Master's thesis

Master's Programme in Mathematics and Statistics

Applying Thompson Sampling to Online Hypothesis Testing

Henri Suominen

January 20, 2021

Supervisor(s): Prof. Vanhatalo, Prof. Puolamäki

Examiner(s): Prof. Vanhatalo
Prof. Puolamäki

University of Helsinki
Faculty of Science

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master's Programme in Mathematics and Statistics

Tekijä — Författare — Author

Henri Suominen

Työn nimi — Arbetets titel — Title

Applying Thompson Sampling to Online Hypothesis Testing

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidantal — Number of pages
Master's thesis	January 20, 2021	56

Tiivistelmä — Referat — Abstract

Online hypothesis testing occurs in many branches of science. Most notably it is of use when there are too many hypotheses to test with traditional multiple hypothesis testing or when the hypotheses are created one-by-one. When testing multiple hypotheses one-by-one, the order in which the hypotheses are tested often has great influence to the power of the procedure.

In this thesis we investigate the applicability of reinforcement learning tools to solve the exploration – exploitation problem that often arises in online hypothesis testing. We show that a common reinforcement learning tool, Thompson sampling, can be used to gain a modest amount of power using a method for online hypothesis testing called alpha-investing. Finally we examine the size of this effect using both synthetic data and a practical case involving simulated data studying urban pollution.

We found that, by choosing the order of tested hypothesis with Thompson sampling, the power of alpha investing is improved. The level of improvement depends on the assumptions that the experimenter is willing to make and their validity. In a practical situation the presented procedure rejected up to 6.8 percentage points more hypotheses than testing the hypotheses in a random order.

Avainsanat — Nyckelord — Keywords

Multiple Hypothesis Testing, Online Hypothesis Testing, Reinforcement Learning

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Contents

1. Introduction

In abundance of data, a great challenge in data analysis is to separate useful information out of resource-wasting random patterns. A common method for limiting the number of false discoveries is to use the help of statistical hypothesis testing when observing a new pattern.

Statistical hypothesis testing allows us to weed out patterns by accepting only those hypotheses that are unlikely to have happened within the current model of the world. When done systematically, the probability of a false discovery can be bounded by an arbitrary constant. This constant is called the significance of the test.

The guard that hypothesis testing offers us against false discoveries is lost when multiple hypothesis tested at the same time. Say a single hypothesis has a 0.05 probability of being a false positive and 20 such test would be independently conducted, the probability of obtaining a false hypothesis would rise up to $1-(1-0.05)^{20} \approx 0.64$. In such case, one would be more likely to obtain a false positive than not! Therefore multiple testing procedures are required. These procedures work by testing each of the individual hypotheses at a lower level of significance in order to still control the total number of false discoveries with an arbitrary number of hypotheses.

Typically multiple hypothesis procedures are not designed to tackle hypotheses that come one at the time. For this type of problems we require online hypothesis testing procedures. A fundamental decision in these types of problems is deciding the order in which the hypotheses are tested at. The correct order of testing the hypotheses allows for more efficient use of significance and even more potentially more powerful tests. In practice this allows for more efficient online hypothesis testing procedures.

Online hypothesis testing is used extensively in multiple areas of science. Most notably, it has been proposed in clinical trials when determining early stopping of the trial for economic and ethical reasons [1]. It has also gained traction in interactive data exploration, where systems use online hypothesis testing procedures to prevent false discoveries [2]. Lastly, streams of hypotheses arise in many areas of science such as genomics and feature selection for high-dimensional models [3] which will benefit from more efficient online hypotheses testing procedures.

In this thesis we explore how reinforcement learning tools and prior assumptions on the correlation structures of the hypotheses can be leveraged to create more powerful online hypothesis testing procedures. We propose the use of Thompson sampling as this balances the exploration – exploitation tradeoff encountered testing multiple hypothesis sequentially. We then examine the size of this effect using both synthetic data and a simulated data obtained from an article studying pollution in different street layouts.

The structure of the thesis is as follows. In Chapter 2, we begin this thesis by going through related work and finally introducing two common online hypothesis procedures called alpha spending and alpha investing. In Chapter 3 we delve deeper into online hypothesis testing and examine its connection with reinforcement learning. We also discuss the exploration – exploration tradeoff that burdens the online hypothesis testing when the order of testing can be chosen during the procedure and present the proposed algorithm to solve this: Thompson sampling. We continue in Chapter 4 by providing empirical evidence that methods mentioned in Chapter 2 can gain statistical power by choosing the order of hypothesis using Thompson sampling. This is done by running three experiments: with synthetic and simulated data. We then examine further questions brought by the thesis and the experiments in Chapter 5. Finally we conclude the thesis in Chapter 6.

2. Background

In this chapter we concentrate on defining the key concepts of probability theory, decision theory and statistical hypothesis testing that are behind online hypothesis testing. Finally we present two important online hypothesis testing procedures: alpha spending and alpha investing.

2.1 Probability Theory

The root of statistical decision theory and thus of hypothesis testing lies in probability theory [4]. Therefore it is worth the time to recap its most integral concepts. Probability theory itself is a means to model randomness behind real-life events.

In order to treat probability formally we begin by defining the fundamental concept of a *probability space*. A probability space $(\Omega, \mathcal{F}, P)$ is a triplet consisting of a *sample space* Ω which refers to all possible outcomes of an experiment, a *σ-algebra* $\mathcal{F}$ which refers to a family of the sample spaces subsets where probability is defined and a *probability measure* P which satisfies the axioms of probability (i.e. P is a measure such that $P(\Omega) = 1$).

An important concept in probability is the one of a random variable. Given a probability space, a random variable is simply a function X from the sample space Ω into another space, often the real number space $\mathbb{R}$. In this thesis, we will only consider real valued random variables. In practice, they are often used to represent the observed sample of an experiments [5]. If X is a random variable and $A \in \mathbb{R}$, the measure $P(X^{-1}(A))$ is called the distribution of X [5].

The expectation of a random variable X is the weighted average of the values taken by that random variable [5]. It is especially useful in summarizing a distribution into a single number. Formally the expected value of a random variable X is defined as the Lebesgue integral with respect to the probability measure, i.e.,

$$E(X) = \int_{\mathbb{R}} X dP.$$

When multiple experiments are conducted, a key concept is independence. We say that two events A and B are independent when

$$P(A \cap B) = P(A)P(B).$$

This assumption is often made to greatly simplify statistical models. In practice this assumption is almost always wrong, but given that it is often not in the interest to test hypotheses that are known to be greatly dependent, it can often be approximately true and thus a justifiable assumption to make.

When multiple random variables are not independent usually information on one provides information about the rest. Conditional probability is the right tool for representing this phenomenon. Conditional probability is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

where A and B are events. The conditional probability can also be taken with respect to a random variable where $P(A \mid X)$ is the conditional probability of an event A given a random variable X and it is itself a random variable dependent on the events of X.

An easy way of dealing with conditional probabilities and a way of making inferences of unobserved random variables given some observations is obtained through the Bayes theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

which follows from the definition of conditional probability. One way of thinking of Bayes theorem is that starting with some *prior* distribution $P(A)$ which corresponds to our initial beliefs of the random variable A, we update the distribution based on the

observed data B. The distribution $P(A|B)$ is called the *posterior* distribution and it expresses our beliefs of A after observing the random variable B [6].

2.2 Decision theory

Another key sub field of science concerning hypothesis testing is the one concerning optimal decision making. We need three elements to define a statistical decision problem: a parameter space Θ, a decision space D and a real valued loss function L defined on $\Theta \times D$. The parameter space Θ represents all possible states of nature for some parameter of interest. The decision space D reflects the actions available for the decision maker. The loss function can be thought of as a penalty that a certain action gives in a given state of nature. The triplet (Θ, D, L) is called a *game* [7].

For any real-world problem, decisions are made based on data. We thus allow for observation of a random variable X whose distribution depends on some state of nature. A *statistical decision problem* is defined as a game coupled with an experiment with a random variable X whose distribution is P_θ where θ is some state of nature [7]. Our goal is, based on that random variable, to make an inference on θ such that the loss function is minimized.

In order to do this we define a decision rule δ which is a function from the sample space Ω to the decision space D. In hypothesis testing, where by definition the decision space is binary i.e. $D = \{d_0, d_1\}$ where d_0 signifies the acceptance of a hypothesis and d_1 the rejection, a decision rule can be thought of as a subset of the real valued space $S_1 \subset \mathbb{R}$, with the understanding that decision d_1 is taken if the random variable of interest X falls in S_1 and d_0 otherwise [7].

A common loss function is the 1-0 loss. It has the value 1 when a correct rejection is done and 0 otherwise. This is what will be minimized in statistical hypothesis testing [4]. The loss function, however, depends heavily on the unknown random variable X. Therefore we need to define a more general measure of optimality. We do this by defining the risk $R(\theta, \delta)$ corresponding to the loss function L as the expected

value over the sample space of L given that θ is the true state of nature. Formally stated:

$$R(\theta, \delta) = E_{p(x|\theta)}(L(\theta, \delta(X))).$$

This performance measure no longer depends on the random variable X but only on the decision rule and the true state of nature.

2.3 Statistical Hypothesis Testing

Statistical hypothesis testing is a form of statistical inference where the truth of a given hypothesis is evaluated based on some data. In the words of statistical decision theory it is a statistical decision problem with a binary decision space ("accept" and "reject") [7]. Statistical hypothesis testing has since its creation gained a foothold as a fundamental part of modern experimental science. It is widely used in areas of ecology, economics, biology, and medical sciences to name a few.

