DECISION MAKING WITH LIMITED DATA

A Dissertation Presented

by

KIEU MY PHAN

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2022

Robert and Donna Manning College of Information and Computer Sciences

DECISION MAKING WITH LIMITED DATA

A Dissertation Presented

by

KIEU MY PHAN

Approved as to style and content by:

Erik Learned-Miller, Chair

Philip Thomas, Member

Justin Domke, Member

John Staudenmayer, Member

James Allan, Chair of the Faculty
Robert and Donna Manning College of Information and Computer Sciences

ACKNOWLEDGMENTS

I am very fortunate to work with my advisor Professor Erik Learned-Miller. Prof. Learned-Miller gave me an interesting problem, suggested many novel ideas and directions, and inspired me to step out of my comfort zones. He also gave me valuable career advice and has been a great role model as a researcher and mentor.

I would like to thank my committee members - Prof. Justin Domke, Prof. Philip Thomas and Prof. John Staudenmayer. Prof. Domke and Prof. Thomas worked with me in papers that resulted in this thesis and provided me valuable feedback.

I am extremely grateful to my collaborators for many fun and fruitful collaborations. I would also like to thank my office-mates at the LLAMA lab and my friends during graduate school for a fun environment and the department office for providing seamless support for graduate students.

ABSTRACT

DECISION MAKING WITH LIMITED DATA

FEBRUARY 2022

KIEU MY PHAN

B.Sc., KAIST

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Erik Learned-Miller

This thesis studies different approaches to decision making with limited data.

First, we study the effects of approximate inference on Thompson sampling in the k-armed bandit problems. Thompson sampling is a successful algorithm but requires posterior inference, which often must be approximated in practice. We show that even small constant inference error (in α-divergence) can lead to poor performance (linear regret) due to under-exploration (for $\alpha < 1$) or over-exploration (for $\alpha > 0$) by the approximation. While for $\alpha > 0$ this is unavoidable, for $\alpha \leq 0$ the regret can be improved by adding a small amount of forced exploration.

Second, we consider the problem of designing a randomized experiment on a source population to estimate the Average Treatment Effect (ATE) on a target population. We propose a novel approach which explicitly considers the target when designing the experiment on the source. Under the covariate shift assumption, we design an

unbiased importance-weighted estimator for the target population's ATE. To reduce the variance of our estimator, we design a covariate balance condition (Target Balance) between the treatment and control groups based on the target population. We show that Target Balance achieves a higher variance reduction asymptotically than methods that do not consider the target during the design phase. Our experiments illustrate that Target Balance reduces the variance even for small sample sizes.

Finally, we examine confidence intervals. Historically, mean bounds for small sample sizes fall into 2 categories: methods with unrealistic assumptions about the unknown distribution (e.g., Gaussianity) and methods like Hoeffding's inequality that use weaker assumptions but produce much looser intervals. In 1969, Anderson (1969a) proposed a mean confidence interval strictly better than or equal to Hoeffding's whose only assumption is that the distribution's support is contained in an interval $[a, b]$. For the first time since then, we present a new family of upper bounds that compares favorably to Anderson's. We prove that each bound in the family holds with probability at least $1 - \alpha$ for all distributions on an interval $[a, b]$. Furthermore, one of the bounds is tighter than or equal to Anderson's for all samples.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS . . . . . **iv**

ABSTRACT . . . . . **v**

LIST OF FIGURES . . . . . **x**

CHAPTER

INTRODUCTION **1**

1. THOMPSON SAMPLING WITH APPROXIMATE INFERENCE **3**

1.1 Introduction 3
1.2 Background and Notations. 4

1.2.1 The k-armed Bandit Problem. 4
1.2.2 Thompson Sampling with Approximate Inference 5
1.2.3 The Alpha Divergence 6
1.2.4 Problem Statement. 7

1.3 Motivating Example 7
1.4 Regret Analysis When $\alpha > 0$ 9
1.5 Regret Analysis When $\alpha < 1$ 11

1.5.1 Algorithms with Sub-linear Regret for $\alpha \leq 0$ 13

1.6 Simulations 16

1.6.1 Adding Forced Exploration to the Motivating Example 16
1.6.2 Simulations of Ensemble Sampling and Variational Inference for 50-armed bandits 17
1.6.3 Discussion 17

1.7 Related Work 18
1.8 Conclusion 19

2. DESIGNING TRANSPORTABLE EXPERIMENTS UNDER S-ADMISSABILITY 21

2.1 Introduction 21
2.2 Problem Setting 23
2.3 Designing for Transportation 26
2.4 Related Work 29
2.5 Analysis 31

2.5.1 Expectation 31
2.5.2 Variance 32

2.5.2.1 Finite Sample Size Variance Reduction for $d = 1$ 32
2.5.2.2 Asymptotic Variance Reduction for $d \geq 1$ 34

2.6 Simulations 37
2.7 Conclusion 41

3. TOWARDS PRACTICAL MEAN BOUNDS FOR SMALL SAMPLES 42

3.1 Introduction 42

3.1.1 Distribution free bounds with guaranteed coverage 44
3.1.2 Bounds that do not exhibit guaranteed coverage 47
3.1.3 Bounds conjectured to have guaranteed coverage 47

3.2 A Family of Confidence Bounds 48

3.2.1 Guaranteed Coverage 51

3.2.1.1 Preview of proof 52
3.2.1.2 Main Result 53

3.3 Computation 57
3.4 Relationships with Existing Bounds 58

3.4.1 Special Case: Bernoulli Distribution 58
3.4.2 Comparisons with Anderson and Hoeffding 59

3.5 Simulations 64

APPENDICES

A. SUPPLEMENTARY MATERIAL: THOMPSON SAMPLING WITH APPROXIMATE INFERENCE 65

B. SUPPLEMENTARY MATERIAL: DESIGNING TRANSPORTABLE EXPERIMENTS UNDER S-ADMISSABILITY **89**
C. SUPPLEMENTARY MATERIAL: TOWARDS PRACTICAL MEAN BOUNDS FOR SMALL SAMPLES **120**

BIBLIOGRAPHY **152**

LIST OF FIGURES

Figure **Page**

1.1 The Gaussian Q which minimizes $D_\alpha(P, Q)$ for different values of α where the target distribution P is a mixture of two Gaussians. Based on Figure 1 from (Minka, 2005) 7

1.2 Approximation Q_t (with high variance) and approximation Z_t (with small variance) are defined in Section 1.3 where $D_1(\Pi_t, Q_t) = 2$ and $D_0(\Pi_t, Z_t) = 1.5$. Arm 1 is the true best arm. 8

1.3 Updating the posterior by exact Thompson sampling or adding forced exploration does not help the over-explored approximation Q_t, but lowers the regrets of the under-explored approximations Z_t, Ensemble sampling and mean-field VI. Shaded regions show 95% confidence intervals. 20

2.1 Bias, Variance and MSE as a function of the sample size. All unweighted estimators are biases because they measure the ATE of the source distribution. As there is no importance weight threshold, all weighted estimators are unbiased (Theorem 4) but the weighted estimator with Target Balance has the lowest variance. *The y axes are in log scale.* 38

2.2 Bias, Variance and MSE as a function of the distance δ (defined in Section 2.6) between the source and the target distribution. Because of the importance weight threshold, the biases of the importance weighted methods increase as the δ increase. If the distance is too large, the bias of the importance weighted estimators is large, leading to high MSE. However when the distance is not too large, the weighted estimator with Target Balance has the lowest MSE. *The y axes are in log scale.* 39

2.3 Bias, Variance and MSE as a function of importance weight threshold. As the threshold increases, the bias of the weighted methods decreases and the variance of the weighted methods increases. Therefore there is a threshold when the MSE is minimized. The weighted estimator with Target Balance has the lowest MSE for a reasonably good threshold. *The y axes are in log scale.* 39

