bar{\mu}_{n_1}^t) \). Thus, on the event that \( \bar{\theta}_{n_1} = \cdots = \bar{\theta}_{n_r} = \theta \),
\[
(\mu_{m_1}^t - \bar{\mu}_{m_1}, \ldots, \mu_{n_r} - \bar{\mu}_{m_r}) | (\bar{\mu}_{n_1}^t, \bar{\mu}_{n_2}^t) \rightarrow N \left( \theta \mathbf{1}, 2 (\Sigma_{11} - \Sigma_{12} \Sigma_{13}^{-1} \Sigma_{12}^T) \right).
\]

\[ \square \]

**Proof.** Proof of Proposition 8. This is a straight-forward extension of the result that the mSPRT controls Type I error for IID data.

First we show that \( L_n(\mathcal{N}) \) is a martingale on the event that \( \bar{\theta}_m = \theta_0 \) for all \( m \in \mathcal{N} \). After conditioning on \( (\bar{\mu}_m^t, \bar{\mu}_m^c) \) and restricting to the desired event, we have that for any \( n \),
\[
(W_1, \ldots, W_n) \sim f_{\theta_0, \mathcal{N}}^n. \] Let \( g_{\theta, \mathcal{N}}^{n+1} \) be the conditional density of \( W_{n+1}|W_1, \ldots, W_n, (\bar{\mu}_m^t, \bar{\mu}_m^c) \).
which is well-defined because \((W_1, \ldots, W_{n+1})|(\bar{\mu}_m^t, \bar{\mu}_m^c)_{m \in \mathcal{N}}\) is a non-degenerate Gaussian. Then

\[
\mathbb{E} \left( \frac{f_{\theta, N}^{n+1}(W_1, \ldots, W_{n+1})}{f_{\theta, N}^n(W_1, \ldots, W_n)} \mid W_1, \ldots, W_n, (\bar{\mu}_m, \bar{\mu}_m^c)_{m \in \mathcal{N}} \right)
\]

\[
= \frac{f_{\theta, N}^n(W_1, \ldots, W_n)}{f_{\theta, N}^n(W_1, \ldots, W_n)} \cdot \mathbb{E} \left( \frac{g_{\theta, N}^{n+1}(W_{n+1}; W_1, \ldots, W_n, (\bar{\mu}_m^t, \bar{\mu}_m^c)_{m \in \mathcal{N}})}{g_{\theta, N}^{n+1}(W_{n+1}; W_1, \ldots, W_n, (\bar{\mu}_m^t, \bar{\mu}_m^c)_{m \in \mathcal{N}})} \right)
\]

\[
= \frac{f_{\theta, N}^n(W_1, \ldots, W_n)}{f_{\theta, N}^n(W_1, \ldots, W_n)} \cdot \int \left( \frac{g_{\theta, N}^{n+1}(w; \ldots)}{g_{\theta, N}^{n+1}(w; \ldots)} \right) g_{\theta, N}^{n+1}(w; \ldots) \, dw
\]

Thus \(f_{\theta, N}^n(W_1, \ldots, W_n)/f_{\theta, N}^n(W_1, \ldots, W_n)\) is a martingale. It follows that, as a mixture of martingales, \(L_n(\mathcal{N})\) is itself a martingale.

The result follows by the Optional Stopping Theorem. Let \(n\) be arbitrary. On the desired event,

\[
1 = L_0(\mathcal{N}) = \mathbb{E} \left\{ L_{T \wedge n}(\mathcal{N}) \mid (\bar{\mu}_m^t, \bar{\mu}_m^c)_{m \in \mathcal{N}} \right\} \geq \alpha \mathbb{P}_K \left( L_{T \wedge n}(\mathcal{N}) \geq 1/\alpha \mid (\bar{\mu}_m^t, \bar{\mu}_m^c)_{m \in \mathcal{N}} \right).
\]

Now take \(n \to \infty\). \(\square\)

**Proof.** Proof of Theorem 4. For all \(m \in \mathcal{N}\) and \(n \leq m\), we have that \(V_{n, \mathcal{N}} \leq V_{n, \{m\}}\). It follows that \(L_n(\{m\}) \leq L_n(\mathcal{N})\) holds a.s. (i.e. viewed as functions of the data, this equality holds at any observed values). Thus for any stopping time \(T\) taking values in \(\mathcal{N}\), we have that \(\max_{n \leq T} L_n(\{T\}) \leq \max_{n \leq T} L_n(\mathcal{N})\). The result now follows from Proposition 8 if we take the stopping time considered in that proposition to be the minimum of \(T\) and the first hitting time of \(1/\alpha\). \(\square\)
Appendix D

Sequential FDR control

To prove Proposition 10, we first prove the following lemma.

Lemma 4.
\[
\sup_{f \in \mathcal{F}} \sum_{k=1}^{m} \frac{1}{k} \int_{(k-1)\alpha/m}^{k\alpha/m} f(x)dx = \frac{\alpha}{m} \sum_{k=1}^{m} \frac{1}{k}
\]
where \( \mathcal{F} = \{ f : [0,1] \to \mathbb{R} : F(x) = \int_{0}^{x} f(t)dt \leq x, F(1) = 1 \}, m \geq 1, \text{and } 0 \leq \alpha \leq 1 \).

Proof. Since \( f \in \mathcal{F} \) are bounded, we restate the optimization in terms of \( F_k = F(\frac{k\alpha}{m}) \), and \( F_0 \equiv 0 \),
\[
\sup_{F_1, \ldots, F_m} \sum_{k=1}^{m} \frac{1}{k} (F_k - F_{k-1})
\]
subject to \( 0 \leq F_j \leq \frac{k\alpha}{m} \), \( F_k \geq F_{k-1} k = 1, \ldots, m \).

The objective can be rearranged as
\[
\sum_{k=1}^{m} \frac{1}{k(k+1)} F_k + \frac{1}{m} F_m
\]
which is clearly maximized by \( F_k = \frac{k\alpha}{m} \) for all \( k \).

Proof. Proof of Proposition 10. Adapting the proof given in [8] now is straight-forward. Translating that proof into the sequential notation of this thesis, the only non-immediate step is to show
\[
\sum_{k=1}^{m} \frac{1}{k} \sum_{r=k}^{m} \mathbb{P} (T_k^i \leq T < T_{k-1}^i, T_r \leq T < T_{r+1}, T \leq \infty)
\leq \sum_{k=1}^{m} \frac{1}{k} \mathbb{P} (\frac{(k-1)\alpha}{m} \leq p_T \leq \frac{k\alpha}{m}) \leq \frac{\alpha}{m} \sum_{k=1}^{m} \frac{1}{k}
\]
for all truly null hypotheses \(i\). The first inequality is a restatement of definitions, and the second follows from Lemma 4 since, by always-validity, \(p_f^i\) is super-uniform.

\[ \text{Proof.} \]

Proof of Theorem 6. We assume wlog that the truly null hypotheses are \(i = 1, \ldots, m_0\). Letting \(V_n\) denote the number of true null rejected at \(n\), the FDR can be expanded as

\[
E \left( \sum_{r=1}^{m} \frac{1}{r} V_T 1\{T_r \leq T < T_r^+\} 1\{T < \infty\} \right) = \sum_{i=1}^{m_0} \sum_{r=1}^{m} \frac{1}{r} P \left( T^i_r \leq T, T_r \leq T < T_r^+ , T < \infty \right).
\]

Note that the sets \(\{T_r \leq T < T_r^+\}\) are disjoint and cover any location of \(T\). Consider the terms in the sum over \(i \in I\) and \(i \notin I\) separately. For \(i \notin I\), we bound the probability in the third equality by

\[
P \left( T^i_r \leq T, T < \infty \right) P \left( T_r \leq T < T_r^+ \mid T^i_r \leq T, T < \infty \right) \leq \frac{\alpha r}{M} P \left( T_r \leq T < T_r^+ \mid T^i_r \leq T, T < \infty \right)
\]

where the first inequality follows from always-validity of sequential p-values, and the last equality because the modified BH procedure on the \(m - 1\) hypothesis other than the \(i\)th makes equivalent rejections at time \(T\) when \(T^i_r \leq T\).