More formally in statistical hypothesis testing we want to make an educated decision of a binary hypothesis concerning a parameter $\theta \in \Theta$ based on a sample X whose distribution is P_θ. The process begins by partitioning the parameter space into two exclusive sets $\Theta_0 \cup \Theta_1 = \Theta$. Commonly we refer to the statement $H_0 : \theta \in \Theta_0$ as the null hypothesis. The opposing statement $H_1 : \theta \in \Theta_1$ is called the alternative hypothesis. The experimenters' problem is to choose the correct hypothesis. Often the null hypothesis is treated very differently from the alternative hypothesis. The null hypothesis is not proved but accepted without evidence. When enough evidence is gathered against it, it can be rejected in favor of the alternative hypothesis. This favoritism of the null hypothesis makes us consider the two kind of errors that can be made when testing a hypothesis differently. The experimenter can either reject a true null hypothesis or fail to reject a false null hypothesis. These are called type I and type II errors, respectively.

As mentioned in Section 2.2, the decision rule can be seen as a subset of the sample space. *Critical region* is defined as the subset $S_1 \subset \mathbb{R}$ of real line where the null

hypothesis is rejected when $X \in S_1$, where X is the the random variable of interest. The complementary region is called the *region of acceptance* $S_0 = S_1^{\complement}$. In order to limit the type I errors the probability of a null hypothesis being rejected given that it is true $P(X \in S_1 \mid \theta \in \Theta_0)$ is bounded from above by some level α. This constant α is called the significance of the test. The significance is arbitrarily chosen since there is no hard limit on the tolerated probability of type I errors [4]. In practice it is by convention often set to 0.05 [8]. In addition to limiting the type I errors we want to minimize the type II errors, or in other words, to maximize power which is defined as $P(X \in S_1 \mid \theta \in \Theta_1)$.

Often instead of working with the sample itself, one calculates a summary of it that discriminates between the null hypothesis and alternative hypothesis. Such transformations of the sample space are called *test statistics*. Otherwise the process is the same.

In single hypothesis testing, experimenters have moved to giving p-values of a test instead of simply informing whether it was accepted or rejected [9]. P-values are defined as the smallest significance at which the null hypothesis would be rejected if the null hypothesis is true. More formally

$$p = \inf(\alpha : X \in S_\alpha)$$

where S_α is a critical region of significance α. P-values are a measure of how extreme the experimenter regards this sample to be, if the null hypothesis is true. P-values not only give us more information of the test but also allow each experimenter to use their individual significance in rejecting the hypothesis [9]. Note that technically p-values need not exist for a test but we only focus on situations where they are readily available.

An important property of p-values is that they are distributed super-uniformly under the null hypothesis. That is

$$P_\theta(p \leq u) \leq u$$

for any $u \in (0, 1)$ and all $\theta \in \Theta_0$ as long as $\sup_{\theta \in \Theta_0} P_\theta(X \in S_\alpha) \leq \alpha$ [4]. This property is assumed for all p-values in this thesis. This means that if a hypothesis is rejected whenever $p \leq u$ the maximum probability of a rejection under the null hypothesis is u. This property is useful to keep in mind as it ensures that if a hypothesis is rejected when the p-value is smaller than a predefined level α, the probability of a falsely rejecting the null hypothesis is bounded from above by α.

2.4 Multiple Hypothesis Testing

In multiple hypothesis testing multiple tests are performed simultaneously. The central problem in multiple hypothesis testing is that using the same significance level for each hypothesis as in single hypothesis testing will result in more of type I errors than is acceptable.

Using the notation by Benjamini and Hochberg [10], we have a family of m hypotheses $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ of which m_0 are true. These hypotheses are then tested at individual levels α_i. In order to control for the number of false hypotheses, the level α_i at which each hypothesis H_i is tested at is often considerably lower than the significance allocated for the complete procedure α.

The summary of the testing situation can be seen in Table 2.1. The term **R** which stands for rejections is the number of rejected hypotheses total. The terms **U**, **V**, **T** and **S** stand for the number of true negative, false positive, false negative, and true positive decisions. Here **R** is an observable random variable compared to **U**, **V**, **T** and **S** which are unobservable random variables.

Table 2.1: Number of errors committed when testing m null hypotheses taken from [10].

	Non-significant	Significant	Total
True null hypothesis	**U**	**V**	m_0
Non-true null hypothesis	**T**	**S**	$m - m_0$
	$m - \mathbf{R}$	**R**	m

As incorrect conclusions might be drawn for every type I error, we wish to minimize the term $\mathbf{V}$ while simultaneously maximizing the number of rejected false hypotheses $\mathbf{S}$. As is with testing single hypothesis, some bound is set for the type I errors while the number of rejections of false hypothesis is maximized. A common way of achieving this for multiple hypotheses is to bound the probability of having a single type I error. This is exactly what the family-wise error rate (FWER) measures [11]. More formally, using the notation of Table 2.1:

$$\text{FWER} = P(\mathbf{V} \geq 1).$$

Control of FWER is important when a single false rejection leads to wrong conclusions [10]. Such is often the case in clinical trials when determining early stopping of a study. In most cases this is too conservative resulting in little statistical power, especially when the number of hypotheses is high.

Another common type of error is the false discovery rate (FDR) [10]. FDR is defined as the expected ratio of false rejections to all rejections that is

$$FDR = E(\frac{\mathbf{V}}{\mathbf{R} \vee 1}).$$

The maximum is taken in order to deal with cases that have zero rejections. This measure is more suitable for cases where false rejections do not affect the conclusions of other hypotheses. Further error rates exist but they have received less attention in the literature.

Both of these error rates have a myriad of variants that have been invented because of specific needs or ease of calculation. Most notably k-FWER is a less restrictive version of FWER that limits the probability of making k type I errors. One notable variant of FDR is the marginal FDR (mFDR). We use a definition from [5] and define

$$mFDR_\eta = \frac{E(\mathbf{V})}{E(\mathbf{R}) + \eta}$$

where $\eta > 0$ is some constant typically chosen as $\eta = 1$ or $\eta = 1 - \alpha$ [2].

We speak of *weak control* when this error rate is controlled under the complete null hypothesis i.e. when all null hypotheses are true. Weak control however in most

cases is inadequate. Alternatively, *strong control* of an error rate means that it is controlled under all combinations of true and false null hypotheses. Note that any procedure that controls FDR controls FWER weakly [10]. In other words, if all null hypotheses are false, the probability of a procedure controlling FDR declaring a type I error is bounded by its significance. Also any procedure that controls $\text{mFDR}_{1-\alpha}$ at level α weakly controls FWER at level α [3].

2.4.1 Bonferroni Procedure

The simplest form of multiple hypothesis correction and one that we are going to present as an example is the Bonferroni correction. The Bonferroni procedure is not an example of online hypothesis testing but of *batch* hypothesis testing which means that all of the hypotheses are tested simultaneously. It is presented since it is widely used and it provides good intuition for the following methods. Bonferroni offers strong control of FWER for a family of m hypotheses at any significance level α. It does this by testing each hypothesis at with the significance of $\frac{\alpha}{m}$. Using the union bound (also known as first-order Bonferroni Inequality) it is straightforward to prove that this method controls FWER:

$$P(\mathbf{V} \geq 1) = P(\bigcup_{i=1}^{m_0} p_i \leq \frac{\alpha}{m}) \leq \sum_{i=1}^{m_0} P(p_i \leq \frac{\alpha}{m}) \leq \sum_{i=1}^{m_0} \frac{\alpha}{m} \leq \alpha \tag{2.1}$$

Although the Bonferroni procedure is simple, more statistical power can be obtained using one of many sequential procedures such as the Holm-Bonferroni method presented by Holm [12].

Notice that in the proposed proof (Equation 2.1), the hypotheses need not be tested at the same level of significance. Testing each hypothesis at level $\alpha\omega_i$ when $\sum_{i=1}^{m} \omega_i = 1$ is called the *weighted Bonferroni procedure*.

2.5 Online Hypothesis Testing

Online hypothesis testing as a form of multiple hypothesis testing differs from the traditional batch testing where the hypotheses are obtained simultaneously by having the hypotheses received one at the time. After a hypothesis is received it must either be accepted or rejected immediately before receiving the next hypothesis. The number of hypotheses can be undetermined before the procedure and may even be infinite. Another motivation for testing each hypothesis one-by-one is that there is a lot of time between obtaining the hypotheses or that the testing of the hypotheses has a high cost.

Online methods can be used to solve traditional batch testing problems. With appropriate prior information on the probability of rejections and carefully designed testing process this can result in more power than using a traditional batch testing method such as the Holm-Bonferroni procedure [13]. Most often, however, the number of conducted tests is not known in advance or the hypotheses are obtained one at the time and thus the use of online hypothesis testing is required.

We investigate a case when the choice of the next hypothesis itself may be dependent on the current and prior rejections. This is the case in interactive data exploration since the experimenter chooses the next hypotheses based on what they have learned from the prior hypotheses. Other real-life use-cases of online hypothesis testing include in A/B testing conducted by internet companies, early stopping of clinical trials and quality-preserving databases for multiple researchers to test multiple hypotheses on the same data [13].

Many online hypothesis procedures exist but we are going to go through the most important to interactive data exploration where the number of hypotheses typically is not known in advance.

2.5.1 Alpha Spending

Alpha Spending can be viewed as the online generalization of the Bonferroni correction [14]. Although it is equivalent to weighted Bonferroni when the number of hypotheses is known it is convenient to view this procedure in terms of "spending" the available significance. Alpha spending begins by choosing an initial amount of *alpha wealth* $W(0) = \alpha$ where α is the significance of the procedure. Instead of spending all of the wealth equally (as in Bonferroni procedure), each individual hypothesis H_i is tested at a level α_i which is chosen before observing the hypothesis such that $0 \leq \alpha_i \leq W(i-1)$. The wealth is correspondingly updated as

$$W(i) = W(i-1) - \alpha_i.$$

The term α_i corresponds to the amount of wealth used for testing the ith hypothesis and $W(i)$ corresponds to the amount of wealth left after testing that hypothesis.