3.1 49

3.2 Illustrations of Section 3.2.1.1. **Left.** The yellow region shows samples of $\mathbf{z} = [z_{(1)}, z_{(2)}]$ such that $b^\alpha(\mathbf{z}) \leq b^\alpha(\mathbf{x}_{max})$. **Right.** The same yellow region, but in the coordinates $\mathbf{u} = F^{-1}(\mathbf{z})$. We will show that the yellow region is a subset of the striped, which contains $\mathbf{u}$ such that $b(\mathbf{x}_{max}, \mathbf{u}) \geq \mu$. 52

3.3 The expected value of the bounds for $\alpha = 0.05$ and $D^+ = [0, 1]$. For each sample size, we sample $\mathbf{X}$ 10,000 times, compute the bound for each sample, and take the average. Our new bound with T being Anderson's bound consistently has lower expected value than Anderson's (Theorem 9), Hoeffding's (Theorem 10) and Maurer and Pontil's. With T being the $l2$-norm, the bound is substantially tighter in these examples, and also has guaranteed coverage. 63

3.4 The α-quantile of the bound distribution for $\alpha = 0.05$ and $D^+ = [0, 1]$. For each sample size, we sample $\mathbf{X}$ 10,000 times, compute the bound for each sample, and take the α quantile. If the α-quantile is below the true mean, the bound does not have guaranteed coverage. For the $\texttt{uniform}(0, 1)$ and $\texttt{beta}(1, 5)$ distribution, when the sample size is small, Student-t does not have guarantee. 63

A.1 Lower bound of regret per time-step as a function of ϵ when $m_1^* - m_2^* = 0.1$ and we sample from the approximation for $T/2$ time-steps in the example construted in the proof of Theorem 1. When α is around 1, the lower bound converges quickly as ϵ increases. 72

A.2 ϵ as a function z that makes the regret linear for different values of α for the example constructed in the proof of Theorem 2. 76

C.1 Finding the upper bound of the mean with $D^+ = [0, 1]$ 124

C.2 Finding the upper bound of the mean with $D^+ = [0, 1]$ 125

C.3 Finding the upper bound of the mean with $D^+ = [0, 10]$ 126

C.4 Finding the upper bound of the mean with $D^+ = (-\infty, 1]$ 127

C.5 Finding the upper bound of the mean with $D^+ = (-\infty, 10]$ 128

C.6 Finding the lower bound of the mean with $D^+ = [0, \infty)$ 129

C.7 The PDFs of $lognorm(0, 1)$ and the sample mean distribution of $lognorm(0, 1)$. The sample mean distribution of $lognorm(0, 1)$ is visibly skewed when the sample size $n = 80$. 130

INTRODUCTION

In this thesis we study different approaches to decision making with limited data. The settings for which one would want to make good decisions with limited data occur frequently in practice. In recommendation systems, one would like to estimate the optimal choice with a small number of user interactions. In small clinical trials (Institute of Medicine, 2001), more data could be either expensive, difficult to obtain or cannot be obtained due to safety reasons. For some problems such as finding mean bounds for small samples, there are few solution existed that have both guarantees and good performances.

There are several aspects to making decisions with limited data. The first is to make decisions that adapt continuously as the data arrives. The multi-armed bandit framework (Lattimore and Szepesvári, 2020) formalizes the problem as the exploration-exploitation trade-offs. In our first project Thompson Sampling with Approximate inference (Chapter 1), we study the performance of Thompson sampling, a multi-armed bandit algorithm, when the posterior samples are approximated. The chapter is based on the paper published in the Conference on Neural Information Processing Systems 2019 (Phan et al., 2019).

The second is to estimate the causal effects among variables to use the causal structure to guide the decisions and therefore making more accurate decisions with fewer data. In our second project Designing Transportable Experiments Under S-Admissablility (Chapter 2), we consider the problem of designing experiments performed on a source population in order to estimate the effect on a target population. The chapter is based on the paper published in the International Conference on Artificial Intelligence and Statistics 2021 (Phan et al., 2021a).

Finally, it is important to have tighter estimates of the value of interest with small samples, so that the decision maker can adapt quickly as the data arrives. In our third project Towards Practical Mean Bounds for Small Samples (Chapter 3), we consider the problem of finding mean bounds for small sample sizes. The chapter is based on the paper published in the International Conference on Machine Learning 2021 (Phan et al., 2021b).

CHAPTER 1

THOMPSON SAMPLING WITH APPROXIMATE INFERENCE

1.1 Introduction

The chapter is based on the paper published in the Conference on Neural Information Processing Systems 2019 (Phan et al., 2019). The stochastic k-armed bandit problem is a sequential decision making problem where at each time-step t, a learning agent chooses an action (arm) among k possible actions and observes a random reward. Thompson sampling (Russo et al., 2018) is a popular approach in bandit problems based on sampling from a posterior in each round. Bandit algorithms are usually evaluated by regret, which is the difference between the best action and the agent's action. Linear regret implies that the agent does not learn the optimal behaviour. Thompson sampling has been shown to have good performance both in term of frequentist regret and Bayesian regret for the k-armed bandit problem under certain conditions.

This chapter investigates Thompson sampling when only an *approximate* posterior is available. This is motivated by the fact that in complex models, approximate inference methods such as Markov Chain Monte Carlo or Variational Inference must be used. Along this line, Lu and Van Roy (2017) propose a novel inference method – Ensemble sampling – and analyze its regret for linear contextual bandits. To the best of our knowledge this is the most closely related theoretical analysis of Thompson sampling with approximate inference.

This chapter analyzes the regret of Thompson sampling with approximate inference. Rather than considering a particular inference algorithm, we parameterize the error

using the α-divergence, a typical measure of inference accuracy. Our contributions are as follows:

- **Even small inference errors can lead to linear regret with naive Thompson sampling.** Given any error threshold $\epsilon > 0$ and any α we show that approximate posteriors with error at most ϵ in α-divergence at all times can result in linear regret (both frequentist and Bayesian). For $\alpha > 0$ and for any reasonable prior, we show linear regret due to over-exploration by the approximation (Theorem 1, Corollary 1). For $\alpha < 1$ and for priors satisfying certain conditions, we show linear regret due to under-exploration by the approximation, which prevents the posterior from concentrating (Theorem 2, Corollary 2).

- **Forced exploration can restore sub-linear regret.** For $\alpha \leq 0$ we show that adding forced exploration to Thompson sampling can make the posterior concentrate and restore sub-linear regret (Theorem 3) even when the error threshold is a very large constant. We illustrate this effect by showing that the performances of Ensemble sampling (Lu and Van Roy, 2017) and mean-field Variation Inference (Blei et al., 2017) can be improved in this way either theoretically (Section 1.5.1) or in simulations (Section 1.6).

1.2 Background and Notations.

1.2.1 The k-armed Bandit Problem.

We consider the k-armed bandit problem parameterized by the mean reward vector $m^* = (m_1^*, ..., m_k^*) \in \mathcal{R}^k$, where m_i^* denotes the mean reward of arm (action) i and Y_t^i denote the reward of choosing action i at time t[1]. At each round t, the learner chooses an action A_t and observes the outcome Y_t which, conditioned on A_t, is independent of the history up to and not including time t, $H_{t-1} = (A_1, Y_1, ..., A_{t-1}, Y_{t-1})$. For a time

[1]We omit the superscript when the context is clear.

horizon T, the goal of the algorithm π is to maximize the expected cumulative reward up to time T.

Let $\Omega \subseteq \mathcal{R}^k$ be the domain of the mean and $\Omega_i \subseteq \Omega$ denote the region where the ith arm is among the arms with the largest mean. For simplicity we assume $P(\Omega_i \cap \Omega_j) = 0$ for any $i \neq j, 1 \leq i, j \leq k$. Let the function $A^* : \Omega \to \{a_1, ..., a_k\}$ denoting the best action be defined as: $A^*(m) = i$ if $m \in \Omega_i$.