For \(i \in I\), arguing as in the proof of Proposition ?? shows

\[
\sum_{r=1}^{m} \frac{1}{r} P \left( T^i_r \leq T, T_r \leq T < T_r^+ , T < \infty \right) \leq \frac{\alpha}{m} \sum_{k=1}^{m} \frac{1}{k}.
\]

The proof is completed on application of (4.6) to the terms in the first expansion with \(i \notin I\) and re-ordering of the resulting terms. \[\square\]
Appendix E

FCR control

Proof. Proof of Theorem 5. By Lemma 1 in [9],

$$FCR = \sum_{i=1}^{m} \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^{BH}| = r, i \in J \cup S^{BH}, \theta^i \notin \tilde{C}^i)$$

On the event $i \in J \cup S^{BH}$, there are two possibilities. If $i \in S^{BH}$, we can say $R^{BH} \leq |J \cup S^{BH}|$. If $i \notin S^{BH}$, we can say further that $R^{BH} + 1 \leq |J \cup S^{BH}|$. In either case, it follows that $\mathcal{C}^i(1 - \alpha|J \cup S^{BH}|/m) \subset \tilde{C}^i$, and so the FCR is at most

$$\sum_{i=1}^{m} \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^{BH}| = r, i \in J \cup S^{BH}, \theta^i \notin \mathcal{C}^i(1 - \alpha r/m))$$

Case 1: $i \notin J$.

$$\{|J \cup S^{BH}| = r, i \in J \cup S^{BH}, \theta^i \notin \mathcal{C}^i(1 - \alpha r/m)\}$$

$$= \{|J \cup (S^{BH})^{-i}| = r - 1, p^i \leq \alpha r/m, \theta^i \notin \mathcal{C}^i(1 - \alpha r/m)\}$$

$$\subset \{|J \cup (S^{BH})^{-i}| = r - 1, \theta^i \notin \mathcal{C}^i(1 - \alpha r/m)\}$$

These two events are independent, so

$$\sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^{BH}| = r, i \in J \cup S^{BH}, \theta^i \notin \mathcal{C}^i(1 - \alpha r/m))$$

$$\leq \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup (S^{BH})^{-i}| = r - 1)\mathbb{P}(\theta^i \notin \mathcal{C}^i(1 - \alpha r/m))$$

$$\leq \frac{\alpha}{m} \sum_{r=1}^{m} \mathbb{P}(|J \cup (S^{BH})^{-i}| = r - 1) = \frac{\alpha}{m}$$
Case 2: $i \in J$.

$$\sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^{BH}| = r, i \in J \cup S^{BH}, \theta^i \notin Cl^i(1 - \alpha r/m))$$

$$\leq \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^{BH}| = r, \theta^i \notin Cl^i(1 - \alpha r/m))$$

$$= \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^{BH}| = r | \theta^i \notin Cl^i(1 - \alpha r/m)) \mathbb{P}(\theta^i \notin Cl^i(1 - \alpha r/m))$$

$$\leq \frac{\alpha}{m} \sum_{r=1}^{m} \mathbb{P}(|J \cup S^{BH}| = r | \theta^i \notin CI^i(1 - \alpha r/m))$$

Since $S^{BH}$ is a function only of the p-values and the data streams are independent, the events \{\{J \cup S^{BH}| = r\} and \{\theta^i \notin CI^i(1 - \alpha r/m)\}\} are conditionally independent given $p^i$. Hence,

$$\mathbb{P}(|J \cup S^{BH}| = r | \theta^i \notin CI^i(1 - \alpha r/m)) \leq \max_{\rho} \mathbb{P}(|J \cup S^{BH}| = r | p^i = \rho)$$

It is easily seen that this maximum must be attained at either $\rho = 0$ or $\rho = 1$, so

$$\mathbb{P}(|J \cup S^{BH}| = r | \theta^i \notin CI^i(1 - \alpha r/m)) \leq \mathbb{P}(|J \cup S^{BH}| = r | p^i = 0) + \mathbb{P}(|J \cup S^{BH}| = r | p^i = 1)$$

$$= \mathbb{P}(|J \cup (S^{BH})_0^i \setminus \emptyset| = r - 1) + \mathbb{P}(|J \cup (S^{BH})_1^i \setminus \emptyset| = r - 1)$$

Thus

$$\sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^{BH}| = r, i \in J \cup S^{BH}, \theta^i \notin Cl^i(1 - \alpha r/m))$$

$$\leq \frac{\alpha}{m} \left\{ \sum_{r=1}^{m} \mathbb{P}(|J \cup (S^{BH})_0^i \setminus \emptyset| = r - 1) + \sum_{r=1}^{m} \mathbb{P}(|J \cup (S^{BH})_1^i \setminus \emptyset| = r - 1) \right\}$$

$$= \frac{2\alpha}{m}$$

Summing over all $i$ now gives the desired result. \qed

**Proof.** Proof of Theorem 7. By the same argument as for Theorem 6, we find that the FCR is at most

$$\sum_{i=1}^{m} \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S^T| = r, i \in J \cup S^T, \theta^i \notin CI^i T(1 - \alpha r/m), T < \infty)$$
**APPENDIX E. FCR CONTROL**

**Case 1:** $i \notin J$. As in Theorem 5, we obtain
\[
\sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S_T^{BH}| = r, i \notin J \cup S_T^{BH}, \theta^i \notin \Theta_T(1-\alpha/m), T < \infty) \\
\leq \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup (S_T^{BH})_0^{-i}| = r-1, \theta^i \notin \Theta_T(1-\alpha/m), T < \infty) \\
= \sum_{r=1}^{m} \frac{1}{r} \mathbb{P}((T_{r-1})^{-i,J}_0 \leq T < (T_{r-1})^{-i,J}_1, T_r^{i,\theta^i} \leq T < \infty) \\
\leq \frac{\alpha}{m} \sum_{r=1}^{m} \mathbb{P}((T_{r-1})^{-i,J}_0 \leq T < (T_{r-1})^{-i,J}_1, T_r^{i,\theta^i} \leq T < \infty) \\
\leq \frac{\alpha}{m}
\]

**Case 2:** $i \in J$. As before,
\[
\sum_{r=1}^{m} \frac{1}{r} \mathbb{P}(|J \cup S_T^{BH}| = r, i \in J \cup S_T^{BH}, \theta^i \notin \Theta_T(1-\alpha/m)) \\
\leq \frac{\alpha}{m} \sum_{r=1}^{m} \mathbb{P}(|J \cup S_T^{BH}| = r \mid \theta^i \notin \Theta_T(1-\alpha/m)) \\
= \frac{\alpha}{m} \sum_{r=1}^{m} \mathbb{P}(|J \cup S_T^{BH}| = r \mid T_r^{i,\theta^i} \leq T < \infty) \\
\leq \frac{\alpha}{m} \sum_{r=1}^{m} \max_{\rho} \mathbb{P}(|J \cup S_T^{BH}| = r \mid p^i_T = \rho, T_r^{i,\theta^i} \leq T < \infty) \\
\leq \frac{\alpha}{m} \left\{ \sum_{r=1}^{m} \mathbb{P}((|J \cup (S_T^{BH})^{-i}_0| = r-1 \mid T_r^{i,\theta^i} \leq T < \infty) \\
+ \sum_{r=1}^{m} \mathbb{P}((|J \cup (S_T^{BH})^{-i}_1| = r-1 \mid T_r^{i,\theta^i} \leq T < \infty) \right\} \\
= \frac{\alpha}{m} \left\{ \sum_{r=1}^{m} \mathbb{P}((T_{r-1})^{-i,J}_0 \leq T < (T_{r-1})^{-i,J}_1, T_r^{i,\theta^i} \leq T < \infty) \\
+ \sum_{r=1}^{m} \mathbb{P}((T_{r-1})^{-i,J}_1 \leq T < (T_{r-1})^{-i,J}_1, T_r^{i,\theta^i} \leq T < \infty) \right\} \\
\leq \frac{2\alpha}{m}
\]

Finally we sum over $i$. 

$\square$
Appendix F

Constructing the experiment design game

Proof. Recall that, by definition, \( \nu_{t+1} \) is the conditional distribution of the vector \( (\omega_{1,t+1}, \ldots, \omega_{M,t+1}) \) given \( f_{t+1} \). When we condition on \( f_t \), there are two sources of randomness in the measure, \( \nu_{t+1} \). The first source is the randomness of \( \omega_{xt} \) given \( \omega_{xt} \). The second comes from the data revealed to the agents after period \( t \), which is contained in \( f_{t+1} \) but not \( f_t \).