As long as the wealth remains non-negative (that is $\alpha_i \leq W(i-1)$), FWER is controlled at level α. The proof as presented in Equation 2.2, resembles much to the one presented in Equation 2.1.

$$P(\mathbf{V} \geq 1) = P(\bigcup_{i=1}^{m_0} p_i \leq \alpha_i) \leq \sum_{i=1}^{m_0} P(p_i \leq \alpha_i) \leq \sum_{i=1}^{m_0} \alpha_i \leq W(0) = \alpha \tag{2.2}$$

The last inequality holds since if the wealth is not allowed to be negative as $\sum_{i=1}^{m_0} \alpha_i \leq \sum_{i=1}^{m_0} W(i-1) - W(i) \leq W(0)$. This means that once the wealth is depleted, no more hypotheses can be tested.

The choice of significance allocated for each hypothesis is a non-trivial question. In clinical trials, where alpha spending is often used, an *alpha spending function* $a(t)$, where t signifies the fraction of information available, is often chosen prior to the experiment. This method introduced by DeMets and Lan [1] allocates a significance of $\alpha_i = a(t_i) - a(t_{i-1})$ for each hypothesis. In clinical trials this method is allows the number of interim analyses and their calendar times to be chosen during the experiment as opposed to before it [1]. This method controls FWER as long as $\alpha(t)$ is an increasing

function such that $a(0) = 0$ and $a(1) = \alpha$. Here the input of the alpha spending function t represents the fraction of information when a statistical test is conducted [1].

For clinical trials this method allows a great deal of flexibility while still offering a statistical guarantee against type I errors. Unfortunately, as is with Bonferroni correction, a drawback of this method is that it often is too conservative resulting in loss of statistical power [15].

Typically the amount of alpha spent on each hypothesis is chosen in advance with an alpha spending function. When this is not the case, one must ensure that the Equation 2.4 holds as is with the procedure presented next.

2.5.2 Alpha Investing

Another method that has been proposed for online hypothesis testing is alpha investing introduced by Foster and Stine [3]. Alpha investing is inspired by alpha spending but unlike alpha spending it controls mFDR_η instead of FWER. This allows for drastically more statistical power.

In alpha investing we start with an initial wealth of $W(0) = \alpha\eta$. When a hypothesis H_i is tested at level α_i the wealth is updated as follows:

$$W(i) = W(i-1) - (1 - R_i)\frac{\alpha_i}{1 - \alpha_i} + R_i\omega$$

where $R_i \in \{0, 1\}$, which stands for rejection, is the outcome of the test H_i i.e.

$$R_i = \begin{cases} 1, \text{ if } p_i \leq \alpha_i \\ \\ 0, \text{ otherwise} \end{cases}$$

and $\omega \leq \alpha$ is a reward gained for rejecting a hypothesis. This reward is customarily set as α as this maximizes the power of the procedure. The name investing comes from the fact that wealth can be gained if a hypothesis is rejected.

The function of the previous rejection

$$\alpha_i = \mathcal{I}_{W(0)}(\{R_1, R_2, \cdots, R_{i-1}\})$$

which determines the level of significance α_i used for hypothesis H_i is called an *alpha investing rule*. The alpha investing rule that was originally presented by Foster and Stine [3] is as follows:

$$\mathcal{I}_{W(0)}(\{R_1, R_2, \cdots, R_{i-1}\}) = \frac{W(i-1)}{1+i-k^*} \tag{2.3}$$

where k^* is the index of the hypothesis that was last rejected. This rule works especially well if the false hypotheses arrive in batches.

This procedure controls mFDR_η at level α if for all tests H_i,

$$P_\theta(R_i = 1 \mid H_{i-1}, H_{i-2}, \cdots, H_1) \leq \alpha_i \tag{2.4}$$

holds for all $\theta \in \Theta_0$. This condition is weaker than independence of the hypotheses although assuming the independence of the hypotheses is a practical way to satisfy this condition [3].

Due to controlling different error rates, alpha investing is considerably more powerful than alpha spending, especially when the proportion of false null hypotheses to true null hypotheses is high. On the other hand the assumption given by Equation 2.4 might be too restricting to be used for all problems and is might lead to more type I errors.

Alpha investing is a special case of *generalized alpha investing* (GAI) [15]. Generalized alpha investing allows for greater freedom for the experimenter in choosing the amount of reward gained from rejecting a hypothesis. More formally under GAI the initial wealth is $W(0) = \alpha\eta$ and after each hypothesis H_i is tested at level α_i the wealth is updated to

$$W(i) = W(i+1) - \varphi_i + R_i\psi_i.$$

Here φ_i refers to the amount of wealth that is invested each test and ψ_i is the reward that is gained on each rejection.

In order to control mFDR_η the generalized alpha investing rules

$$(\alpha_i, \varphi_i, \psi_i) = \mathcal{I}_{W(0)}(\{R_1, R_2, \cdots, R_{i-1}\})$$

must further satisfy the following three inequalities: $\psi_i \leq \varphi_i + \alpha$, $\psi_i \leq \frac{\varphi_i}{\alpha_i} + \alpha - 1$ and $\varphi_i \leq W(i-1)$ [16].

The GAI procedure can also be used to control FDR at level $\alpha(1+\eta)$ when the p-values are independent [16]. Adjusting the method to control FDR at lower levels is possible, but this comes with a substantial loss of power.

2.6 Reinforcement Learning

As this thesis explored similarities between online hypothesis testing and reinforcement learning, a section is dedicated to a brief summary of reinforcement learning. The connection is further and the exploration – exploitation problem are further examined in the following chapter.

Reinforcement learning is one important paradigm of machine learning. Its goal is to learn the mapping from situations to actions that maximizes a reward signal without a direct input on the optimal action but it must learn the reward maximizing action by trial-and-error [17]. It simultaneously refers to the computational problem, solutions to that problem, and the field that studies that problem [17].

Reinforcement learning has taken large inspiration from biological systems [17]. In fact out of all forms of machine learning, reinforcement is the closest to the way that humans learn [17].

What distinguishes reinforcement learning from the other main paradigms of machine learning is that in it, an agent learns directly from interacting with an environment. An agent refers to the decision maker trying to learn the mapping from situations to actions while interacting with an environment. When interacting with the environment, the agent gains access to reward signals which guide the agents future actions. The goal of the agent in reinforcement learning is to maximize the cumulative reward.

Formally, in addition to the agent and the environment, the main subelements of a reinforcement problem are a *policy*, a *reward signal*, a *value function*, and, optionally,

a *model of the environment* [17]. The policy refers to the way of the agent to act in a specific situations. It can be thought as a mapping from state to actions. Choosing the correct policy may depending on the state may allow the agent to obtain a reward signal.

The reward signal is the value that ought to be maximized. The reward signal provides a way to inform the agent what should be achieved [17]. For example, the reward of an agent learning a game (such as chess or go) could be 1 when a game is won and 0 otherwise (in chess, a reward of $\frac{1}{2}$ could be awarded in case of a tie to prevent the agent from learning to take desperate actions in drawn positions). It is vital to define the reward to match the underlying goal of the agent.

Value is the future expected reward. The value function indicates the long term reward while the reward signal is immediately acquired form the environment. The sole purpose of modeling the value function is to achieve better long term reward [17].

The fourth and optional element of reinforcement learning is model of the environment. The model the agent allows to make predictions of the reward without directly interacting with the environment. This is not strictly necessary as the agent may directly interact with the environment but it may help it to generalize faster.

One defining challenge in reinforcement learning problems is the *exploration – exploitation trade-off* [17]. Making use of the most promising policies (*exploiting* the current knowledge) may lead to missing even higher reward policies. On the other hand, using all the available time for exploration leads to the agent to try suboptimal policies in order to learn more about their true value function in order to find higher reward policies (*exploring*). This challenge, although heavily researched for decades, is not fully solved [17].

Given that this problem has an exploration – exploitation trade-off, it is very natural to look towards reinforcement learning which is partly characterized by this problem.

The three elements of reinforcement learning can also be found from online hy-

pothesis testing. Using the terminology of reinforcement learning, an agent (experimenter) must choose a policy (the amount of wealth used to test the next hypothesis) while receiving a reward signal (1 if the hypothesis is rejected, 0 otherwise) while optimizing the value function which is the expected number of rejections of the procedure. False rejections need not to be worried about since the using online hypothesis procedures such as alpha investing guarantees that the ratio of false rejections is bounded by an acceptable level. This means that we can state the online hypothesis problem as a reinforcement learning problem and use already established methods to solve it. The fourth element, model of the environment, can optionally be defined in order for the algorithm to generalize faster.

3. Methods

In this chapter we define our research problem formally as a computational problem. We investigate the relationship and applicability of reinforcement learning methods for online hypothesis testing through the exploration – exploitation tradeoff. We then present different natural correlation structures which (if true) can be harnessed to gain power in online hypothesis testing. Finally we shortly discuss another significant way of gaining power in already established online hypothesis testing procedures: choosing the optimal investing rule.