In the frequentist setting we assume that there exists a true mean m^* which is fixed and unknown to the learner. Therefore, a policy π^* that always chooses $A^*(m^*)$ will get the highest reward. The performance of policy π is measured by its expected regret compared to an optimal policy π^*, which is defined as:

$$\text{Regret}(T, \pi, m^*) = T m^*_{A^*(m^*)} - \mathbb{E} \sum_{t=1}^{T} m^*_{A_t} \ . \tag{1.1}$$

If the regret is linear, the agent chooses a sub-optimal arm for a constant number of time-steps, which implies that the agent does not learn the optimal arm. An naive algorithm that achieves linear regret is one that chooses an arm uniformly at random at every time step.

On the other hand, in the Bayesian setting, an agent expresses her beliefs about the mean vector in terms of a prior Π_0, and therefore, the mean is treated as a random variable $M = (M_1, ..., M_k)$ distributed according to the prior Π_0. The Bayesian regret is the expectation of the regret under the prior of parameter M:

$$\text{BayesRegret}(T, \pi) = \mathbb{E}_{\Pi_0} \text{Regret}(T, \pi, M) \ . \tag{1.2}$$

1.2.2 Thompson Sampling with Approximate Inference

In the frequentist setting, in order to perform Thompson sampling we define a prior which is only used in the algorithm. On the other hand, in the Bayesian setting the prior is given.

Let Π_t be the posterior distribution of $M|H_{t-1}$ with density function $\pi_t(m)$. Thompson sampling obtains a sample $\widehat{m}$ from Π_t and then selects arm A_t as follow: $A_t = i$ if $\widehat{m} \in \Omega_i$. In each round, we assume an approximate sampling method is available that generates sample from an approximate distribution Q_t. We use q_t to denote the density function of Q_t.

Popular approximate sampling methods include Markov Chain Monte Carlo (MCMC) (Andrieu et al., 2003), Sequential Monte Carlo (Doucet and Johansen, 2009) and Variational Inference (VI) (Blei et al., 2017). There are packages that conveniently implement VI and MCMC methods, such as Stan (Carpenter et al., 2017), Edward (Tran et al., 2016), PyMC (Salvatier et al., 2016) and infer.NET (Minka et al., 2018).

To provide a general analysis of approximate sampling methods, we will use the α-divergence (Section 1.2.3) to quantify the distance between the posterior Π_t and the approximation Q_t.

1.2.3 The Alpha Divergence

The α-divergence between two distributions P and Q with density functions $p(x)$ and $q(x)$ is defined as:

$$D_\alpha(P, Q) = \frac{1 - \int p(x)^\alpha q(x)^{1-\alpha} dx}{\alpha(1-\alpha)}. \tag{1.3}$$

α-divergence generalizes many divergences, including $KL(Q, P)$ ($\alpha \to 0$), $KL(P, Q)$ ($\alpha \to 1$), Hellinger distance ($\alpha = 0.5$) and χ^2 divergence ($\alpha = 2$) and is a common way to measure errors in inference methods. MCMC errors are measured by the Total Variation distance, which can be upper bounded by the KL divergence using Pinsker's inequality ($\alpha = 0$ or $\alpha = 1$). Variational Inference tries to minimize the reverse KL divergence (information projection) between the target distribution and the approximation ($\alpha = 0$). Ensemble sampling (Lu and Van Roy, 2017) provides

error guarantees using reverse KL divergence ($\alpha = 0$). Expectation Propagation tries to minimize the KL divergence ($\alpha = 1$) and χ^2 Variational Inference tries to minimize the χ^2 divergence ($\alpha = 2$).

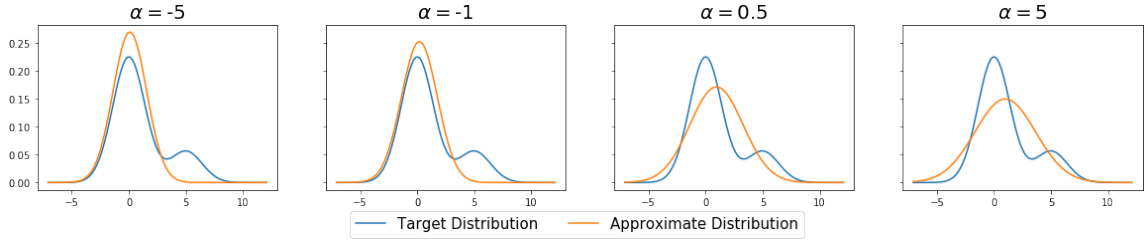


Figure 1.1: The Gaussian Q which minimizes $D_\alpha(P, Q)$ for different values of α where the target distribution P is a mixture of two Gaussians. Based on Figure 1 from (Minka, 2005)

When α is small, the approximation fits the posterior's dominant mode. When α is large, the approximation covers the posterior's entire support (Minka, 2005) as illustrated in Figure 1.1. Therefore changing α will affect the exploration-exploitation trade-off in bandit problems.

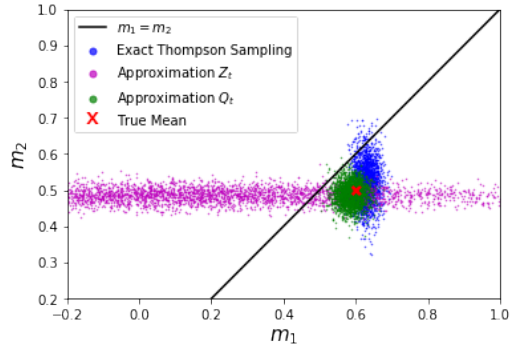
1.2.4 Problem Statement.

Problem Statement. For the k-armed bandit problem, given α and $\epsilon > 0$, if at all time-steps t we sample from an approximate distribution Q_t such that $D_\alpha(\Pi_t, Q_t) < \epsilon$, will the regret be sub-linear in t?

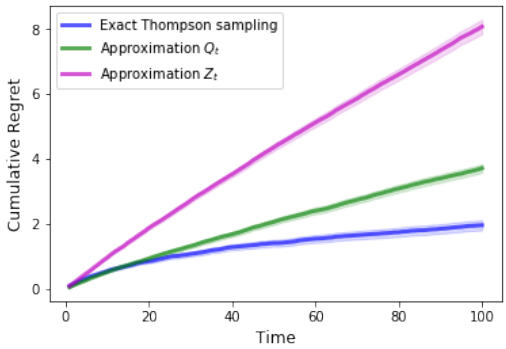
1.3 Motivating Example

In this section we present a simple example to show the effects of inference errors on the frequentist regret. We design the example to illustrate the effects of under-exploration and over-exploration. While in practice, it is possible for the posterior to over-explore in some dimensions and under-explore in others, the example is designed to focus on analyzing each type of errors separately.

Example. Consider a 2-armed bandit problem where the reward distributions are $\text{Norm}(0.6, 0.2^2)$ and $\text{Norm}(0.5, 0.2^2)$ for arm 1 and 2 respectively. The prior Π_0 is Norm $\left(\mu_0^T, 0.5^2 I\right)$ where $\mu_0 = [0.1, 0.9]$ is the vector of prior means of arm 1 and 2 respectively, and I denotes the identity matrix.



(a) Over-dispersed (approximation Q_t) and under-dispersed sampling (approximation Z_t) yield different posteriors after $T = 100$ time-steps. m_1 and m_2 are the means of arms 1 and 2. Q_t picks arm 2 more often than exact Thompson sampling and Z_t mostly picks arm 2. The posteriors of exact Thompson sampling and Q_t concentrate mostly in the region where $m_1 > m_2$ while Z_t's spans both regions.