We address the irreducible noise in transitioning from \( \omega_{xt} \) to \( \omega_{x,t+1} \) first. We define the innovations \( Z_{x,t+1} \) such that:

\[
\omega_{x,t+1} = \sqrt{1 - \varepsilon^2} \omega_{xt} + \varepsilon Z_{x,t+1}
\]

Note that \( Z_{x,t+1} \sim N(0, 1) \), and it is independent of \( f_{t+1} \).

Let \( A_1, \ldots, A_M \subset \mathbb{R}^d \) be arbitrary. Define the sets \( B_x(z) \) such that, when \( Z_{x,t+1} = z \), we have that \( \omega_{x,t+1} \in A_x \) if and only if \( \omega_{xt} \in B_x(z) \).

\[
\nu_{t+1}(A_1, \ldots, A_M) = P(\omega_{x,t+1} \in A_x \forall x | f_{t+1})
\]

\[
= P(\omega_{xt} \in B_x(Z_{x,t+1}) \forall x | f_{t+1})
\]

\[
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(\omega_{xt} \in B_x(z_x) \forall x | f_{t+1}) \phi(z_1) \cdots \phi(z_M) dz_1 \cdots dz_M
\]

where \( \phi \) is the standard normal density.

Next we address the contraction of \( \nu_t \) as \( Y_t \) provides more information about \( \omega_{xt} \). After fixing \( E_t = \{x_{it}, n_{it} \forall i \} \), \( Y_t = \{Y_{1xt}, \ldots, Y_{nxt} \forall i \} \) has an everywhere positive density

\[
q(y_1^1, \ldots, y_{n_1}^1, \ldots, y_1^N, \ldots, y_{n_N}^N; \omega_{1t}, \ldots, \omega_{Mt}) = \prod_{i=1}^{N} \prod_{j=1}^{n_{it}} \int_{\Theta} dF_{\theta}(y_j^i) G(\omega_{xit}, \theta)(d\theta)
\]

where \( \Theta \) is the parameter space.

99
APPENDIX F. CONSTRUCTING THE EXPERIMENT DESIGN GAME

Construct a new vector $Y'$ from $Y_t$ by leaving each entry unchanged if it is included in the filtered data set $Y_t' = \sigma(\nu_t, E_t, Y_t)$; else the entry is replaced by a standard normal random variable, which is sampled independently of everything else. $Y'$ has some new everywhere positive density $q'(\cdot; \nu_t, \sigma, E_t, \omega_1, \ldots, \omega_M)$ over $\mathbb{R}^{k \times \Pi_{i=1}^N n_i}$.

Now for any sets $B_1, \ldots, B_M \subset \mathbb{R}^d$, we have

$$
\begin{align*}
\mathbb{E} \{ \mathbb{P} \left( \omega_{xt} \in B \forall x | f_{t+1} \right) | f_t \} \\
= \mathbb{E} \left\{ \mathbb{P} \left( \omega_{xt} \in B \forall x | f_t \cup Y_t' \right) | f_t \right\} \\
= \mathbb{E} \left\{ \mathbb{P} \left( \omega_{xt} \in B \forall x | f_t \cup Y \right) | f_t \right\} \\
= \mathbb{E} \left\{ \frac{\mathbb{P} \left( \omega_{xt} \in B \forall x, Y' | f_t \right)}{\mathbb{P} (Y' | f_t)} \right\} \\
= \frac{\int_{B_1 \times \cdots \times B_M} q'(y'; \nu_t, \sigma, E_t, \omega_1, \ldots, \omega_M) d\nu_t(\omega_1, \ldots, \omega_M)}{\int q'(y'; \nu_t, \sigma, E_t, \omega_1, \ldots, \omega_M) d\nu_t(\omega_1, \ldots, \omega_M)} \\
\end{align*}
$$

Finally we put the two components together. For any sets $A_1, \ldots, A_M \subset \mathbb{R}^d$,

$$
\begin{align*}
\mathbb{E} \{ \nu_{t+1}(A_1, \ldots, A_M) | f_t \} \\
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{E} \{ \mathbb{P} \left( \omega_{xt} \in B \forall x | f_{t+1} \right) | f_t \} \phi(z_1) \cdots \phi(z_M) dz_1 \cdots dz_M \\
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \frac{\int_{B(z_1) \times \cdots \times B(z_M)} q'(y'; \nu_t, \sigma, E_t, \omega_1, \ldots, \omega_M) d\nu_t(\omega_1, \ldots, \omega_M)}{\int q'(y'; \nu_t, \sigma, E_t, \omega_1, \ldots, \omega_M) d\nu_t(\omega_1, \ldots, \omega_M)} \right\} \prod_{x=1}^{M} \phi(z_x) dz_x \\
\end{align*}
$$

Proof. Proof of Proposition 34 Wlog we prove the result for $i = 1$. By definition,

$$
V_{1t}(f_t, L_{1, t}; \chi, \eta_1, \eta_1, \eta_1, \sigma) = (1 - \beta) \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mathbb{E} \chi_{\tau, 1, \tau, \eta_1, \eta_1, \eta_1, \sigma, \tau} \left[ r_{1\tau} | f_t, L_{1, t} \right].
$$

Note that, under $\chi_1, \eta_1, \eta_1, \eta_1$ and $\sigma$,

$$
\begin{align*}
r_{1\tau} & = (1 + \rho L_{1, \tau, \tau}) R_0(\nu_\tau, x_{1, \tau}, n_{1, \tau}) \\
& = (1 + \rho L_{1, \chi, (\nu_\tau, L_\tau, \tau)} \mathbb{E} \chi_{\tau, 1, \tau, \eta_1, \eta_1, \eta_1, \sigma, \tau} \left[ r_{1\tau} | f_t, L_{1, t} \right].
\end{align*}
$$
For $\tau \geq t + 1$, we get
\[
\mathbb{E} \left[ r_{1\tau} | f_t, L_{1:t} \right] = \mathbb{E} \left\{ \mathbb{E} \left[ B(\nu, L_{1:t}) | f_{\tau-1}, L_{1:(\tau-1)} \right] | f_t, L_{1:t} \right\} \\
= \mathbb{E} \left\{ A[B](\nu_{\tau-1}, L_{1:(\tau-1)}) | f_t, L_{1:t} \right\}.
\]
where for any function $F : B(\mathbb{R}^{d \times M}) \times \{0, 1\}^M \rightarrow \mathbb{R}$,
\[
A[F](\nu, L) = 2^{-MN} \sum_{L', L_2, \ldots, L_N \in \{0, 1\}^M} \int F(\nu', L') \, dP(\nu', \sigma, E(\nu, L, L_2, \ldots, L_N)),
\]
and
\[
E(\nu, L, L_2, \ldots, L_N) = \{ \chi_1(\nu, L), \eta_1(\nu, L), \chi_2(\nu, L_2), \eta_2(\nu, L_2), \ldots, \chi_N(\nu, L_N), \eta_N(\nu, L_N) \}.
\]
Iterating we get that $\mathbb{E} [r_{1\tau} | f_t, L_{1:t}] = A^{\tau-t}[B](\nu_t, L_{1:t})$.