3.1 Stating The Problem

When it comes to online hypothesis testing there are two ways of improving the power of an existing procedure. The first is in improving the order of the hypotheses. The ordering is important as many procedures get more powerful after rejecting hypotheses. If the hypotheses that are likely to be rejected are tested first, more power will be gained for later experimenting. Also, many investing rules test earlier hypotheses at a higher level of significance due to the uncertainty towards the number of hypotheses. The second one is in finding a more efficient way of distributing the wealth. We focus on improving the order of the hypotheses based on the information that is acquired during the hypothesis testing process.

We define the goal explicitly as a computational formal problem as follows:

Problem 1. *Given a set of hypotheses* $\mathcal{H} = \{H_1, H_2, \dots\}$ *and the corresponding test statistics* $T = \{T_1, T_2 \cdots\}$*, which order of testing the hypothesis will maximize the*

expected power of a given online testing procedure?

Such problems arise often in interactive data exploration when the experimenter needs to decide the next hypothesis to be tested based on the already tested hypotheses while still controlling the false discovery rate. Another case when hypotheses are tested sequentially comes when there is a high cost associated with each test. In this case the experimenter wishes to obtain a rejection as soon as possible. More use cases are explored in Chapter 4.

The hypotheses should generally be tested in an descending order based on their likelihood. In their article, Foster and Stine [3] coin this the "Best-foot-forward policy". This is mainly because many methods (such as alpha-investing) gain more power after they reject hypotheses, but also because many investing rules (at least when the number of hypotheses is unknown) test each hypothesis with a decreasing amount of significance, resulting in more power given to the early hypotheses. This is done in order not to deplete of the alpha wealth and to ensure that most of it is used if the number of hypotheses ends up being small.

3.2 Exploration – Exploitation Trade-off

The optimal order of the hypotheses depends on the information we gain during the process of hypothesis testing. The information comes through modeling the joint distribution of the test statistics. This way testing a single hypothesis (and thus observing its test statistic) allows us to calculate the posterior distribution of the test statistics using the Bayes rule. Based on this posterior distribution we can then choose the next hypothesis to test (and thus which test statistic to observe next).

Since testing new hypotheses reveals new information on the rest of the hypotheses there is an exploration – exploitation trade-off inherently built in to the problem. It would be natural to choose the hypothesis that is most likely to be rejected. However this may leave some regions of the joint distribution completely unexplored resulting in a lack of power. A balance must be struck between exploiting the current knowledge

of the joint distribution of the test statistics and learning more about it.

One tool in reinforcement learning where the same exploration – exploitation problem is especially visible is in *multi-armed bandits*. The basic version of a multi-armed bandit is an algorithm that has K possible actions to choose from (which are often referred as arms) [18] and T rounds. During each round the algorithm must choose an arm. From each action it gains a random (but with a fixed distribution) reward specific to that arm [18]. The name multi-armed bandit is inspired by a scenario where a gambler must choose from several slot machines that yield different amount of payoff [18].

Online hypothesis testing bears many similarities to multi-armed bandits when each hypothesis is seen as an arm. If the hypothesis is rejected a reward is obtained. The main difference between our problem and the multi-armed bandits is that each hypothesis (arm) is tested (pulled) only once. Since each arm is only pulled once, some assumptions must be made on how the information is "leaked" to the other arms. These correlation structures are discussed in Section 3.4.

Seeing the connection with the multi-armed bandit problem gives us justification to use already existing algorithms in order to obtain an approximate solution to the computational problem above.

Many solutions to multi-armed bandits have been proposed. The naive solution is to sample each arm uniformly (uniform exploration strategy). A slightly more enticing but still naive solution is to draw the most promising arm with a probability of $1-\epsilon$ and choose the arm uniformly with a probability of ϵ (ϵ-greedy strategy). These, however, is not suitable for determining the order of the hypotheses since each arm can only be sampled once. More refined solutions are Thompson sampling and Upper confidence bound (UCB) -algorithms. We suggest Thompson sampling as it is easy to adapt for hypothesis testing in general situations as shown below.

3.3 Thompson Sampling

Thompson sampling starts by specifying a prior for all arms being the best arm. An arm is tested with the probability that it is the best arm. When an arm is tested the posterior probability of each arm being the best arm is computed and the process repeats until a potential time limit is reached.

As an input it requires the set of all hypotheses of interest $\mathcal{H}$ and a prior joint reward distribution P_0. The reward can for example be chosen as the probability of rejecting a hypothesis or as the absolute value of the test statistic. When using alpha investing, we have to define the alpha-investing function $\mathcal{I}_{W(0)}$ that defines the amount of significance assigned for each hypotheses. This input is often not for Thompson sampling necessary when applied to other use-cases than online hypothesis testing.

Thompson Sampling $(\mathcal{H}, P_0, \mathcal{I}_{W(0)})$

Choose a hypothesis $H_1 \in \mathcal{H}$ uniformly to be tested first.

for *each hypothesis* $H_i, i = 1, 2, \cdots$ **do**

- Test hypothesis H_i at level $\alpha_i = \mathcal{I}_{W(0)}(T_i, \cdots, T_1)$.
- Update the posterior distribution $P_i = P_{i-1}(\cdot \mid H_i)$.
- Sample the reward μ_t from the posterior distribution P_i.
- Choose the next hypothesis to correspond to the highest μ_i.

end

Algorithm 1: The pseudocode for Thompson sampling

The prior distribution and likelihood can be chosen freely but especially fast algorithms exist for choosing the arm if it follows the the beta-binomial or Gaussian distribution [18].

Thompson sampling deals with the exploration – exploitation tradeoff by concentrating on the more promising hypotheses. It does however have a smaller positive probability to choose each hypothesis providing a chance for exploring even the more improbable hypotheses.

Thompson sampling also has theoretical properties that make sure that the result

is adequate. The important results come from reinforcement learning literature where it is shown that the procedures *regret* can be bounded. Regret at time T is defined as the reward

$$R(T) = \mu^* \cdot T - \sum_{t=1}^{t=T} \mu(a_t)$$

where $\mu(a_t)$ is the reward from the action chosen at time t and μ^* is the expected reward from the best arm [18]. Regret is linearly dependent on the reward (which in turn is determined by the loss) meaning that bounding expected regret allows us to give bounds to expected rewards if the expected reward from the best arm is known.

It can be shown that Thompson sampling with 0-1 rewards and an independent uniform priors achieves an expected regret

$$E(R(T)) \leq O(KT \log T)$$

where K is the number of arms and T is the number of time steps [18]. The same bound is also achieved with independent Gaussian priors and unit-variance Gaussian rewards [18]. Unfortunately this bound is not very useful if one views each arm as a hypothesis, since K would be very large.

3.4 Structure of the Hypotheses

If we want to learn which hypotheses are the most promising, we need to make assumptions on their structure. This is demonstrated with the fact that no learning can happen when the hypotheses are independent. In this section, we present different types of assumptions that are useful in real life cases. To be precise, we inspect three cases: independent hypotheses, the test statistics follow a topic model and the test statistics are a Gaussian process.

3.4.1 Independent Hypotheses

Although independence is a very strong assumption, it is often made to simplify the decision process. Unfortunately, in this case, the information of testing other hypothe-

ses cannot be used to model the future hypotheses. This is exemplified by the following equation:

$$P(T_m|T_1, \cdots, T_{m-1}) = P(T_m)$$

which hold for each m.

Thompson sampling would therefore not update the posterior distribution but only sample the order of hypotheses based on the prior. Since no learning is happening, using Thompson sampling is not suggested. This demonstrates that some assumptions on the structure of the hypotheses are indeed necessary in order to learn a more optimal ordering of the hypotheses during the testing process.

For this reason, in this special case, the importance of the amount alpha wealth spent for each hypothesis becomes much more important.

3.4.2 Topic model

Another structure of the hypothesis that we are going to inspect is the *topic model*. It assumes that the hypotheses come from K independent topics. All the hypotheses are then conditionally independent given the topic.

This setting mimics the setting of a traditional multi-arm bandit. Each topic (arm) has a sequence of independent hypotheses with a differing proportion of null hypotheses to alternative hypotheses (which results to a differing payoff). The difference between this structure and traditional multi armed bandit problem is that an arm can deplete, meaning that all of the hypotheses of a certain topic can run out.

Formally this model can be described followingly:

$$\begin{aligned} P(T_m \mid T_1, \cdots, T_k) &= \sum_i P(T_m \mid T_1, \cdots, T_k, z_i) P(z_i \mid T_1, \cdots, T_k) \\ &= \sum_i P(T_m \mid z_i) P(z_i \mid T_1, \cdots, T_k) \\ &\propto \sum_i P(T_m \mid z_i) \prod_{l=1}^{k} P(T_l \mid z_i) P(z_i) \end{aligned} \tag{3.1}$$

where $P(T_m \mid z)$ is the distribution of the test statistic given the topic z.

In practice such structures are often assumed for text analysis where the probability of each word is dependent on the topic of the text. In physical sciences a confounding factor such as the season or the time of day (day or night) can cause observations to follow such pattern.

3.4.3 Gaussian Processes

In nature it is common for all hypotheses to be correlated to each other at various levels. We model this situation by assuming that the joint distribution of the test statistics corresponding to the hypotheses is normally distributed with a known mean and covariance matrix. In other words we use Gaussian processes to model the test statistics.

This type of correlation structure occurs in nature often due to spatial or temporal location. For example, Nearby pollution detectors are most likely have similar levels of pollution while more distant detectors are less correlated. For this reason predictions by Gaussian processes have been widely used in sciences such as meteorology and geostatistics where the method is known as *kriging* [19].

The idea behind Gaussian processes is to define a distribution over functions [19]. This distribution is then conditioned on the training set points [19]. The predictions can then be sampled from the resulting posterior distribution.