(b) The regret of sampling from the approximations Q_t and Z_t are both larger than that of exact Thompson sampling from the true posterior Π_t. Shaded regions show 95% confidence intervals.

Figure 1.2: Approximation Q_t (with high variance) and approximation Z_t (with small variance) are defined in Section 1.3 where $D_1(\Pi_t, Q_t) = 2$ and $D_0(\Pi_t, Z_t) = 1.5$. Arm 1 is the true best arm.

Let $\Pi_t = \text{Norm}(\mu_t, \Sigma_t)$ be the posterior at time t. Approximations Q_t and Z_t are calculated such that $\text{KL}(\Pi_t, Q_t) = 2$ and $\text{KL}(Z_t, \Pi_t) = 1.5$ by multiplying the covariance Σ_t by a constant: $Q_t = \text{Norm}(\mu_t, 4.5^2\Sigma_t)$ and $Z_t = \text{Norm}(\mu_t, 0.3^2\Sigma_t)$.

We perform the following simulations 1000 times and plot the mean cumulative regret up to time $T = 100$ in Figure 1.2b using three different policies:

1. (**Exact Thompson Sampling**) At each time-step t, sample from the true posterior Π_t.

2. (**Approximation** Q_t) At each time-step t, compute Q_t from Π_t and sample from Q_t.
3. (**Approximation** Z_t) At each time-step t, compute Z_t from Π_t and sample from Z_t.

The regrets of sampling from the approximations Q_t and Z_t are in both cases larger than that of exact Thompson sampling. Intuitively, the regret of Q_t is larger because Q_t explores more than the true posterior (Figure 1.2a). In Section 1.4 we show that when $\alpha > 0$ the approximation can incur this type of error, leading to linear regret. On the other hand, the regret of Z_t is larger because Z_t explores less than the exact Thompson sampling algorithm and therefore commits to the sub-optimal arm (Figure 1.2a). In Section 1.5 we show that when $\alpha < 1$ the approximation can change the posterior concentration rate, leading to linear regret. We also show that adding a uniform sampling step can help the posterior to concentrate when $\alpha \leq 0$, and make the regret sub-linear.

1.4 Regret Analysis When $\alpha > 0$

In this section we analyze the regret when $\alpha > 0$. Our result shows that the approximate method might pick the sub-optimal arm with constant probability in every time-step, leading to linear regret even when the algorithm only samples from the approximation for $0.01T$ time-steps (where 0.01 stands for a very small positive constant) and use any optimal policy in the remaining $0.99T$ time-steps. The result implies that $\alpha > 0$ should be used for at most $o(T)$ time-steps if we want to achieve sub-linear regret.

Theorem 1 (Frequentist Regret)**.** *Let $\alpha > 0$, the number of arms be $k = 2$ and $m_1^* > m_2^*$. Suppose that at any time t, the posterior Π_t has density π_t. For any error threshold $\epsilon > 0$, there is a deterministic mapping $f(\Pi)$ such that for all $t \geq 0$:*

1. *Sampling from $Q_t = f(\Pi_t)$ chooses arm 2 with a constant probability.*

2. $D_\alpha(\Pi_t, Q_t) < \epsilon$.

Therefore sampling from Q_t for $c \cdot T$ time-steps for any constant $c > 0$ and using any policy for the remaining time-steps will cause linear frequentist regret.

Typically, approximate inference methods minimize divergences. Broadly speaking, this theorem shows that making a divergence a small constant, alone, is not enough to guarantee sub-linear regret. We do not mean to imply that low regret is *impossible* but simply that making an α-divergence a small constant alone is not sufficient.

At every time-step, the mapping f constructs the approximation Q_t from the posterior Π_t by moving probability mass from the region Ω_1 where $m_1 > m_2$ to the region Ω_2 where $m_2 > m_1$. Then Q_t will choose arm 2 with a constant probability at every time-step.

Therefore, if we sample from $Q_t = f(\Pi_t)$ for $0.1T$ time steps and use any policy in the remaining $0.9T$ time steps, we will still incur linear regret from the $0.1T$ time-steps. On the other hand, when $\alpha \leq 0$, we show in Section 1.5.1 that sampling an arm uniformly at random for $\log T$ time-steps and sampling from an approximate distribution that satisfies the divergence constraint for $T - \log T$ time-steps will result in sub-linear regret.

Agrawal and Goyal (2013) show that the frequentist regret of exact Thompson sampling is $O(\sqrt{T})$ with Gaussian or Beta priors and bounded rewards. Theorem 1 implies that when the assumptions in (Agrawal and Goyal, 2013) are satisfied but there is a small constant inference error at every time-step, the regret is no longer guaranteed to be sub-linear.

If the assumption $m_1^* > m_2^*$ in Theorem 1 is satisfied with a non-zero probability ($\mathbb{P}_{\Pi_0}(M_1 > M_2) > 0$), the Bayesian regret will also be linear:

Corollary 1 (Bayesian Regret)**.** *Let $\alpha > 0$ and the number of arms be $k = 2$. Let Π_0 be a prior where $\mathbb{P}_{\Pi_0}(M_1 > M_2) > 0$. Suppose that at any time t, the posterior Π_t has*

density π_t. Then for any error threshold $\epsilon > 0$, there is a deterministic mapping $f(\Pi)$ such that for all $t \geq 0$ the two statements in Theorem 1 hold.

Therefore sampling from Q_t for cT time-steps for any constant $c > 0$ and using any policy for the remaining time-steps will cause linear Bayesian regret.

Russo and Roy (2016) prove that the Bayesian regret of Thompson sampling for k-armed bandits with sub-Gaussian rewards is $O(\sqrt{T})$. Corollary 1 implies that even when the assumptions in (Russo and Roy, 2016) are satisfied, under certain conditions and with approximation errors, the regret is no longer guaranteed to be sub-linear.

1.5 Regret Analysis When $\alpha < 1$

In this section we analyze the regret when $\alpha < 1$. Our result shows that for any error threshold, if the posterior Π_t places too much probability mass on the wrong arm then the approximation Q_t is allowed to avoid the optimal arm. If the sub-optimal arms do not provide information about the arms' ranking, the posterior Π_{t+1} does not concentrate. Therefore Q_{t+1} is also allowed to be close in α-divergence while avoiding the optimal arm, leading to linear regret in the long term.

Theorem 2 (Frequentist Regret)**.** *Let $\alpha < 1$, the number of arms be $k = 2$ and $m_1^* > m_2^*$. Let Π_0 be a prior where M_2 and $M_1 - M_2$ are independent and $\mathbb{P}_{\Pi_0}(M_2 \geq M_1) > 0$. Suppose that at any time t, the posterior Π_t has density π_t. There is a deterministic mapping $f(\Pi)$ such that for all $t \geq 0$:*

1. *Sampling from $Q_t = f(\Pi_t)$ chooses arm 2 with probability 1.*
2. *For any $\epsilon > 0$, there exists $0 < z < 1$ such that if $\mathbb{P}_{\Pi_0}(M_2 \geq M_1) = z$ and arm 2 is chosen at all times before t then $D_\alpha(\Pi_t, Q_t) < \epsilon$.*

   For any $0 < z < 1$, there exists $\epsilon > 0$ such that if $\mathbb{P}_{\Pi_0}(M_2 \geq M_1) = z$ and arm 2 is chosen at all times before t then $D_\alpha(\Pi_t, Q_t) < \epsilon$.

Therefore sampling from Q_t at all time-steps results in linear frequentist regret.