All in all,
\[
V_{1t}(f_t, L_{1:t}; \chi_1, \eta_1, \chi_{-1}, \eta_{-1}, \sigma) = (1 - \beta) \sum_{\tau=t}^{\infty} \beta^{\tau-t} A^{\tau-t}[B](\nu_t, L_{1:t}) \\
= (1 - \beta) \sum_{\tau'=0}^{\infty} \beta^{\tau'} A^{\tau'}[B](\nu_t, L_{1:t}) \\
:= W_1(\nu_t, L_{1:t}; \chi_1, \eta_1, \chi_{-1}, \eta_{-1}, \sigma).
\]

\[\square\]

\textbf{Proof.} Proof of Proposition \[\text{2}]. By the definition of $W$ (see Proposition \[\text{3}\]), and noting that $h_0 = f_0$, we have
\[
\bar{W}_1(N_0(h_0); \chi^N, \eta^N, \sigma) = 2^{-M} \sum_{L \in \{0, 1\}^M} W_1(N_0(h_0), L; \chi^N, \eta^N, \sigma) \\
= 2^{-M} \sum_{L \in \{0, 1\}^M} V_{10}(h_0, L; \chi, \eta, \chi_{-1}^N, \eta_{-1}^N, \sigma) \\
= 2^{-M} \sum_{L \in \{0, 1\}^M} (1 - \beta) \mathbb{E}_{\chi^N, \eta^N, \sigma} \left[ \sum_{t=0}^{\infty} \beta^t r_{1t} \left| h_0, \{L_{1:0} = L\} \right. \right] \\
= (1 - \beta) \mathbb{E}_{\chi^N, \eta^N, \sigma} \left[ \sum_{t=0}^{\infty} \beta^t r_{1t} \left| h_0 \right. \right] \\
= U_0(h_0; \chi, \eta, \sigma). \]

\[\square\]
APPENDIX F. CONSTRUCTING THE EXPERIMENT DESIGN GAME

Proof. Proof of Proposition \( \mathbb{F} \)

1. Fix any mechanism \( \sigma \), and let

\[
\chi^*_t : F_t(\sigma) \to \{1, \ldots, M\}, \quad \eta^*_t : F_t(\sigma) \to \mathbb{N}.
\]

be a public strategy that maximizes the high-impact payoff, \( U^*_0(h_0; (\chi^*_t)^*, (\eta^*_t)^*), \sigma \). Now suppose that every agent plays \( (\chi^*_t)^*, (\eta^*_t)^* \), and let \( h^*_t \) and \( f^*_t = \Sigma_t(h^*_t; \sigma) \) be the unfiltered and filtered histories generated, starting from the initial history \( h_0 \). Then the chosen experiment designs are given by \( x^*_t = \chi^*_t(f^*_t) \), \( n^*_t = \eta^*_t(f^*_t) \) (these designs are common to all agents since all agents play the same public strategy). Denoting the decisions selected in each experiment by \( \delta^*_t \), we have that

\[
U^*_0(h_0; (\chi^*_t)^*, (\eta^*_t)^*), \sigma) = (1 - \beta)(1 + \rho)E_{(\chi^*_t)^*, (\eta^*_t)^*, \sigma} \left[ \sum_{t=0}^{\infty} \beta^t R(n^*_t, \delta^*_t, \theta_{\chi^*_t, \eta^*_t}) \right] h_0
\]

\[
= (1 - \beta)(1 + \rho) \sum_{t=0}^{\infty} \beta^t \mathbb{E} \left[ R(n^*_t, \delta^*_t, \theta_{\chi^*_t, \eta^*_t}) \right] \left[ f^*_t, Y^{\uparrow t}_1, \ldots, Y^{\uparrow t}_{n^*_t} \right] | h_0
\]

where the last line follows from the definition of \( \delta^*_t \).

We need to prove that there exists a public strategy \( (\tilde{\chi}_t), (\tilde{\eta}_t) \) under \( \sigma^0 \), such that for all \( h_0 \),

\[
U^*_0(h_0; (\chi^*_t)^*, (\eta^*_t)^*), \sigma) \leq U^*_0(h_0; (\tilde{\chi}_t), (\tilde{\eta}_t), \sigma^0).
\]

Once we have done so, we can say that

\[
U^*_0(h_0; (\chi^*_t)^*, (\eta^*_t)^*), \sigma) \leq \max_{(\chi^*_t)^*, (\eta^*_t)^*} U^*_0(h_0; (\chi^*_t)^*, (\eta^*_t)^*), \sigma^0),
\]

and so taking a max over the mechanism \( \sigma \) will deliver the desired result.

To that end, we take

\[
\tilde{\chi}_t = \Sigma_t(\cdot; \sigma) \circ \chi^*_t, \quad \tilde{\eta}_t = \Sigma_t(\cdot; \sigma) \circ \eta^*_t.
\]

That is, at each period, this new strategy first maps the filtered history (equivalently the unfiltered history) obtained under the non-filtering mechanism to the filtered history that would have been available to the agents under \( \sigma \), before applying \( \chi^*_t \) and \( \eta^*_t \) to arrive at an experiment design. We note here that each \( \Sigma_t(\cdot; \sigma) \) is defined on the set \( H_t \), which coincides with the set \( F_t(\sigma^0) \). Thus \( \tilde{\chi}_t \) and \( \tilde{\eta}_t \) define functions from \( F_t(\sigma^0) \) to \( \{1, \ldots, M\} \) and \( \mathbb{N} \) respectively, as is required for a public strategy under \( \sigma^0 \).
Appendix F. Constructing the Experiment Design Game

2. Suppose for contradiction that there exist initial histories, $h^*_0$, $h^*_0 \in H_0$, where $N_0(h^*_0) = N_0(h^*_0) = \nu_0$ (for some $\nu_0$) but $U^*_0(h^*_0) < U^*_0(h^*_0)$.

Let $(\chi^*_t), (\eta^*_t)$ be a public strategy under $\sigma^0$ such that

$$U^*_0(h^*_0) = U^*_0(h^*_0; (\chi^*_t), (\eta^*_t), \sigma^0).$$
In the case that the company selects \( \sigma^0 \), the initial history is given by \( h_0 = h_0^B \), and all agents play \((\chi^*, \eta^*)\), let \( h_t = h_t^B \) be the sequence of histories generated. Every agent then chooses the experiment designs, \( x_t^* = \chi_t^*(h_t^B), n_t^* = \eta_t^*(h_t^B) \). To distinguish the dependence of these designs on the initial history from dependence on data collected at \( t \geq 0 \), we define \( \tilde{h}_t = h_t \backslash h_0 \) and write
\[
x_t^* = \chi_t^*(h_0^B, \tilde{h}_t^B), \quad n_t^* = \eta_t^*(h_0^B, \tilde{h}_t^B).
\]

Preceding as in the first part of this proof, we find that
\[
U_0^*(h_0^B) = U_0^*(h_0^B; (\chi^*_t), (\eta^*_t), \sigma^0)
\]
\[
= (1 - \beta)(1 + \rho) \sum_{t=0}^{\infty} \beta^t \left[ \max_{\delta} \mathbb{E} \left[ R(n_{1t}, \delta, \theta_{1x_t};) \mid h_t = h_t^B, Y_{1t}^1, \ldots, Y_{nt}^1 \right] \mid h_0 = h_0^B \right]
\]
\[
= (1 - \beta)(1 + \rho) \sum_{t=0}^{\infty} \beta^t \mathbb{E} \left[ J(h_0^B, \tilde{h}_t^B, Y_{1t}^1, \ldots, Y_{nt}^1; (\chi^*_t), (\eta^*_t)) \mid h_0 = h_0^B \right]
\]
\[
= (1 - \beta)(1 + \rho) \sum_{t=0}^{\infty} \beta^t \int \mathbb{E} \left[ J(h_0^B, \tilde{h}_t^B, Y_{1t}^1, \ldots, Y_{nt}^1; (\chi^*_t), (\eta^*_t)) \mid \omega_0 \right] d\nu_0(\omega_0),
\]
where for any \( h_t = (h_0, h_t) \in H_t \), and any public strategy \((\chi_t), (\eta_t)\),
\[
J(h_0^B, \tilde{h}_t^B, Y_{1t}^1, \ldots, Y_{nt}^1; (\chi_t), (\eta_t))
\]
\[
= \max_{\delta} \mathbb{E} \left[ R(\eta_t(h_0, \tilde{h}_t), \delta, \theta_{1x_t(h_0, \tilde{h}_t)}; \mid h_0, \tilde{h}_t, Y_{1t}^1, \ldots, Y_{nt}^1 \right]
\]
\[
= \max_{\delta} \left\{ \int \int R(\eta_t(h_0, \tilde{h}_t), \delta, \theta) d\tilde{G}(\omega; N_t(h_0)) \right\}.
\]
Here \( \tilde{G}(\cdot; \nu) \) denotes the posterior for the vector \( \omega_t = (\omega_{1t}, \ldots, \omega_{Mt}) \) when \( \omega_0 \sim \nu \), after conditioning on \( \tilde{h}_t \) and \( Y_{1t}^1, \ldots, Y_{nt}^1 \). \( \tilde{G}(\cdot; \theta) \) denotes the posterior for \( \theta \), when \( \theta \sim G(\cdot) \) holds marginally and \( Y_1^1, \ldots, Y_{nt}^1 \) has been observed.