Since usually only a modest number of hypotheses are tested at the time, we employ a form of kriging which is called simple kriging in geostatistics. This assumes both the mean and covariance function to be known. Specifically, the covariance matrix is generated by a covariance function K which specifies the correlation between two points. The mean is conventionally set to 0 as this has a lesser impact on the resulting model. Under these assumption, conditioning the distribution on the observed results, one can obtain the following posterior distribution for the unobserved test statistics:

$$f^* \mid X_*, X, f \sim \mathcal{N}(K(X_*, X)K(X, X)^{-1}f, \\ K(X_*, X_* - K(X_*, X)K(X, X)^{-1}K(X, X_*))), \tag{3.2}$$

where f^* and f are vectors containing the unknown and known samples respectively, X_* and X are their corresponding features, and K is the covariance function [19].

The covariance function quickly presented above is an approach to encode the assumptions on the correlations of the data points based on some features [19]. Simply put, it defines the similarity of the data points.

A commonly used covariance function is the squared exponential function. The squared exponential covariance function is defined followingly:

$$K_{se}(r) = \exp(\frac{r^2}{2l^2}),$$

where l is called the *characteristic length-scale* and $r = |x_1 - x_2|$ is the distance between the two data points [19]. The characteristic length-scale corrects for the scale of the points.

In practice this means that points which are nearby in terms of their covariates are highly correlated while far-away points have very low correlations. The squared exponential function is infinitely differentiable resulting in it appearing very smooth [19]. In theory this level of smoothing is most often unrealistic, but it is nevertheless very popular [19].

3.5 Optimal Investing Rule

The other consideration that must be made is the amount of wealth that is spent for each hypothesis. The basic idea is that spending too much significance for each hypothesis results in the significance depleting before all the hypotheses are tested. On the other hand, spending too little results in loss of power as the remaining significance ends up being wasted.

The situation when alpha wealth ends prematurely ending the exploration process has been referred to as *alpha-death* [13]. This alpha death can be avoided by employing *thrifty* investing rules, meaning rules that never use all of their alpha wealth. This is recommended if the number of hypotheses is unknown. The problem is not, however,

completely solved by employing thrifty investing rules as the amount of significance left may not be able to be enough to reject any of the remaining hypotheses. Although important, when compared to the ordering of the hypotheses the investing rule is less important with alpha investing [3].

Multiple different investing rules have been proposed in the literature on top of the one presented in Equation 2.3. A comparison of multiple different investing rules can be found by Zhao et al. [2] with some more presented in the original article by Foster and Stine [3].

The optimality of investing rules naturally depends on the assumptions that the experimenter is willing to make. For example, the investing rule presented by Foster and Stine [3] and in 2.3 works best when the hypotheses are clustered. For a general purpose a simple investing rule is to spend a fixed proportion $1 - \beta$ of significance for each hypothesis while saving a proportion of β of the significance for future tests. This is known as the β-farsighted rule and it has been found to be the best policy if the number of hypotheses is unbounded [2]. On top of its good performance it is easy to implement and, due to its simplicity, to justify. The β-farsighted rule test each hypothesis at significance level of

$$\frac{W(i-1)(1-\beta)}{1+W(i-1)(1-\beta)}.$$

The parameter beta controls how long the testing procedure lasts. When there are expected to be only few hypothesis a large value for beta should be chosen. Conversely if there are a large number of hypotheses a large value for beta will make sure that enough wealth is saved for the testing of the later hypothesis

4. Experiments

In this chapter we experiment on how much statistical power can be gained by exploiting the knowledge of the structure of the data when comparing it to testing the hypothesis in a random order. Firstly, we demonstrate that this is the case in the simplest non-trivial case with synthetic data following a mixture model. The more realistic synthetic data is created to better understand the methods and to find its power in an ideal situation. Finally the methods are tested in a more practical setting to investigate their usability in practice.

We begin this chapter by presenting the data. After this we go through each experimental set-up and their goals. Finally we present and discuss the results of each experiment separately.

4.1 Materials

Three datasets are explored in this thesis, two synthetic data set created for the purpose of the experiments and a real-life dataset.

4.1.1 Synthetic Topic Model Data

In order to estimate the effectiveness of incorporating prior structural information in online hypothesis testing, we begin by creating data matching the topic model. This situation is created to resemble the multi armed bandit problem and it serves as a proof of concept. The synthetic model has K different topics. The number of topics K is varied during the experiment. For each topic, we create $n = 100$ independent p-values

following a mixed beta binomial distribution

$$G_z(x) \sim \pi_z U(0,1) + (1 - \pi_z)F_1$$

where π_z refers to the probability of a null hypothesis, U to the uniform distribution and F_1 to the distribution of p-values under the alternative hypothesis. The value of π_z is different under each K topics. To simulate the distribution F_1 of p-values when under the alternative hypothesis, we use a beta distribution. The parameters of the beta distribution ($\alpha = 0.064, \beta = 1.517$) are chosen from a prior study estimating a empirical distribution of p-values obtained from RNA-sequences [20], a common use-case for multiple hypothesis testing.

The distribution of p-values used for the experiment can be seen from Figure 4.1.

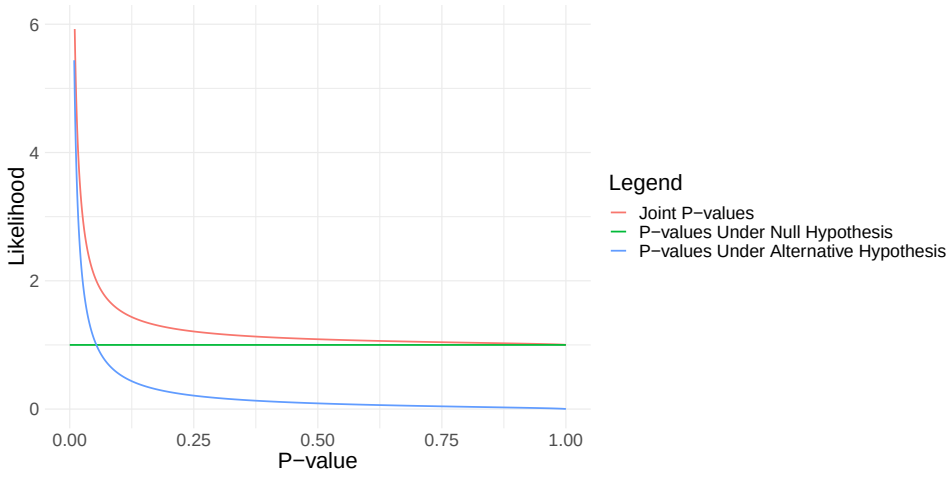

Figure 4.1: Raster of mean particles in the data set explored in experiment 3.

4.1.2 Synthetic Simplex Noise Data

A more practical situation in hypothesis testing is that all of the hypotheses have certain correlations with each other. For this purpose we create a data set inspired by the following the simulated data set so the ideal setting can be created and explored.

We use simplex noise to represent the ground truth test statistic values. Simplex noise creates a smooth looking data set where data points have similar values with nearby points. In order to add a greater element of randomness, some Gaussian noise

(having the mean of 0 and standard deviation of 1) is added on each data point. Example of the data can be seen from Figure 4.2. The strength of the simplex noise (referred as signal strength) is varied through the experiment. The signal strength of k means that the simplex noise is scaled between 0 and k.

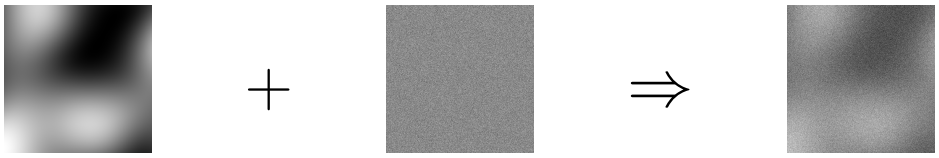

Figure 4.2: Example of the synthetic data. The simplex noise on the left is added with the Gaussian noise to create the synthetic data used on the right. Signal strength of 3 was used to create the figure.

Such raster of 500×500 is created and random points are chosen from this raster uniformly. The Euclidean distance between these random points is used to create a correlation matrix for Thompson sampling. The number of random points is chosen to be $n = 20$. The noisy values of the simplex noise serves as the test statistics. Notice that given how this data is created the test statistics are correlated with each other depending on the distance between them.

4.1.3 Simulated Data

For the practical experiment we are going to use data set derived from a simulated data set computed for the Kurppa's masters thesis [21]. The original article studies the effect of different city plans on pollution using a large-eddy simulation. The data involves four city plans (including the heights of buildings, the surrounding terrain and information on the tree canopy) for two different wind directions. From this dataset we choose one layout in which the amount of simulated particles in every $2m \times 2m \times 1m$ block for an area of $770m \times 634m \times 30m$ measured every $5s$ for an duration of an hour. The first $100s$ averaged serves as our dataset. For our purposes we only inspect the ground level (4 meters over the ground) as this was done in the original study. Figure 4.3 shows a visualization of the used dataset. The data set has large areas with

0 particles mainly on the top left and bottom right corners of the map. These points are not investigated since they are not of interest for a researcher. Buildings are also excluded since none of these points have any particles at the required height level.