We discuss why the above results are not immediately obvious. When $\alpha \to 0$, the α-divergence becomes $\mathrm{KL}(Q_t, \Pi_t)$. We might believe that the regret should be sub-linear in this case because the posterior Π_t becomes more concentrated, and so the total variation between Q_t and Π_t must decrease. For example, Ordentlich and Weinberger (2004) show the distribution-dependent Pinsker's inequality between $\mathrm{KL}(Q, P)$ and the total variation $\mathrm{TV}(P, Q)$ for discrete distributions P and Q as follows:

$$\mathrm{KL}(Q, P) \geq \phi(P) \cdot \mathrm{TV}(P, Q)^2 \,. \tag{1.4}$$

Here, $\phi(P)$ is a quantity that will increase to infinity if P becomes more concentrated. However, the algorithm in Theorem 2 constructs an approximation distribution that never picks the optimal arm, so the posterior Π_t can not concentrate and the regret is linear. The error threshold ϵ causing linear frequentist regret is correlated with the probability mass the prior places on the true best arm.

With some assumptions on the rewards, Gopalan et al. (2014) show that the problem-dependent frequentist regret is $O(\log T)$ for finitely-supported, correlated priors with $\pi_0(m^*) > 0$. Liu and Li (2016) study the prior-dependent frequentist regret of 2-armed-and-2-models bandits, and show that with some smoothness assumptions on the reward likelihoods, the regret is $O(\sqrt{T/\mathbb{P}_{\Pi_0}(M_2 > M_1)}$ if arm 1 is the better arm. Theorem 2 implies that when the assumptions in Gopalan et al. (2014) or Liu and Li (2016) are satisfied, if M_2 and $M_1 - M_2$ are independent and there are approximation errors, the regret is no longer guaranteed to be sub-linear.

If the assumption $m_1^* > m_2^*$ in Theorem 2 is satisfied with a non-zero probability ($\mathbb{P}_{\Pi_0}(M_1 > M_2) > 0$), the Bayesian regret wil also be linear:

Corollary 2 (Bayesian Regret)**.** *Let $\alpha < 1$ and the number of arms be $k = 2$. Let Π_0 be a prior where M_2 and $M_1 - M_2$ are independent, $\mathbb{P}_{\Pi_0}(M_1 > M_2) > 0$ and*

$\mathbb{P}_{\Pi_0}(M_2 \geq M_1) > 0$. *Suppose that at any time* t*, the posterior* Π_t *has density* π_t*. Then there is a deterministic mapping* $f(\Pi)$ *such that for all* $t \geq 0$ *the 2 statements in Theorem 2 hold.*

Therefore sampling from Q_t *at all time-steps results in linear Bayesian regret.*

Russo and Roy (2016) prove that the Bayesian regret of Thompson sampling for k-armed bandits with sub-Gaussian rewards is $O(\sqrt{T})$. Corollary 2 implies that even when the assumptions in Russo and Roy (2016) are satisfied, under certain conditions and with approximation errors, the regret is no longer guaranteed to be sub-linear.

We note that, unlike the case when $\alpha > 0$, if we use another policy in $o(T)$ time-steps to make the posterior concentrate and sample from Q_t for the remaining time-steps, the regret can be sub-linear. We provide a concrete algorithm in Section 1.5.1 for the case when $\alpha \leq 0$.

1.5.1 Algorithms with Sub-linear Regret for $\alpha \leq 0$

In the previous section, we see that when $\alpha < 1$, the approximation has linear regret because the posterior does not concentrate. In this section we show that when $\alpha \leq 0$, it is possible to achieve sub-linear regret even when ϵ is a very large constant by adding a simple exploration step to force the posterior to concentrate (the case of $\alpha > 0$ cannot be improved according to Theorem 1). We first look at the necessary and sufficient condition that will make the posterior concentrate, and then provide an algorithm that satisfies it. Russo (2016) and Qin et al. (2017) both show the following result under different assumptions:

Lemma 1 (Lemma 14 from (Russo, 2016))**.** *Let* $m^* \in \mathcal{R}^k$ *be the true parameter and let* $a^* = A^*(m^*)$ *be the true best arm. If for all arms* i*,* $\sum_{t=1}^{\infty} P(A_t = i|H_{t-1}) = \infty$*, then*

$$\lim_{t\to\infty} P(A^*(M) = a^*|H_{t-1}) = 1 \text{ with probability } 1 \,. \tag{1.5}$$

If there exists arm i such that $\sum_{t=1}^{\infty} P(A_t = i|H_{t-1}) < \infty$, then $\liminf_{t\to\infty} P(A^*(M) = i|H_{t-1}) > 0$ with probability 1.

Russo (2016) make the following assumptions, which allow correlated priors:

Assumption 1. *Let the reward distributions be in the canonical one dimensional exponential family with the density: $p(y|m) = b(y)exp\{mT(y) - A(m)\}$ where b, T and A are known function and $A(m)$ is assumed to be twice differentiable. The parameter space $\Omega = (\overline{m}, \underline{m})$ is a bounded open hyper-rectangle, the prior density is uniformly bounded with $0 < \inf_{m\in\Omega} \pi_0(m) < \sup_{m\in\Omega} \pi_0(m) < \infty$ and the log-partition function has bounded first derivative with $\sup_{\theta\in[\overline{m},\underline{m}]} |A'(m)| < \infty$.*

Qin et al. (2017) make the following assumptions:

Assumption 2. *Let the prior be an uncorrelated multivariate Gaussian. Let the reward distribution of arm i be $\text{Norm}(m_i, \sigma^2)$ with a common known variance σ^2 but unknown mean m_i.*

Even though we consider the error in sampling from the posterior distribution, the regret is a result of choosing the wrong arm. We define $\overline{\Pi}_t$ as the posterior distribution of the best arm and $\overline{Q}_t$ as the approximation of $\overline{\Pi}_t$ with the density functions

$$\overline{\pi}_t(i) = P(A^* = i|H_{t-1}) \text{ and } \overline{q}_t(i) = P(A_t = i|H_{t-1}).$$

We now define an algorithm where each arm will be chosen infinitely often, satisfying the condition of Lemma 1.

Theorem 3 (Bayesian and Frequentist Regret)**.** *Consider the case when Assumption 1 or 2 is satisfied. Let $\alpha \leq 0$ and $p_t = o(1)$ be such that $\sum_{t=1}^{\infty} p_t = \infty$. For any number of arms k, any prior Π_0 and any error threshold $\epsilon > 0$, the following algorithm has $o(T)$ frequentist regret: at every time-step t,*

- *with probability* $1 - p_t$*, sample from an approximate posterior* Q_t *such that* $D_\alpha(\overline{\Pi}_t, \overline{Q}_t) < \epsilon$*,*
- *with probability* p_t*, sample an arm uniformly at random.*

Since the Bayesian regret is the expectation of the frequentist regret over the prior, for any prior if the frequentist regret is sub-linear at all points the Bayesian regret will be sub-linear.

The following lemma shows that the error in choosing the arms is upper bounded by the error in choosing the parameters. Therefore whenever the condition $D_\alpha(\Pi_t, Q_t) < \epsilon$ is satisfied, the condition $D_\alpha(\overline{\Pi}_t, \overline{Q}_t) < \epsilon$ will be satisfied and Theorem 3 is applicable.

Lemma 2.

$$D_\alpha(\overline{\Pi}_t, \overline{Q}_t) \leq D_\alpha(\Pi_t, Q_t)\,.$$

We also note that we can achieve sub-linear regret even when ϵ is a very large constant. We revisit Eq. 1.4 to provide the intuition: $\mathrm{KL}(Q, P) \geq \phi(P) \cdot \mathrm{TV}(P, Q)^2$. Here, $\phi(P)$ is a quatity that will increase to infinity if P becomes more concentrated. Hence, if $KL(\overline{Q}_t, \overline{\Pi}_t) < \epsilon$ for any constant ϵ and $\overline{\Pi}_t$ becomes concentrated, the total variation $\mathrm{TV}(\overline{Q}_t, \overline{\Pi}_t)$ will decrease. Therefore, $\overline{Q}_t$ will become concentrated, resulting in sub-linear regret.