Now we will present another public strategy \((\tilde{\chi}_t), (\tilde{\eta}_t)\), under \( \sigma^0 \). Intuitively, the new strategy maps the current history at each time to an imaginary history where \( h_0 \) is replaced by \( h_0^B \), before selecting an experiment design according to \((\chi^*_t, \eta^*_t)\). Formally, for any \( h_t = (h_0, \tilde{h}_t) \in H_t \), we set
\[
\tilde{\chi}_t(h_0, \tilde{h}_t) = \chi_t^*(h_0^B, \tilde{h}_t), \tilde{\eta}_t(h_0, \tilde{h}_t) = \eta_t^*(h_0^B, \tilde{h}_t).
\]

Suppose that every agent plays this new strategy. By induction over \( t \), we find that for any initial history \( h_0 \), we obtain \( \tilde{h}_t = h_t^B \). Further, every agent selects the sequence of experiment designs, \((x_t^*, n_t^*)\). We will obtain a contradiction, because under the same experiment designs the same average rewards are obtained, whether we condition on \( h_0 = h_0^A \) or \( h_0 = h_0^B \).
As above,

\[
U^T_0(h^A_0; \hat{\chi}_t, (\hat{\eta}_t), \sigma^0)
= (1 - \beta)(1 + \rho) \sum_{t=0}^{\infty} \beta^t \mathbb{E} \left[ \max_{\hat{\delta}} \mathbb{E} \left[ R(n^A_{1t}, \hat{\delta}, \theta_{1x_{t+1}}, t) \right] \left| h_0 = h^A_0, \hat{h}_t = h^B_t, Y^1_{1t}, \ldots Y^T_{nt} \right] \right] \left| h_0 = h^A_0 \right]
= (1 - \beta)(1 + \rho) \sum_{t=0}^{\infty} \beta^t \int \mathbb{E} \left[ J(h^A_0, h^B_t, Y^1_{1t}, \ldots Y^T_{nt}; (\hat{\chi}_t), (\hat{\eta}_t)) \right] \mid \omega_0 \rangle \, d\nu_0(\omega_0).
\]

Further,

\[
J(h^A_0, h^B_t, Y^1_{1t}, \ldots Y^T_{nt}; (\hat{\chi}_t), (\hat{\eta}_t))
= \max_{\hat{\delta}} \left\{ \int \int R(\eta^A_{t+1}, h^B_t, \hat{\delta}, \theta) \, d\tilde{G}(\omega_{\hat{\chi}^0(h^A_0, h^B_0, \hat{\delta}, \theta)(\hat{\eta}_0)}) \, d\tilde{v}_t(\omega_t; \nu_0) \right\}
= \max_{\hat{\delta}} \left\{ \int \int R(\eta^A_{t+1}, h^B_t, \hat{\delta}, \theta) \, d\tilde{G}(\omega_{\hat{\chi}^0(h^A_0, h^B_0, \hat{\delta}, \theta)(\hat{\eta}_0)}) \, d\tilde{v}_t(\omega_t; \nu_0) \right\}
= J(h^B_t, h^B_t, Y^1_{1t}, \ldots Y^T_{nt}; (\hat{\chi}_t), (\hat{\eta}_t)).
\]

We conclude that

\[
U^T_0(h^A_0) \geq U^T_0(h^A_0; \hat{\chi}_t, (\hat{\eta}_t), \sigma^0) = U^T_0(h^B_t; (\hat{\chi}_t), (\hat{\eta}_t), \sigma^0) = U^T_0(h^B_t),
\]

which gives the desired contradiction.

\[\square\]

**Proof.** Proof of Proposition 14. First suppose for contradiction that there exist \( t \geq 0, \nu_0 \in B(\mathbb{R}^{d \times M}), h^B_t \in H_t, \) and a public strategy \((\chi^*, \eta^*)\) under \( \sigma^0, \) such that the following both hold:

1. \( N_t(h^B_t; \sigma^0) = \nu_0. \)
2. \( W^*(\nu_0) < U^T_0(h^B_t; (x^*_{t+\tau})_{\tau \geq 0}, (\eta^*_{t+\tau})_{\tau \geq 0}, \sigma^0). \)

Then let \( h^A_0 \in H_0 \) be any initial history such that \( N_0(h^A_0) = \nu_0. \) Let \( (h^B_{t+\tau})_{\tau \geq 0} \) be the histories generated from period \( t \) onwards, when the company selects \( \sigma^0, \) all agents play \((\chi^*, \eta^*), \) and \( h_t = h^B_t. \) Thus every agent selects the experiment designs, \( x^*_{t+\tau} = \chi^*_{t+\tau}(h^B_{t+\tau}), n^*_{t+\tau} = \eta^*_{t+\tau}(h^B_{t+\tau}). \)

As in the proof of Proposition 14, for each \( \tau \geq 0, \) we define \( \tilde{h}_\tau = h_{t+\tau}. \) Let \( \hat{H}_\tau \) be the set of all values that the set \( \tilde{h}_\tau \) may take, given arbitrary sequences of experiment designs and any values for the observed data. Further, for each \( \tau \geq 0, \) we let \( \hat{h}_\tau = h_{t+\tau} \setminus h_0. \) Note that the set of possible values for \( \hat{h}_\tau \) corresponds to the same set, \( \hat{H}_\tau. \) In all, each element of \( H_{t+\tau} \) is thus decomposed into three disjoint (possibly empty) sets: \( h_{t+\tau} = h_0 \cup \hat{h}_t \cup \hat{h}_\tau, \) where \( h_0 \in H_0, \hat{h}_t \in \hat{H}_t, \) and \( \hat{h}_\tau \in \hat{H}_\tau. \)
We define a new public strategy \((\tilde{\chi}, \tilde{\eta})\) under \(\sigma^0\), where for each \(\tau \geq 0\) and \(h_\tau = h_0 \cup \tilde{h}_\tau \in H_\tau\),

\[
\tilde{\chi}(h_0 \cup \tilde{h}_\tau) = \chi^*_t(h^B_0, \tilde{h}^B_\tau, \tilde{h}_\tau), \quad \tilde{\eta}(h_0 \cup \tilde{h}_\tau) = \eta^*_t(h^B_0, \tilde{h}^B_\tau, \tilde{h}_\tau).
\]

Informally, at each time \(\tau\), the new strategy imagines that the current period is \(t + \tau\) and constructs a fake history up to that period. This fake history supposes that the data, which truly was collected between times 0 and \(\tau\), in fact was collected between \(t\) and \(t + \tau\). This is combined with \(h^B_\tau\), which is taken as the history up to time \(t\). Finally, \(\chi^*_t\) and \(\eta^*_t\) are applied to this fake history to obtain an experiment design.

Now suppose that the company selects \(\sigma^0\), all agents play \((\tilde{\chi}, \tilde{\eta})\), and \(h_0 = h^A_0\). By induction on \(\tau\), we see that the histories generated are given by \(h_\tau = h^B_{t+\tau}\), and that the agents select the experiment designs, \(x_{i\tau} = x^*_{i+\tau}, n_{i\tau} = n^*_{i+\tau}\). Proceeding by the same arguments as in part 2. of Proposition \(\Box\), we find the rewards obtained are equal in expectation to the rewards obtained from period \(t\) onwards, in the case that all agents play \((\chi^*, \eta^*)\), and \(h_t = h^B_t\). Specifically, we find that

\[
U^*_t(h^A_0; \tilde{\chi}, \tilde{\eta}, \sigma^0) = U^*_t(h^B_t; (\chi^*)_{\tau \geq 0}, (\eta^*)_{\tau \geq 0}, \sigma^0).
\]

This gives a contradiction, because

\[
U^*_0(h^A_0; \tilde{\chi}, \tilde{\eta}, \sigma^0) \leq U^*_0(h^A_0) = W^*(\mathcal{N}_0(h^A_0)) = W^*(\nu_0).
\]

Conversely, suppose that there exist \(t \geq 0, \nu_0 \in \mathcal{B}(\mathbb{R}^{d \times M})\), and \(h^B_t \in H_t\), such that:

1. \(\mathcal{N}_t(h^B_t; \sigma^0) = \nu_0\).
2. \(U^*_t(h^B_t) < W^*(\nu_0)\).