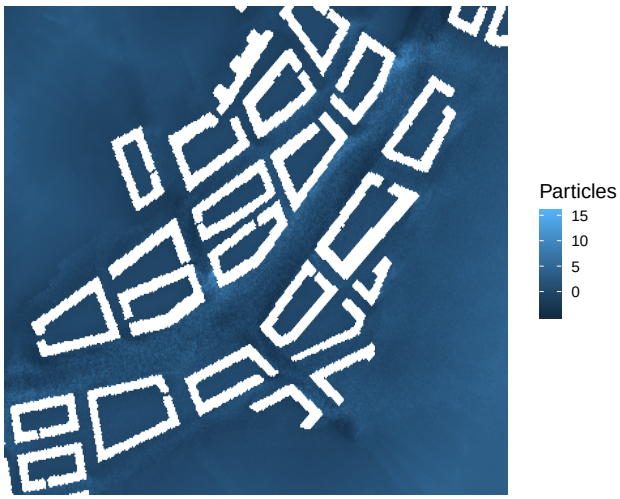

Figure 4.3: Raster of mean particles in the data set explored in experiment 3.

4.2 Experimental set-up

For our first proof-of-concept experiment we use the data generated in subsection 4.1.1. As we model the p-values themselves instead of the test statistics, we do not specify a statistical test as it works generally for any test. The hypotheses (to which the p-value refers) are rejected when the p-value is below the significance that it is tested at.

The order of the tested hypotheses is determined by the Thompson sampling as described in algorithm 1. The level of significance controlled by alpha investing is set to $\alpha = 0.1$. We use β-farsighted strategy where $\beta = 0.9$ as this parameter is suggested to work well by Zhao et al. [2]. This is contrasted with the same hypotheses being tested in a uniformly random order. The experiment is conducted $m = 10000$ times and the power is then averaged in order to get more accurate results.

The effect of having different number of topics is experimented by repeating the

experiment for each number of topics from $K = 1$ to $K = 100$. The class probabilities of each topic is sampled uniformly between 0 and 1. Averaging over multiple trials mitigate the randomness that this results in.

To recap, a total of $n = 100$ hypothesis are tested from K topics with different probabilities of rejection. Thompson sampling automatically learns a testing strategy that we hypothesize to be better than testing the same set hypotheses in a random order which it is contrasted with.

The second experiment is an idealized version of a real-life dataset. Here each hypothesis is properly treated as an arm in terms for Thompson sampling. The information of the test statistic then leaks into the other arms trough the updating of the posterior distribution. Each test statistic (corresponding to the hypotheses) are assumed to have a fixed correlation structure generated by the squared exponential kernel introduced in subsection 3.4.3. The characteristic length-scale parameter of the squared exponential kernel is varied during the experiment.

The null hypothesis is that each test statistic is fully comprised of random Gaussian noise allowing us to calculate the one sided p-values for the online hypothesis testing procedure using the following equations $p_i = 1 - \phi(T_i)$ where T_i is the test statistic and ϕ is the Gaussian cumulative distribution function. In other words, we are testing if the data is fully comprised of the random noise and we reject the hypotheses if it has values larger than would be expected by the null hypothesis.

As with the first experiment the parameters for alpha investing are fixed at $\alpha = 0.1$ and β-farsighted strategy is used with $\beta = 0.9$. The number of repetitions is again fixed at $m = 10000$ in order to get a more accurate result.

Two scenarios are examined with the second data set. First we vary the characteristic length-scale parameter of the correlation function. This is an important hyperparameter as it controls the amount that each hypothesis affects the other hypotheses posterior probabilities. To be precise the values of kernel length varies between 1 and 250 every 25 steps while signal strength is kept at 3. In the second scenario the best

kernel length is used to examine the effect of different levels of signal strength. In the second scenario the kernel length is set as 50 since this performed the best in the first scenario while the signal strength is varied between 1 and 10 every 1 step. Running the tests with this way allows us to identify the effects of both variables separately.

The third experiment is a practical one. A regression model is trained over the data set to predict the particle densities. We use a linear regression model trained on the whole data set. This is done in order to leverage our current level of knowledge. If the raw values were inspected instead of residuals we would not have a reason to expect the data to be a Gaussian process with squared exponential kernel since it has different areas which are known to have different level of pollutants, e.g. The courtyards have on average less particles than the smaller streets which has less particles than the central boulevard. The absolute difference between the true value and the prediction serves as the test statistic. Because of the massive size of the raster, $m = 100$ points are chosen randomly to be examined. In practice such situations are common since the value of particle density is often only available from sensors which are sparsely located. The ability test each statistic one by one is especially useful in cases where testing a hypothesis has high costs involved.

The hypothesis is that the residuals of this model follows the standard normal distribution. This means that the method of obtaining p-values is identical to the one used in experiment 2. Practically this method allows us to find areas where the model does not match the assumptions laid down. The standard deviation of this distribution can be arbitrarily chosen to match the requirements of the model. In our case the standard deviation is incorrect enough to give a good amount of rejections with significance of $\alpha = 0.1$ to compare the two methods. The amount of rejections strongly depends on the null hypothesis which must be chosen on a case by case basis.

4.3 Results

Three experiments are run as described in Section 4.2. The main results are visualized in the following figures.

The results of the first experiment can be seen in Figure 4.4. Increasing the number of arms creates the power to resemble an inverted U-shape curve. This is due to the way the data is created. The more topics there are, the "better" the best topic is. This is because the probability of a hypothesis not following the null distribution is sampled uniformly. When the number of topics grows too large, Thompson sampling does not have enough time to explore all its possibilities and therefore it resembles more and more of random sampling. As expected, random sampling performs approximately uniformly. The best performance is obtained when there were 7 different arms. After this the performance slowly decays. It should be noted that even when the number of arms equals the number of tested hypotheses, Thompson sampling performs considerably better than random sampling meaning that it does not get stuck in the exploration phase even in such an extreme case.

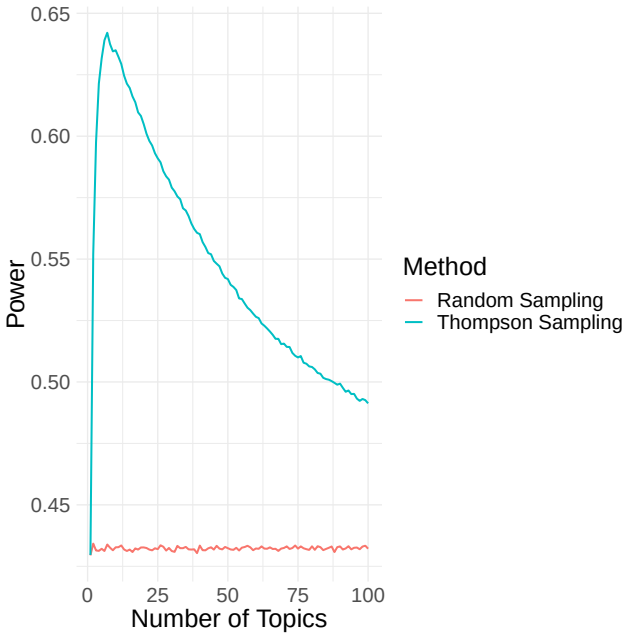

Figure 4.4: Comparison of the statistical power of alpha investing when the hypotheses are ordered using Thompson sampling and with random ordering with different number of topics.

This simple artificial example shows that there exists situations where Thompson sampling improves online hypothesis testing. In practice this type of situation could arise from hypotheses belonging in fixed families which have different probabilities of rejection.

The result of the second experiments first result can be seen in 4.5. Here the characteristic length-scale of the kernel function is varied to see its effect on statistical power of alpha investing. The characteristic length-scale has a distinct effect on Thompson sampling's performance. When the length-scale is too small, meaning that the posterior distribution is not updated enough after testing each new hypothesis, Thompson sampling performs at the same level as random sampling. When, the length-scale is increased, the performance of Thompson sampling improves to be much better than the one of random sampling. When the length-scale grows the performance of Thompson sampling slowly decreases closer to the performance of random sampling. This is likely due to the posterior probability of the non-correlated hypothesis are being updated as well resulting in the real correlations being drowned by the noise. Ordering the hypotheses randomly results in a approximately uniform power. With the best length scale, Thompson sampling rejected 1.5 percentage points more of the hypothesis than random sampling. The improvement gained by Thompson sampling is modest but not too sensitive to the correct length scale kernel as long as it is in the right scale.

In the second scenario, the signal strength is varied. The results can be seen from Figure 4.6. Thompson sampling beats random sampling consistently with each signal strength although, again, the result is very modest. On average Thompson sampling beats random sampling by 1.2%. The effect is largest with reasonable sized signal strength and it is less pronounced with both very small and very large signal strengths. In Figure 4.6 it looks like the effect is largest the larger signal strength. This is due to there being more rejections overall and thus the difference although relatively small appears large.

The results of the third experiment resembles the second one but they are more

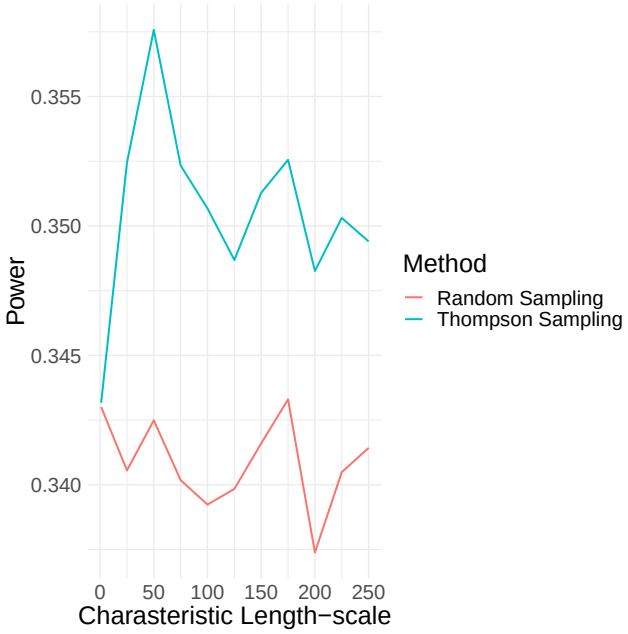

Figure 4.5: Comparison of the statistical power of alpha investing with Thompson sampling to random ordering when the kernel length of the covariance function is varied.

pronounced. Similar to experiment 2, Figure 4.7 shows a clear difference between the performance of the two methods. The effect is again non-existent with too small of a characteristic length-scale of the kernel function. When the length-scale of the kernel is increased to a reasonable length, the results are more pronounced. The same overlearning phenomena is seen in the simulated data as in experiment 2 when the length scale is too large, most likely for the same reasons.