Application. Lu and Van Roy (2017) propose an approximate sampling method called Ensemble sampling where they maintain a set of $\mathcal{M}$ models to approximate the posterior and analyze its regret for the linear contextual bandits when $\mathcal{M}$ is $\Omega(\log(T))$. For the k-armed bandit problem and when $\mathcal{M}$ is $\Theta(\log(T))$, Ensemble sampling satisfies the condition $\mathrm{KL}(\overline{Q}_t, \overline{\Pi}_t) < \epsilon$ in Theorem 3 with high probability. In this case, Lu and Van Roy (2017) show a regret bound that scales linearly with T. We can apply Theorem 3 to get sub-linear regret with Ensemble sampling when $\mathcal{M}$ is $\Theta(\log(T))$.

1.6 Simulations

For each approximation method we repeat the following simulations for 1000 times and plot the mean cumulative regret, using five different policies.

1. (**Exact Thompson sampling**) Use exact posterior sampling to choose an action to calculate the reward and update the posterior (for reference).
2. (**Approximation method**) Use the approximation method to choose an action to calculate the reward and update the posterior. We use the approximation naively without any modification.
3. (**Forced Exploration**) With a probability (the exploration rate), choose an action uniformly at random to calculate the reward and update the posterior. Otherwise, use the approximation method to choose an action to calculate the reward and update the posterior. This is the method suggested by Thm. 3.
4. (**Approximate Reward**) Use the approximation method to choose an action to calculate the reward. Use exact posterior sampling to choose an action to update the posterior.
5. (**Approximate Update**) Use exact posterior sampling to choose an action to calculate the reward. Use the approximate method to choose an action to update the posterior.

The last two policies are performed to understand how the approximation affects the posterior (discussed in Section 1.6.3). We update the posterior using the closed-form formula when both the prior and reward distribution are Gaussian.

1.6.1 Adding Forced Exploration to the Motivating Example

In this section we revisit the example in Section 1.3. We apply Q_t, Z_t and Ensemble sampling with $\mathcal{M} = 2$ models to the bandit problem described in the example. We set the exploration rate at time t to be $1/t$, $T = 100$ and show the results in Figure 1.3a and discuss them in Section 1.6.3.

1.6.2 Simulations of Ensemble Sampling and Variational Inference for 50-armed bandits

Now we add forced exploration to mean-field Variational Inference (VI) and Ensemble Sampling with $\mathcal{M} = 5$ models for a 50-armed bandit instance. We generate the prior and the reward distribution as follows: the prior is $\text{Norm}(\mathbf{0}, \Sigma_0)$. To generate a positive semi-definite matrix Σ_0, we generate a random matrix A of size (k, k) where entries are uniformly sampled from $[0, 1)$ and set $\Sigma_0 = A^T A/k$. The true mean m^* is sampled from the prior. The reward distribution of arm i is $\text{Norm}(m_i^*, 1)$.

Mean-field VI approximates the posterior by finding an uncorrelated multivariate Gaussian distribution Q_t that minimizes $KL(\Pi_t, Q_t)$. If the posterior is $\Pi_t = \text{Norm}(\mu_t, \Sigma_t)$ then Q_t has the closed-form solution $Q_t = \text{Norm}(\mu_t, \text{Diag}(\Sigma_t^{-1})^{-1})$, which we used to perform the simulations. We set the exploration rate at time t to be $50/t$, $T = 3000$, show the results in Figure 1.3b and discuss them in Section 1.6.3.

1.6.3 Discussion

We observe in Figure 1.3a that the regret of Q_t and the regret of Approximate Reward with Q_t are similar, and the regret of Approximate Update with Q_t and the regret of exact Thompson sampling are similar. These two observations imply that Q_t has the same effect on the posterior as exact Thompson sampling. Therefore adding forced exploration is not helpful.

On the other hand, in Figures 1.3a and 1.3b the regrets of Approximate Reward with Z_t, Ensemble sampling and mean-field VI decrease significantly, with the regrets of Approximate Update with these 3 methods are high. This behaviour is likely because the approximation causes the posterior to concentrate in the wrong region[2].

[2]Note that in the case where there are 2 arms (Figure 1.3a), Approximate Update has slightly lower regret than naively using the approximate method. This is only because there are only 2 regions, so exact reward sampling explores more than the approximation in the other region, which happens to be the correct one.

In combination, these two observations suggest that these methods do not explore enough for the posterior to concentrate. Therefore adding forced exploration is helpful, which is compatible with the result in Theorem 3.

1.7 Related Work

There have been many works on sub-linear Bayesian and frequentist regrets for exact Thompson sampling. We discussed relevant works in detail in Section 1.4 and Section 1.5.

Ensemble sampling (Lu and Van Roy, 2017) gives a theoretical analysis of Thompson sampling with one particular approximate inference method. Lu and Van Roy (2017) maintain a set of $\mathcal{M}$ models to approximate the posterior, and analyzed its regret for linear contextual bandits when $\mathcal{M}$ is $\Omega(\log(T))$. For the k-armed bandit problem and when $\mathcal{M}$ is $\Theta(\log(T))$, Ensemble sampling satisfies the condition $\mathrm{KL}(\overline{Q}_t, \overline{\Pi}_t) < \epsilon$ in Theorem 3 with high probability. In this case, the regret of Ensemble sampling scales linearly with T.

We show in Theorem 2 that when the constraint $\mathrm{KL}(Q_t, \Pi_t) < \epsilon$ is satisfied, which implies by Lemma 2 that $\mathrm{KL}(\overline{Q}_t, \overline{\Pi}_t) < \epsilon$ is satisfied, there can exist approximation algorithms that have linear regret in T. This result provides a linear lower bound, which is complementary with the linear regret upper bound of Ensemble Sampling in (Lu and Van Roy, 2017). We can apply Theorem 3 to get sub-linear regret with Ensemble sampling with $\Theta(\log(T))$ models.

In reinforcement learning, there is a notion that certain approximations are "stochastically optimistic" and that this has implications for regret (Osband et al., 2016). This is similar in spirit to our analysis in terms of α-divergence, in that the characteristics of inference errors are important.

There has been a number of empirical works using approximate methods to perform Thompson sampling. Riquelme et al. (2018) implement variational inference, MCMC,

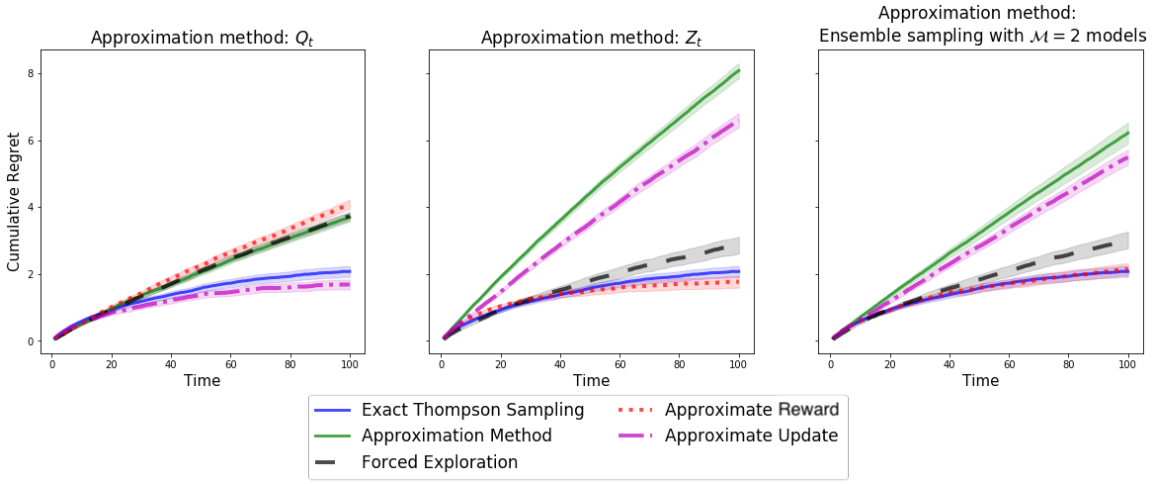
Gaussian processes and other methods on synthetic and real world data sets and measure the regret. Urteaga and Wiggins (2018) derive a variational method for contextual bandits. Kawale et al. (2015) use particle filtering to implement Thompson sampling for matrix factorization.