Again let \(h^A_0 \in H_0\) be any initial history such that \(\mathcal{N}_0(h^A_0) = \nu_0\). Since \(W^*(\nu_0) = U^*_0(h^A_0)\), this implies that there exists a public strategy \((\chi^*, \eta^*)\) under \(\sigma^0\), such that

\[
U^*_t(h^B_t) < U^*_0(h^A_0; \chi^*, \eta^*, \sigma^0).
\]

Now we define a new public strategy \((\tilde{\chi}, \tilde{\eta})\) as follows. Up to period \(t\), this strategy can be set arbitrarily. For \(\tau \geq 0\) and \(h_{t+\tau} = (h_0, \tilde{h}_t, \tilde{h}_\tau) \in H_{t+\tau}\), we set

\[
\tilde{\chi}_{t+\tau}(h_0, \tilde{h}_t, \tilde{h}_\tau) = \chi^*_t(h^A_0, \tilde{h}_\tau), \quad \tilde{\eta}_{t+\tau}(h_0, \tilde{h}_t, \tilde{h}_\tau) = \eta^*_t(h^A_0, \tilde{h}_\tau).
\]

Arguing as before that the new strategy produces the same experiment designs and expected rewards, albeit at a lag of \(t\) periods, we find:

\[
U^*_t(h^B_t; \tilde{\chi}, \tilde{\eta}, \sigma^0) = U^*_0(h^A_0; \chi^*, \eta^*, \sigma^0).
\]
This gives a contradiction because $U^*_t(h^B) \geq U^*_t(h^B; \tilde{x}, \tilde{n}, \sigma^0)$.

Proof. Proof of Proposition 16. For part 1., since employees are interchangeable, it is sufficient to prove that $(\chi, \eta)$ is the optimal strategy for agent 1 when $\chi_{-1} = \chi^{N-1}, \eta_{-1} = \eta^{N-1}$. For this, it is sufficient to prove that there is no time where she can improve her payoff through a “one-step deviation”. Specifically, we need to show that for all $t, f_t \in F_t(\sigma)$ and $L_{1,t} \in \{0, 1\}^M$, the continuation payoff

$$V_{1t}(f_t, L_{1,t}; (x, \chi), (n, \eta), \chi^{N-1}, \eta^{N-1}, \sigma)$$

is maximized at $x = \chi(N_t(f_t), L_{1,t}), n = \eta(N_t(f_t), L_{1,t})$. Here $(x, \chi)$ and $(n, \eta)$ denote the strategy of selecting $x$ and $n$ at time $t$, before playing $\chi, \eta$ from time $t+1$ onwards.

To that end,

$$V_{1t}(f_t, L_{1,t}; (x, \chi), (n, \eta), \chi^{N-1}, \eta^{N-1}, \sigma)$$

$$= (1 - \beta)E_{(x, \chi), (n, \eta), \chi^{N-1}, \eta^{N-1}, \sigma} \left[ \sum_{\tau = t}^{\infty} \beta^{\tau-t} r_{1\tau} \left| f_t, L_{1,t} \right. \right]$$

$$= (1 - \beta)(1 + \rho L_{1xt})R_0(\nu_t, x, n)$$

$$+ \beta E_{x,n,\sigma} \left\{ (1 - \beta)E_{\chi^{N}, \eta^{N}, \sigma} \left[ \sum_{\tau = t+1}^{\infty} \beta^{\tau-(t+1)} r_{1\tau} \left| f_{t+1}, \right. \right] f_t \right\}$$

$$= (1 - \beta)(1 + \rho L_{1xt})R_0(\nu_t, x, n) + \beta E_{x,n,\sigma} \left\{ \bar{W}_1(\nu_{t+1}, \chi^N, \eta^N, \sigma) \right\} f_t$$

$$= (1 - \beta)(1 + \rho L_{1xt})R_0(\nu_t, x, n)$$

$$+ 2^{-M(N-1)} \beta \sum_{L_2, \ldots, L_N \in \{0, 1\}^M} \int \bar{W}_1(\nu'; \chi^N, \eta^N, \sigma) dP(\nu'; \nu_t, \sigma, E(x, n, \nu_t, L_2, \ldots, L_N))$$

which is maximized at $x = \chi(\nu_t, L_{1,t}), n = \eta(\nu_t, L_{1,t})$ by assumption.

Part 2. follows because a discounted dynamic program with a finite action space must have a unique solution (see e.g. [52]).

Proof. Proof of Proposition 17.
1. Fix $\nu$, and let $h_0 \in H_0$ be an initial history such that $N_0(h_0) = \nu$. Then we have:

$$W^*(\nu) = \max_{(\chi_t), (\eta_t)} U^*_0(h_0; (\chi_t), (\eta_t), \sigma^0)$$

$$= (1 - \beta)(1 + \rho) \max_{x, n, (\chi_t), (\eta_t)} E_{(x, \chi_t)^N, (n, \eta_t)^N, \sigma^0} \left[ \sum_{t=0}^{\infty} \beta^t R(n_{1t}, \delta_{1t}, \theta_{1t}) \big| h_0 \right]$$

$$= (1 - \beta)(1 + \rho) \max_{x, n} \left\{ R_0(\nu, x, n) + \beta \mathbb{E}_{x, n, (\chi_t), (\eta_t)} \left[ \max_{(\chi_t), (\eta_t)} E_{(x, \chi_t)^N, (n, \eta_t)^N, \sigma^0} \left[ \sum_{t=1}^{\infty} \beta^{t-1} R_0(n_{1t}, \delta_{1t}, x_{1t}) \big| h_1 \right] \big| h_0 \right\}$$

$$= \max_{x, n} \left\{ (1 - \beta)(1 + \rho) R_0(\nu, x, n) \right\}$$

$$= \max_{x, n} \left\{ (1 - \beta)(1 + \rho) R_0(\nu, x, n) + \beta \mathbb{E}_{x, n, \nu, \sigma^0} \left[ W^*_1(h_1; (\chi_t), (\eta_t), \sigma^0) \big| h_0 \right] \right\}$$

$$= \max_{x, n} \left\{ (1 - \beta)(1 + \rho) R_0(\nu, x, n) + \beta \int W^*(\nu') d\mathbb{P}_0 (\nu'; \nu, \{x, n\}^N) \right\}$$

The penultimate equality follows from Proposition 15, while the final equality holds by Proposition 11.

2. This follows from standard theory on discounted dynamic programming.
Appendix G

Performance of the Minimum Learning Mechanism

We begin with a lemma, which we shall leverage several times when proving Theorem 8. The MLM punishes deviations from \((\chi^{opt}, \eta^{opt})\) by threatening to share less data with the employees. In short, this lemma says that filtering data constitutes a threat, because in expectation it leads to a lower continuation payoff for each agent at the next period.

**Lemma 5.** Let \(E = \{x_1, n_1, \ldots, x_N, n_N\}\) and \(E' = \{x_1', n_1', \ldots, x_N', n_N'\}\) be any collections of experiment designs, where for each employee \(i\) the experiment types coincide but \(n_i \leq n_i'\). Then the benchmark payoff \(W^*\) satisfies the following inequality: for any \(\nu_0 \in B(\mathbb{R}^{d \times M})\),

\[
\int W^*(\nu')dP_0(\nu'; \nu_0, E) \leq \int W^*(\nu')dP_0(\nu'; \nu_0, E').
\]

**Proof.** Let \(h_0^A \in H_0\) be any initial history such that \(N_0(h_0^A) = \nu_0\), and let \(h_1 = h_1^F\) be the unfiltered history generated, when \(h_0 = h_0^A\) and the agents choose designs \(E\) at \(t = 0\). By Proposition 11, followed by Proposition 15, we have that

\[
\int W^*(\nu')dP_0(\nu'; \nu_0, E) = \mathbb{E} \left[ W^*(N_1(h_1^F; \sigma^0)) \middle| h_0 = h_0^A \right]
\]

\[
= \mathbb{E} \left\{ \max_{\chi, \eta} U_1^T(h_1^F; \chi, \eta, \sigma^0) \middle| h_0 = h_0^A \right\}.
\]