The third experiment performed at best 6.8 percentage points better than random sampling. This was achieved with a characteristic length-scale of the kernel function of 750. This proves that real gain in power can be obtained in practical situations. The difference between the power gain in experiment 2 and 3 is likely due to the distribution of the average particle residuals being smoother than in the synthetic data and such the Gaussian processes are better able to model the joint distribution of the test statistics.

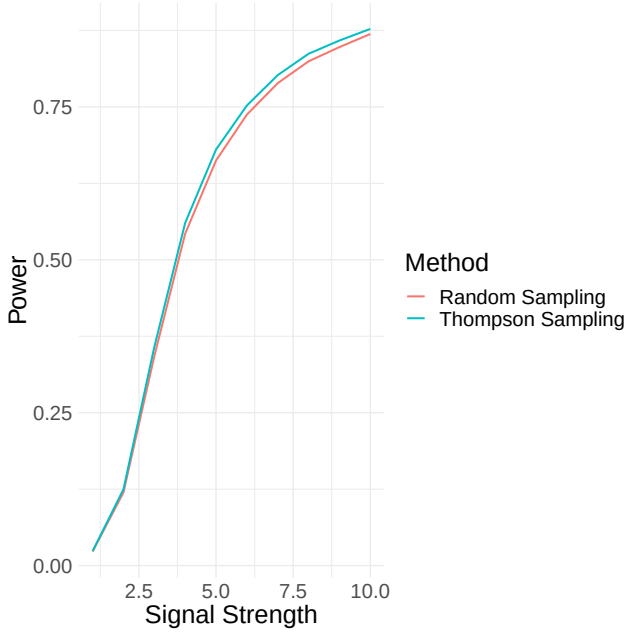

Figure 4.6: Comparison of the statistical power of alpha investing with Thompson sampling to random ordering when the signal strength is varied.

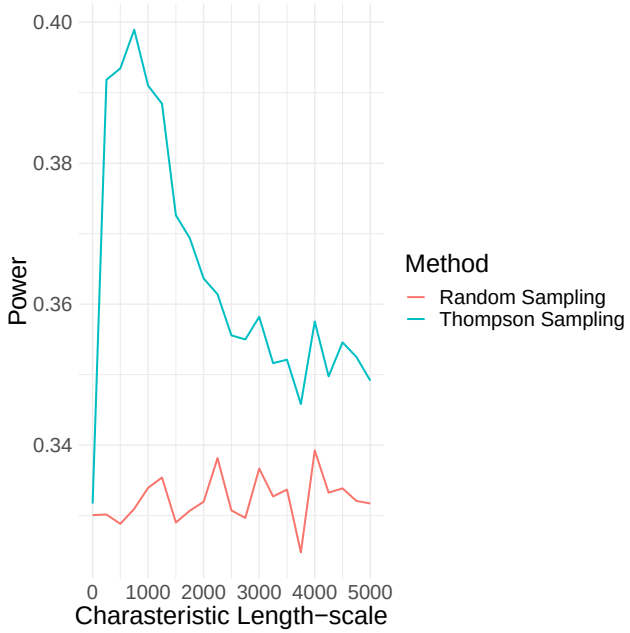

Figure 4.7: Comparison of the statistical power of alpha investing with Thompson sampling to random ordering with the simulated data when the kernel length is varied.

5. Discussion

In this chapter we discuss topics raised forth by this thesis. We begin by inspecting the applicability of the methods raised forth in practice which is the main topic of this chapter. We finalize it by talking of possible avenues for future research.

5.1 Are all of The Assumptions Warranted?

The benefit of the methods presented can only be expected if the assumptions laid down are valid. Thompson sampling only requires the calculation of the posterior distribution of the test statistics. The applicability of Thompson sampling therefore depends on the ability to correctly calculate the posterior distribution of the test statistics.

In the example with simulated data the joint distribution of the test statistics was assumed to follow a Gaussian distribution. The correlation of the test statistics was also assumed to be dependent only on the distance between the two test statistics. This is most likely incorrect since the structures on the geography of the raster causes the residual points of the regression to have much more complicated correlation structures. Even if that was not the case, having the correlations being defined by a squared exponential covariance function is not completely warranted. It is however good to remember that the point of modeling is not necessarily to be perfectly accurate as long as it is useful. This setting is naturally a simplification but the fact that Thompson sampling outperformed random sampling serves as evidence that this assumption is useful in modeling such a situation.

Applying Thompson sampling to alpha investing presents another set of problems.

In experiment 1 the reward distribution is non-stationary. This is because the amount of alpha wealth invested each turn can change depending on the past rejections. In practice this does not seem to be too important in this case, but it is not advisable to spend the wealth too quickly as this would result the spent alpha changing too quickly and the probability of rejection might not be comparable to the one of the last tested hypothesis. In the later experiments this problem was circumvented by modeling the test statistics which is stationary instead of the reward distribution.

Another question that using alpha investing is in equation 2.4. In order to guarantee that mFDR is controlled at the required level this equation must hold. The first experiment's structure is broad enough for this to hold. Alpha investing naturally works even without this assumption but since the statistical guard is lost, some method of estimating the number of falsely rejected hypotheses would be desirable. Constructing independent hypotheses would be the natural way make sure that this constraint is met but that would in turn make any methods of learning the ordering of the hypotheses impossible. If none of the results above, one can always use another method of online hypothesis testing such as alpha spending which does not require this equation to hold.

5.2 Arguments Against P-Values

Although p-values are a fundamental part of modern science they have gained a lot of criticism during the past decades. The main issue in modern science is the lack of reproducible results. Ionnidis [22] in fact argues that most published research is false. This reproducibility crisis stems from the the misuse of p-values.

The issue spurred the American statistician association (ASA) to publish a statement on p-values [8]. They bring forth multiple key issues in usage of p-values. ASA points that p-values are commonly both misused and misunderstood [8]. P-values are notoriously hard to interpret. A common misunderstanding is that p-values measure the probability that a given hypothesis is true [8]. This is hardly surprising as they are

often used as if that would be the case rejecting a null hypothesis solely based on the p-value surpassing some arbitrary threshold (often 0.05). They also do not measure the probability that the data were produced by random chance alone [8]. On top of that, the interpretation of the rejected null hypothesis is often lacking. A rejected hypothesis does not give information on the size or importance of an effect. Another problem in using the p-values as the main tool in research is selective reporting. A researcher is encouraged to find results that are significant at an arbitrarily level while not reporting all the other experiments performed behind closed doors. This results in another multiple hypothesis problem.

Most notably, these arguments have made the Basic and Applied Social Psychology -journal to ban publications involving p-values [23]. In the editorial [23], the null hypothesis significance testing is outright called "invalid".

In light of all this well deserved criticism it should be kept in mind that correctly used p-values are an invaluable tool that have brought to a modern era of science. It is hard to imagine the level of reproducibility crisis had no such statistical tool be used at all. Even in light of the alternatives, the authors of an ASA special issue [24] mentioned applications in which a "highly automated decision rule is needed and the costs of erroneous decisions can be carefully weighed when specifying the threshold" to be an example of a situation where it is warranted to use p-values. Multiple of these types of applications exists for online hypothesis testing to tackle.

5.3 Comparison of The Error Rates

With all the error rates presented in Section 2.4 and many more existing in the literature, it is important to examine which situation should each error rate be used.

The control of the FWER is important in cases where a rejection of an individual null hypothesis results in a false conclusion of a study [10]. This is the case for example in clinical trials where a single interim rejection may lead into the approval of a drug. A variant of FWER, k-FWER is likewise natural when k individual rejections beget a

false conclusion.

Often FWER is however too stringent and a single false rejection is not worth the loss in power. In most cases FDR is very enticing as it still assures that the rejections are expected to be true positives to a degree chosen prior to the experiment. Although formulated slightly differently, mFDR has the same appeal. After the procedure is over the experimenter can be confident that only a small proportion of the rejections are false.

In general, FDR and mFDR can be very different as shown by Javanmard and Montanari in [16] with examples of highly correlated data. On the other hand, for the most basic textbook example they behave very similarly as argued by Foster and Stine [3].

Yet another option is not to use any multiple hypothesis correction (i.e. per comparison error rate). This however is not suggested as it results in many false positives.

5.4 Future Research

Online multiple hypothesis procedures are and will remain a heavily research area of science. However the connection to reinforcement learning inspires future directions of research.

One direction where such learning systems can be applied is in interactive data exploration systems to suggest the next hypothesis. Estimating the posterior probability of each hypothesis that could be tested next allows for the experimenter to have more information and consequently more power when testing hypotheses. This information can be visualized in tandem with the hypotheses themselves in such systems.

The performance of such human experimenters (given the posterior probability of the next possible hypotheses) conducting interactive data exploration should perform close to Thompson sampling presented in chapter 3. This is because humans have been observed to act similarly to Thompson sampling by intuitively matching probabilities

when dealing with uncertainty [25].

In this thesis we have only used Thompson sampling. Although it has many preferable qualities, other methods may situationally outperform Thompson sampling. A complete breakdown and comparison of different methods would provide insight into even better ordering of the hypotheses.