Finally, if exact inference is not possible, it remains an open question if it is better to use Thompson sampling with approximate inference, or to use a different bandit method that does not require inference with respect to the posterior. For example Kveton et al. (2019) propose an algorithm based on the bootstrap.

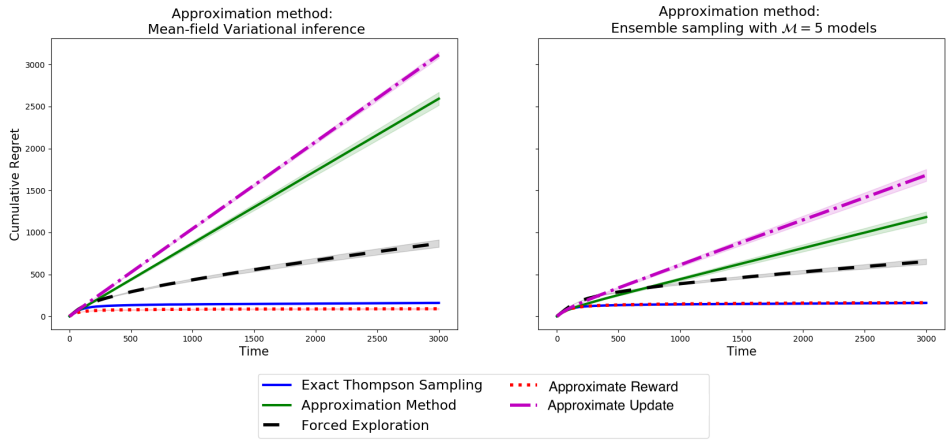
1.8 Conclusion

In this chapter we analyzed the performance of approximate Thompson sampling when at each time-step t, the algorithm obtains a sample from an approximate distribution Q_t such that the α-divergence between the true posterior and Q_t remains at most a constant ϵ at all time-steps.

Our results have the following implications. To achieve a sub-linear regret, we can only use $\alpha > 0$ for $o(T)$ time-steps. Therefore we should use $\alpha \leq 0$ with forced exploration to make the posterior concentrate. This method theoretically guarantees a sub-linear regret even when ϵ is a large constant.



(a) Applying approximations Q_t, Z_t and Ensemble Sampling to the motivating example (Section 1.6.1).



(b) Applying mean-field Variational Inference (VI) and Ensemble sampling on a 50-armed bandit (Section 1.6.2).

Figure 1.3: Updating the posterior by exact Thompson sampling or adding forced exploration does not help the over-explored approximation Q_t, but lowers the regrets of the under-explored approximations Z_t, Ensemble sampling and mean-field VI. Shaded regions show 95% confidence intervals.

CHAPTER 2

DESIGNING TRANSPORTABLE EXPERIMENTS UNDER S-ADMISSABILITY

2.1 Introduction

The chapter is based on the paper published in the International Conference on Artificial Intelligence and Statistics 2021 (Phan et al., 2021a). The problem of generalization is present everywhere that experiments are run. In the online environment, tests are run with the users who show up on the product while the experiment is running (and are therefore highly active users), while inferences about user experience are most useful on the full set of users (both highly active and less active) (Wang et al., 2019). In clinical research, it is an omnipresent problem to recruit minorities into randomized trials (Fisher and Kalbaugh, 2011), thus making it difficult to assume that the measured effects will generalize to the larger population of interest (e.g. the United States as a whole, or people afflicted with a particular health condition). In lab experiments, the sample is often one of convenience such as undergraduates in rich countries or from pools of potential subjects available online (Henrich et al., 2010). Field experiments in governance or development such as Dunning et al. (2019) are conducted in particular countries or in particular communities, but the policy implications of such work stretch far beyond the borders of the study population. As in Dunning et al. (2019), the desire is not just to understand how Burkina Faso voters respond to more information about their political leaders, but to understand how voters across the world might respond to similar informational treatments. The same is true for experiments in development economics, such as microfinance (Meager,

2019) and for studies of internet phenomena (Munger, 2018). In these cases, it isn't a *surprise* after running an experiment that generalizing the knowledge is important; indeed, generalization of knowledge to a broader population is core to the motivation for the experiment in the first place.

We pre-suppose that an experimenter knows ex-ante the population on which they wish to draw broader inferences. The task we consider, therefore, is to design an experiment that best allows the generation of causal knowledge on this inferential target. While previous work (Hartman et al., 2015; Dehejia et al., 2019; Stuart et al., 2011; DuGoff et al., 2014) has examined corrections on the analysis side to extrapolate estimates from sample to target population, the novelty of this work is in doing this through a *design-based* solution. That is, if you know your goal is to generalize to a target population, we consider how that should modify experimental design. We focus in particular on the "S-admissability" condition for transportability, in which the outcome distribution conditional on a set of covariates is the same in both the source and target distributions (Pearl and Bareinboim, 2011).

Contributions. Using the Mahalanobis distance and importance weighting, we design an estimator with a balancing condition for the target distribution's ATE that is unbiased and has low variance.

- In Section 2.3, we introduce an importance-weighted estimator with a balance condition called Target Balance that explicitly considers the target distribution in the design phase.
- In Section 2.5.1 we show that using the importance-weighted estimator with Target Balance results in an unbiased estimator of the target distribution's ATE (Theorem 4)
- We analyze the variance assuming a linear model. In Section 2.5.2.1, we show that when the dimension of the covariates $d = 1$, for a finite sample size n, Target Balance reduces the variance (Corollary 3). Moreover, among all balance criteria with rejection probability at most α (including balancing by only considering the

source distribution, which we call Source Balance), Target Balance achieves the optimal variance reduction (Theorem 5). When $d \geq 1$ (Section 2.5.2.2), when the sample size is large, Target Balance reduces the variance (Theorem 6) and achieves a lower variance than Source Balance (Theorem 7).

- In Section 2.6 we perform experiments[1] to show that Target Balance has small mean-squared errors even for $d > 1$, small sample size and non-linear model.

2.2 Problem Setting

We first fix notation before proceeding to the problem setting. Upper-case letters are used to denote random variables, lower-case letters are used to denote values taken by them. We use bold-faced letters to denote n samples and normal letters to denote a single sample. For example, $X_i \in \mathcal{R}^d$ is a random variable denoting the covariates of sample i. $\mathbf{X} = (X_1, ..., X_n)^T \in \mathcal{R}^{n \times d}$ is the random variables $X_1, .., X_n$ concatenated together. $x_i \in \mathcal{R}^d$ is a value of X_i, and $\mathbf{x} = (x_1, ..., x_n)^T \in \mathcal{R}^{n \times d}$ is a value of $\mathbf{X}$.

Some random variables, like X, can have two different distributions, either source distribution or target distribution. In that case, we use $\mathbb{E}^S X$, $\text{var}^S X$ and $cov^S X$ to denote the expectation, variance and covariance with respect to the source distribution, and $\mathbb{E}^T X$, $\text{var}^T X$ and $cov^T X$ to denote with respect to the target distribution. We use no superscripts when there is no confusion. For example $\mathbb{E}A_i$ is the expectation of the treatment assignment A_i of sample i. For a random variable R, we use $\mathbb{E}_R$, var_R and cov_R to denote the expectation, variance and covariance over the randomness of R. For example, $\mathbb{E}^S_{\mathbf{X}}$ denote the expectation over the randomness of $\mathbf{X} = (X_1, ..., X_n)^T$ according to the source distribution. We omit the subscripts when it's clear.