Let \((\chi^*, \eta^*)\) be a public strategy where the max is obtained. For \(t \geq 1\), let \(h_t^*\) be the histories generated and let \(x_t^*, n_t^*\) be the common designs chosen, in the case that the company selects \(\sigma^0\), all agents play \((\chi^*, \eta^*)\) from \(t = 1\) onwards, and \(h_1 = h_1^F\). Then by the same argument as in the proof
of Proposition 10, we find that

$$U_1^t(h_1^E; x^*, \eta^*, \sigma^0)$$

$$= (1 - \beta)(1 + \rho) \sum_{i=1}^{\infty} \beta^{t-1} \mathbb{E} \left[ \max_{\delta} \mathbb{E} \left[ R(n^*_t, \delta, \theta_1 x_t^*) \mid h_t = h^*_1, Y_1^t, \ldots, Y_{n^*_t}^t \right] \mid h_1 = h^E_1 \right].$$

Thus,

$$\int W^*(\nu')dP_0(\nu' ; \nu_0, E)$$

$$= (1 - \beta)(1 + \rho) \sum_{i=1}^{\infty} \beta^{t-1} \mathbb{E} \left[ \max_{\delta} \mathbb{E} \left[ R(n^*_t, \delta, \theta_1 x_t^*) \mid h_t = h^*_1, Y_1^t, \ldots, Y_{n^*_t}^t \right] \mid h_0 = h^A_0 \right].$$

For any $h_1 \in H_1$, let $\tilde{h}_1 = h_1 \setminus h_0$, and for any $t \geq 1$ and $h_t \in H_t$, let $\hat{h}_t = h_t \setminus h_1$. Thus for any $t \geq 1$, each element of $H_t$ is decomposed as $h_t = h_0 \cup \tilde{h}_1 \cup \hat{h}_t$. Now we define a new public strategy $(\tilde{\chi}, \tilde{\eta})$ under $\sigma^0$ as follows. $\tilde{\chi}_0$ and $\tilde{\eta}_0$ are set arbitrarily. For any $t \geq 1$ and $h_t = (h_0, \tilde{h}_1, \hat{h}_t) \in H_t$, we set

$$\tilde{\chi}_t(h_t) = \chi^*_t(h_0, \tilde{h}_1 \cap (h^E_1 \setminus h^A_0), \hat{h}_t), \quad \tilde{\eta}_t(h_t) = \eta^*_t(h_0, \tilde{h}_1 \cap (h^E_1 \setminus h^A_0), \hat{h}_t).$$

That is, for any $t \geq 1$, the new strategy maps the history available to a smaller pseudo-history by disregarding any data that was collected at $t = 0$, which would not have been observed if the agents had selected designs $E$ at $t = 0$. Then $(\chi^*, \eta^*)$ is applied to this pseudo-history to obtain an experiment design.

Suppose that the company selects $\sigma^0$, $h_0 = h^A_0$, and the agents choose the designs $E'$ at $t = 0$, whereafter they all play $(\tilde{\chi}, \tilde{\eta})$. Let $h_1 = h^E_1'$ be the history obtained at $t = 1$. For any realizations of the observable data, we have that $h_1^E \subset h_1^{E'}$, so the experiment designs at $t = 1$ are given by

$$\tilde{\chi}_1(h_1^{E'}) = \chi^*_1(h_0^A, h_1^E \setminus h^A_0) = x^*_1, \quad \tilde{\eta}_1(h_1^{E'}) = \eta^*_1(h_0^A, h_1^E \setminus h^A_0) = n^*_1.$$

In other words, the designs at $t = 1$ coincide with the designs that were selected in the original case, where the agents choose $E$ at $t = 0$ before all playing the strategy $(\chi^*, \eta^*)$. Since the same data is collected at $t = 1$, we get that $h_2 = h_1^{E'} \cup \tilde{h}_2 = h^*_2 \cup h^+_1$, where $h^+_1 = h_1^{E'} \setminus h^E_1$. In fact, induction establishes that for all $t \geq 1$, the generated history is given by $h_t = h^*_t \cup h^+_1$, while every agent selects the design, $x^*_t, n^*_t$.

Finally, we get the desired inequality by arguing that, while $E'$ followed by $(\tilde{\chi}, \tilde{\eta})$ obtains the same experiment designs as $E$ followed by $(\chi^*, \eta^*)$, the former obtains better expected rewards,
because it makes more data available at each period for optimizing decisions:

\[
\int W^*(\nu')dP_0(\nu';\nu_0, E) \\
= \mathbb{E} \left[ W^*(\mathcal{N}_1(h_1^{E'}, \sigma^0)) \mid h_0 = h_0^A \right] \\
\geq \mathbb{E} \left\{ U_1^{h_1^{E'}, \hat{x}, \hat{n}, \sigma^0} \mid h_0 = h_0^A \right\} \\
= (1 - \beta)(1 + \rho) \sum_{t=1}^{\infty} \beta^{t-1} \mathbb{E} \left[ \max_\delta R(n_{t1}^*, \delta, \theta_{1\chi_{t1}}) \mid h_t = h_t^* \cup h_1^+, Y_{1t}^l, \ldots Y_{n_{t1}}^l \right] \mid h_0 = h_0^A \\
\geq (1 - \beta)(1 + \rho) \sum_{t=1}^{\infty} \beta^{t-1} \mathbb{E} \left[ R(n_{t1}^*, \delta, \theta_{1\chi_{t1}}) \mid h_t = h_t^* \cup h_1^+, Y_{1t}^l, \ldots Y_{n_{t1}}^l \right] \mid h_0 = h_0^A \\
= \int W^*(\nu')dP_0(\nu';\nu_0, E).
\]

\[\square\]

**Proof.** Proof of Theorem 8. By Proposition 11, we need to show that \( x = \chi^{\text{opt}}(\nu), n = \eta^{\text{opt}}(\nu) \), and \( w(\nu) = W^*(\nu) \) solve the following dynamic program:

\[
w(\nu) = \max_{x,n} \left\{ (1 - \beta)R_0(\nu, x, n) + \int w(\nu')dP(\nu';\nu, \sigma^{MLM}, E(x,n,\nu)) \right\},
\]

where

\[
E(x,n,\nu) = \{ x, n, \chi^{\text{opt}}(\nu), \eta^{\text{opt}}(\nu), \ldots, \chi^{\text{opt}}(\nu), \eta^{\text{opt}}(\nu) \}.
\]

Let

\[
A(x,n,\nu) = (1 - \beta)R_0(\nu, x, n) + \int W(\nu')dP(\nu';\nu, \sigma^{MLM}, E(x,n,\nu)).
\]

We break down evaluating \( A(x,n,\nu) \) into three cases. By the definition of the MLM, we have that

\[
P(\nu';\nu, \sigma^{MLM}, E(x,n,\nu)) = \begin{cases} 
\mathcal{P}_0(\cdot;\nu, \{x,n'\}^N) & x = \chi^{\text{opt}}(\nu) \\
\mathcal{P}_0(\cdot;\nu, \{x,n\} \times \{0\}^{N-1}) & R_0(\nu, x, n) > R_0(\nu, \chi^{\text{opt}}(\nu), \eta^{\text{opt}}(\nu)) \\
\mathcal{P}_0(\cdot;\nu, E(\cdot,0,\nu)) & R_0(\nu, x, n) < R_0(\nu, \chi^{\text{opt}}(\nu), \eta^{\text{opt}}(\nu)) 
\end{cases}
\]

where \( n' = \min(n, \eta^{\text{opt}}(\nu)) \). Dots in the place of experiment types in the above expression correspond to the fact the experiment type is irrelevant if the sample size is chosen to be zero.

**Case I:** \( x = \chi^{\text{opt}}(\nu) \).
By Lemma 1, we have that

\[ A(\chi_{\text{opt}}(\nu), n, \nu) = (1 - \beta)R_0(\nu, \chi_{\text{opt}}(\nu), n) + \beta \int W^*(\nu') dP_0(\nu, \{\chi_{\text{opt}}(\nu'), n\}^N) \]

\[ \leq (1 - \beta)R_0(\nu, \chi_{\text{opt}}(\nu), n) + \beta \int W^*(\nu') dP_0(\nu, \{\chi_{\text{opt}}(\nu), n\}^N). \]

If \( n = \eta_{\text{opt}}(\nu) \), we have equality here. In that case, the value is equal to \( W^*(\nu) \) (see Proposition 1).

Case II: \( R_0(\nu, x, n) > R_0(\nu, \chi_{\text{opt}}(\nu), \eta_{\text{opt}}(\nu)) \).