Another way of improving power is in choosing better alpha investing strategies. Taking the same information into account not only in the ordering of the hypotheses but also by creating an adaptive investing rule should result in better performance.

6. Conclusions

In this thesis we have investigated the applicability of reinforcement learning tools to solve the exploration – exploitation problem that often arises in online hypothesis testing. To be precise we used investigated alpha investing as the method of online hypothesis testing. We used Thompson sampling to improve the order where the hypotheses are tested during the testing process.

We created two synthetic data sets to explore the applicability of Thompson sampling when ordering hypotheses. These were compared to a situation where the same hypotheses are tested in a random order. First under a topic model and the second one following simplex noise. We show that when the data is divided in distinct topics with different probability of rejection, Thompson sampling performs a lot better when compared to random sampling. This is not surprising as Thompson sampling is often used to solve the similar multi-armed bandit problem. When the data follows simplex noise (against the assumptions of Thompson sampling), Thompson sampling still performs better than random ordering of the hypotheses although the gain in power is modest.

In addition to the synthetic data sets the method was tested with a real life data set proving that ordering the hypotheses with Thompson sampling performs better than the random baseline in real life situations.

To complement the experiments, the applicability of Thompson sampling is discussed when used with online hypothesis testing. Finally some avenues of future research are discussed.

Bibliography

[1] D. L. Demets and K. K. G. Lan, "Interim analysis: The alpha spending function approach," *Statistics in Medicine*, 1994.

[2] Z. Zhao, L. D. Stefani, E. Zgraggen, C. Binnig, E. Upfal, and T. Kraska, "Controlling false discoveries during interactive data exploration," 2016.

[3] D. P. Foster and R. A. Stine, "α-investing: a procedure for sequential control of expected false discoveries," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 2, pp. 429–444, 2008.

[4] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*. Springer Texts in Statistics, New York: Springer, third ed., 2005.

[5] B. Fristedt and L. Gray, "A modern approach to probability theory," 1996.

[6] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for Machine Learning*. Cambridge University Press, 2020.

[7] T. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. Probability and mathematical statistics, Acad. Press, 1973.

[8] R. L. Wasserstein and N. A. Lazar, "The asa statement on p-values: Context, process, and purpose," *The American Statistician*, vol. 70, no. 2, pp. 129–133, 2016.

[9] J. P. Shaffer, "Multiple hypothesis testing," *Annual Review of Psychology*, vol. 46, no. 1, pp. 561–584, 1995.

[10] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate - a practical and powerful approach to multiple testing," *J. Royal Statist. Soc., Series B*, vol. 57, pp. 289 – 300, 11 1995.

[11] S. Dudoit, J. P. Shaffer, and J. C. Boldrick, "Multiple hypothesis testing in microarray experiments," *Statist. Sci.*, vol. 18, pp. 71–103, 02 2003.

[12] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[13] A. Ramdas, F. Yang, M. J. Wainwright, and M. I. Jordan, "Online control of the false discovery rate with decaying memory," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 5650–5659, Curran Associates, Inc., 2017.

[14] J. Tian and A. Ramdas, "Online control of the familywise error rate," 2019.

[15] E. Aharoni and S. Rosset, "Generalized alpha investing: Definitions, optimality results, and application to public databases," 2013.

[16] A. Javanmard and A. Montanari, "Online rules for control of false discovery rate and false discovery exceedance," *Ann. Statist.*, vol. 46, pp. 526–554, 04 2018.

[17] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, second ed., 2018.

[18] A. Slivkins, "Introduction to multi-armed bandits," *Foundations and Trends in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.

[19] C. E. Rasmussen, *Gaussian Processes in Machine Learning*, pp. 63–71. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[20] P. Chakraborty, C. Ma, J. Grego, and J. Lynch, “Exploratory data analysis for large-scale multiple testing problems and its application in gene expression studies,” 2019.

[21] M. Kurppa, “Assessing pollutant ventilation in city planning alternatives using a large-eddy simulation,” Master’s thesis, University of Helsinki, 2016.

[22] J. P. A. Ioannidis, “Why most published research findings are false,” *PLoS Med*, vol. 2, p. e124, 08 2005.

[23] D. Trafimow and M. Marks, “Editorial,” *Basic and Applied Social Psychology*, vol. 37, no. 1, pp. 1–2, 2015.

[24] R. L. Wasserstein, A. L. Schirm, and N. A. Lazar, “Moving to a world beyond "p < 0.05",” *The American Statistician*, vol. 73, no. sup1, pp. 1–19, 2019.

[25] B. R. Newell, D. J. Koehler, G. James, T. Rakow, and D. van Ravenzwaaij, “Probability matching in risky choice: the interplay of feedback and strategy availability.,” *Memory & cognition*, vol. 41, pp. 329–38, Apr 2013.

Appendix A. Tables From the Experiments

The in-depth tables resulting from the experiments conducted in Section 4 are listed below in the same order as presented in the text:

k	Random Sampling	Thompson Sampling
1	0.4294	0.4296
2	0.4342	0.5519
3	0.4314	0.5975
4	0.4313	0.6214
5	0.4321	0.6316
6	0.4312	0.6390
7	0.4339	0.6421
8	0.4326	0.6375
9	0.4315	0.6345
10	0.4327	0.6350
11	0.4328	0.6323
12	0.4335	0.6294
13	0.4318	0.6246
14	0.4312	0.6214
15	0.4318	0.6196
16	0.4308	0.6162
17	0.4323	0.6138
18	0.4318	0.6097
19	0.4327	0.6082
20	0.4327	0.6049
21	0.4324	0.6010
22	0.4318	0.5981

23	0.4315	0.5962
24	0.4323	0.5931
25	0.4320	0.5910
26	0.4335	0.5894
27	0.4329	0.5857
28	0.4314	0.5837
29	0.4325	0.5823
30	0.4312	0.5791
31	0.4309	0.5776
32	0.4333	0.5755
33	0.4324	0.5744
34	0.4324	0.5707
35	0.4329	0.5698
36	0.4319	0.5675
37	0.4318	0.5645
38	0.4319	0.5623
39	0.4304	0.5607
40	0.4334	0.5602
41	0.4316	0.5569
42	0.4315	0.5549
43	0.4323	0.5524
44	0.4327	0.5520
45	0.4318	0.5492
46	0.4333	0.5481
47	0.4322	0.5471
48	0.4319	0.5441
49	0.4329	0.5424

50	0.4324	0.5419
51	0.4319	0.5395
52	0.4317	0.5387
53	0.4325	0.5374
54	0.4314	0.5340
55	0.4326	0.5337
56	0.4328	0.5319
57	0.4333	0.5302
58	0.4327	0.5292
59	0.4316	0.5278
60	0.4323	0.5265
61	0.4321	0.5260
62	0.4331	0.5238
63	0.4322	0.5229
64	0.4321	0.5218
65	0.4327	0.5205
66	0.4320	0.5192
67	0.4321	0.5175
68	0.4313	0.5176
69	0.4322	0.5154
70	0.4325	0.5156
71	0.4331	0.5143
72	0.4320	0.5142
73	0.4324	0.5118
74	0.4335	0.5107
75	0.4322	0.5100
76	0.4331	0.5105

77	0.4323	0.5078
78	0.4320	0.5074
79	0.4317	0.5063
80	0.4331	0.5061
81	0.4317	0.5051
82	0.4332	0.5037
83	0.4328	0.5034
84	0.4316	0.5017
85	0.4321	0.5012
86	0.4325	0.5009
87	0.4331	0.5003
88	0.4308	0.4996
89	0.4328	0.4989
90	0.4331	0.4993
91	0.4318	0.4976
92	0.4323	0.4960
93	0.4332	0.4965
94	0.4319	0.4951
95	0.4325	0.4952
96	0.4326	0.4932
97	0.4319	0.4923
98	0.4330	0.4931
99	0.4334	0.4926
100	0.4322	0.4913

Table A.1: Tabulated values of experiment 1.

Characteristic Length-Scale	Thompson Sampling	Random Sampling

1	0.3432	0.3430
25	0.3524	0.3406
50	0.3576	0.3425
75	0.3523	0.3402
100	0.3507	0.3392
125	0.3487	0.3398
150	0.3513	0.3416
175	0.3526	0.3433
200	0.3483	0.3374
225	0.3503	0.3405
250	0.3494	0.3414

Table A.2: Tabulated values of experiment 2: section 1

Signal Strength	Thompson Sampling	Random Sampling
1	0.0237	0.0230
2	0.1248	0.1197
3	0.3576	0.3425
4	0.5603	0.5428
5	0.6803	0.6624
6	0.7523	0.7374
7	0.8020	0.7888
8	0.8369	0.8246
9	0.8586	0.8478
10	0.8775	0.8692

Table A.3: Tabulated values of experiment 2: Scenario 2

Characteristic Length-Scale	Thompson Sampling	Random Sampling

1	0.3317	0.3301
250	0.3918	0.3302
500	0.3934	0.3288
750	0.3989	0.3310
1000	0.3910	0.3339
1250	0.3884	0.3354
1500	0.3726	0.3290
1750	0.3694	0.3307
2000	0.3636	0.3320
2250	0.3614	0.3382
2500	0.3556	0.3307
2750	0.3550	0.3297
3000	0.3582	0.3367
3250	0.3516	0.3327
3500	0.3521	0.3337
3750	0.3458	0.3248
4000	0.3575	0.3392
4250	0.3498	0.3332
4500	0.3546	0.3339
4750	0.3525	0.3321
5000	0.3492	0.3317

Table A.4: Tabulated values of experiment 3