Appealing to Lemma 1 again,

\[ A(x, n, \nu) = (1 - \beta)R_0(\nu, x, n) + \beta \int W^*(\nu') dP_0(\nu, \{x, n\} \times \{0\}^{N-1}) \]

\[ \leq (1 - \beta)R_0(\nu, x, n) + \beta \int W^*(\nu') dP_0(\nu, \{x, n\}^N) \]

\[ \leq (1 - \beta)R_0(\nu, \chi_{\text{opt}}(\nu), \eta_{\text{opt}}(\nu)) + \beta \int W^*(\nu') dP_0(\nu, \{\chi_{\text{opt}}(\nu), \eta_{\text{opt}}(\nu\}^N) \]

\[ = W^*(\nu). \]

The penultimate line holds because \( \chi_{\text{opt}} \) and \( \eta_{\text{opt}} \) are defined to be maximizers of that expression.

Case III: \( R_0(\nu, x, n) < R_0(\nu, \chi_{\text{opt}}(\nu), \eta_{\text{opt}}(\nu)) \).

\[ A(x, n, \nu) = (1 - \beta)R_0(\nu, x, n) + \beta \int W^*(\nu') dP_0(\nu, \{x, n\} \times \{0\}^{N-1}) \]

\[ \leq (1 - \beta)R_0(\nu, x, n) + \beta \int W^*(\nu') dP_0(\nu, \{\chi_{\text{opt}}(\nu), \eta_{\text{opt}}(\nu\}^N) \]

\[ = W^*(\nu). \]

\[ \Box \]

Proof. Proof of Proposition 13. By Proposition 13, it is sufficient to find a function \( w(\nu, L) \) that, together with \( n(0) = n(1) = \eta^t(\nu) \), solves the following dynamic program:

\[ w(\nu, L) = \max_n \left\{ (1 - \beta)(1 + \rho L)R_0(\nu, n) + \beta \int \tilde{w}(\nu') dP(\nu'; \nu, \sigma^{\text{MLM}}, \{n, \eta^t(\nu)^{N-1}\}) \right\} \]

By the definition of the MLM, this dynamic program is equivalent to the following:

\[ w(\nu, L) = \max_n \left\{ (1 - \beta)(1 + \rho L)R_0(\nu, n) + \beta \int \tilde{w}(\nu') dP_0(\nu'; \nu, \min\{n, \eta^t(\nu)\}) \right\} \]
We will show that it is solved by

\[ w(\nu, 0) = \bar{W}_1^*(\nu) - \frac{1}{2}\rho(1 - \beta)R_0(\nu, \eta^*(\nu)), \quad w(\nu, 1) = \bar{W}_1^*(\nu) + \frac{1}{2}\rho(1 - \beta)R_0(\nu, \eta^*(\nu)). \]

In that case, \( \bar{w}(\nu) = \bar{W}_1^*(\nu) \).

Let

\[ A(n, \nu, L) = (1 - \beta)(1 + \rho L)R_0(\nu, n) + \beta \int \bar{W}_1^*(\nu')dP_0(\nu'; \nu, \min\{n, \eta^*(\nu)\})^N. \]

First we consider \( L = 0 \). For \( n \leq \eta^*(\nu) \), we have

\[ A(n, \nu, 0) = (1 - \beta)R_0(\nu, n) + \beta \int \bar{W}_1^*(\nu')dP_0(\nu'; \nu, n^N) \]

\[ = \left\{ (1 - \beta)(1 + \rho)R_0(\nu, n) + \beta \int \bar{W}_1^*(\nu')dP_0(\nu'; \nu, n^N) \right\} - \rho(1 - \beta)R_0(\nu, n) \]

By the definition of the high-impact strategy, the expression in parentheses is maximized over all \( n \) at \( n = \eta^*(\nu) \). The subtracted term is non-increasing in \( n \) as rewards are non-increasing in the sample size. Thus \( n = \eta^*(\nu) \) maximizes \( A(n, \nu, 0) \) over this range. For \( n \geq \eta^*(\nu) \),

\[ A(n, \nu, 0) = (1 - \beta)R_0(\nu, n) + \beta \int \bar{W}_1^*(\nu')dP_0(\nu'; \nu, \eta^*(\nu)^N) \]

is decreasing in \( n \), so we conclude that \( n = \eta^*(\nu) \) maximizes \( A(n, \nu, 0) \) over all \( n \). The maximal value is given by

\[ A(\eta^*(\nu), \nu, 0) = (1 - \beta)R_0(\nu, \eta^*(\nu)) + \beta \int \bar{W}_1^*(\nu')dP_0(\nu'; \nu, \eta^*(\nu)^N) \]

\[ = \bar{W}_1^*(\nu) - \frac{1}{2}\rho(1 - \beta)R_0(\nu, \eta^*(\nu)). \]

Now consider \( L = 1 \).

\[ A(n, \nu, 0) \leq (1 - \beta)(1 + \rho)R_0(\nu, n) + \beta \int \bar{W}_1^*(\nu')dP_0(\nu'; \nu, n^N) \]

with equality at \( n = \eta^*(\nu) \). This expression is maximized at \( n = \eta^*(\nu) \) by definition. In this case, the maximal value is seen to be \( \bar{W}_1^*(\nu) + \frac{1}{2}\rho(1 - \beta)R_0(\nu, \eta^*(\nu)) \) as desired.

\[ \square \]

Proof. Proof of Theorem 4. By propositions \( \text{K} \) and \( \text{K} \), we have that

\[ U_0(h_0; \eta^*, \sigma^{MLM}) = \bar{W}_1(N_0(h_0); (\eta^*)^N, \sigma^{MLM}) = \bar{W}_1(N_0(h_0)). \]
From Definition 1, we have that for any $\nu \in B(\mathbb{R}^{d \times M})$,

$$W^\uparrow_1(\nu) = \max_n \left\{ (1 - \beta)(1 + \rho)R_0(\nu, n) + \beta \int W^\uparrow_1(\nu')dP_0(\nu'; \nu, n^N) \right\} - \frac{1}{2}(1 - \beta)\rho R_0(\nu, \eta^\uparrow(\nu)).$$

Since rewards are at most one, the subtracted term is at most $\frac{1}{2}(1 - \beta)\rho$. Thus, comparing against $n = \eta^{opt}(\nu)$, we conclude:

$$W^\uparrow_1(\nu) \geq \left\{ (1 - \beta)(1 + \rho)R_0(\nu, \eta^{opt}(\nu)) + \beta \int W^\uparrow_1(\nu')dP_0(\nu'; \nu, \eta^{opt}(\nu)^N) \right\} - \frac{1}{2}(1 - \beta)\rho.$$

Iterating this inequality, we get that for any $\nu_0 \in B(\mathbb{R}^{d \times M})$,

$$W^\uparrow_1(\nu_0) \geq \left\{ (1 - \beta)(1 + \rho)R_0(\nu_0, \eta^{opt}(\nu_0)) + \beta \int W^\uparrow_1(\nu')dP_0(\nu'; \nu_0, \eta^{opt}(\nu_0)^N) \right\} - \frac{1}{2}(1 - \beta)\rho$$

$$\geq (1 - \beta)(1 + \rho)R_0(\nu_0, \eta^{opt}(\nu_0))$$

$$+ \beta \int \left\{ (1 - \beta)(1 + \rho)R_0(\nu_1, \eta^{opt}(\nu_1)) + \beta \int W^\uparrow_1(\nu')dP_0(\nu'; \nu_1, \eta^{opt}(\nu_1)^N) \right\} dP_0(\nu_1; \nu_0, \eta^{opt}(\nu_0)^N)$$

$$- \frac{1}{2}(1 - \beta)\rho - \frac{1}{2}\beta(1 - \beta)\rho$$

$$\geq \ldots$$

$$\geq (1 - \beta)(1 + \rho) \sum_{t=0}^{\infty} \mathbb{E}_{(\eta^{opt})^N, \sigma^0} \left[ R_0(\nu_t, \eta^{opt}(\nu_t)) \bigg| \nu_0 \right] - \frac{1}{2}(1 - \beta)\rho \sum_{t=0}^{\infty} \beta^t$$

$$= W^\ast(\nu_0) - \rho/2.$$

Setting $\nu_0 = N_0(h_0)$ gives the desired result. \qed