The problem considered in this chapter is as follows. We assume that we are presented with two populations, referred to as the source and target populations, with

[1]Code for this chapter is available at `https://github.com/myphan9/Designing_Transportable_Experiments`

corresponding densities p_S and p_T, respectively. We further assume that we observe a set of pre-treatment covariates from the source population, $x_1, \dots x_n \sim p_S$. We assume that we are freely able to assign treatment, $a_1, \dots, a_n \in \{0, 1\}$ to individuals observed in the source population and observe their outcomes, $y_1, \dots, y_n \in \mathcal{R}$. The estimand of interest is the average treatment effect for the *target* population (the population of individuals which were not subject to an experiment),

$$\tau_Y^T = \mathbb{E}^T[Y^{A=1} - Y^{A=0}]. \tag{2.1}$$

Where $Y^{A=0}$, $Y^{A=1}$ are the potential outcomes (Rubin, 2011), i.e., the values of Y that would have been observed had treatment been observed at $A = 1$ or $A = 0$, respectively. We use Y to denote (Y^0, Y^1) and Y^* to denote the *observed* outcome.

In order to make this problem tractable we will assume the following throughout the remainder of the chapter:

Assumption 3. *Equality of conditional densities, i.e., $p_S(Y|X) = p_T(Y|X)$ (note $p_S(X) \neq p_T(X)$ in general).*

This assumption places identification of the transportability of effects under the rubric of S-admissability (Pearl and Bareinboim, 2011).

Assumption 4. *Overlap between source and target distributions, i.e., $p_T(X) > 0 \implies p_S(X) > 0$.*

Assumption 5.

$$Y^1 = \psi(X)^T \beta_1 + \mathcal{E}_1 \qquad Y^0 = \psi(X)^T \beta_0 + \mathcal{E}_0$$

where ψ is a basis function and $\mathcal{E}_1, \mathcal{E}_0$ are mean zero random variables.

The linear model is a simple but fundamental model. We leave the extension to more complex models to future works. To reduce notational clutter, and without loss of generality, we will assume that ψ is the identity function for the remainder of chapter so that we can write X instead of $\psi(X)$.

Assumption 6. *The ratio of the pdfs, $p_T(X)/p_S(X)$, is known.*

In a nested trial design (Dahabreh et al., 2019), the experimenter is given a set of trial-eligible individuals (called the actual population), which is assumed to be random samples from a population (the target population). In the "census of the actual population" variant, each individual in the actual population is sampled to be in the trial with an indicator random variable S where $S = 1$ indicates that the individual is selected to be in the trial (the source population) and $S = 0$ indicates that the individual is not selected. The ratio of the pdfs can be calculated as

$$\frac{p_T(X)}{p_S(X)} = \frac{p(X)}{p(X|S=1)} = \frac{p(S=1)}{p(S=1|X)}.$$

In general the assumptions, though nontrivial, are common throughout the literature on transportability (Stuart et al., 2011; Hartman et al., 2015; Pearl and Bareinboim, 2011). We conjecture that similar results to those in this chapter will hold in the case in which importance weights are estimated with parametric convergence rates. We leave this extension as future work.

For a sample i, let X_i, A_i, Y_i^a and Y_i^* be the covariates, treatment, outcome of treatment a, and observed outcome. Let n_0 be the size of the control group (where $A_i = 0$) and n_1 be the size of the treatment group (where $A_i = 1$). Similar to common practice (c.f., Stuart et al. (2011); Hartman et al. (2015); Rudolph and van der Laan (2017); Buchanan et al. (2018)), we infer τ_Y^T with importance weights,

$$
\begin{aligned}
\hat{\tau}_Y^T &\stackrel{\text{def}}{=} \frac{1}{n_1}\sum_{i:A_i=1} W_i Y_i^* - \frac{1}{n_0}\sum_{i:A_i=0}^{n} W_i Y_i^* \\
&= \frac{1}{n_1}\sum_{i:A_i=1} W_i Y_i^1 - \frac{1}{n_0}\sum_{i:A_i=0}^{n} W_i Y_i^0 \\
&= \frac{1}{n_1}\sum_{i:A_i=1} W_i A_i Y_i^1 - \frac{1}{n_0}\sum_{i:A_i=0}^{n} W_i (1-A_i) Y_i^0 \\
&= \frac{1}{n_1}\sum_{i=1}^{n} W_i A_i Y_i^1 - \frac{1}{n_0}\sum_{i=1}^{n} W_i (1-A_i) Y_i^0 \qquad (2.2)
\end{aligned}
$$

where $W_i = \frac{p_T(X_i)}{p_S(X_i)}$. While equation 2.2 is unbiased, the estimate can incur large variance in the presence of large importance weights.

For ease of notation we define $Z_i = 2A_i - 1 \in \{-1, 1\}$ and let $\mathbf{Z}$ be the $n \times 1$ vector of random variables $Z_1, ..., Z_n$ and $\mathbf{z}$ be a value taken by $\mathbf{Z}$. $Y_i = (Y_i^0, Y_i^1)$ is a random variable denoting all possible outcomes of sample i. $\mathbf{Y} = (Y_1, ..., Y_n)^T \in \mathcal{R}^{n\times 2}$ is the random variables $Y_1, .., Y_n$ concatenated together. $y_i \in \mathcal{R}^2$ is a value of Y_i, and $\mathbf{y} = (y_1, ..., y_n)^T \in \mathcal{R}^{n\times 2}$ is a value of $\mathbf{Y}$. Let $w_i = \frac{p_T(x_i)}{p_S(x_i)}$ and $\mathbf{w}$ be the $n \times n$ diagonal matrix with $\mathbf{w}(i,i) = w_i$. For matrix $\mathbf{a}$, we use $\tilde{\mathbf{a}}$ to denote $\mathbf{wa}$ where each row i of $\mathbf{a}$ is multiplied by w_i.

2.3 Designing for Transportation

We consider $n_0 = n_1 = n/2$ throughout the chapter. The core contribution of this work is a procedure to estimate equation 2.1 which explicitly considers the target population when designing the experiment for the source population. We focus on adapting re-randomization, an experimental design procedure which optimizes balance, i.e., the difference in means of X between treatment and control groups. Specifically, the experimenter specifies a *balance criterion* and then repeatedly randomly assign individuals to treatment and control group until the balance criterion is satisfied. A rejection threshold is the probability that the assignment is rejected. We define the following balance criterion.

Definition 1 (Target Balance)**.** *With a rejection threshold α, define the balance condition:*

$$\phi_T^\alpha(\mathbf{x}, \mathbf{Z}) = \begin{cases} 1, & \text{if } M(\frac{2}{n}(\mathbf{wx})^T\mathbf{Z}) < a(\mathbf{x}) \\ 0, & \text{otherwise} \end{cases}$$

where $M(\frac{2}{n}(\mathbf{wx})^T\mathbf{Z})$ is a distance (defined below in Eq. 2.3) between the covariates associated with treatment and control given by $\mathbf{Z}$ and $a(\mathbf{x})$ is chosen such that $\mathbb{P}(\phi_T^\alpha = 1|\mathbf{x}) = 1 - \alpha$.

We omit α and simply write ϕ_T when α is not necessary for exposition. We omit $\mathbf{x}$ and write a when there is no confusion. We also reuse the notation and use $\phi_T^{a(\mathbf{x})}$ instead of ϕ_T^α (recalled that $a(\mathbf{x})$ is chosen such that $\mathbb{P}(\phi_T^\alpha = 1|\mathbf{x}) = 1 - \alpha$).

The full assignment procedure is then

1. Assign A randomly for each person $1, \ldots, n$ in the source population such that $\sum_i a_i = n_1$. There are $\binom{n}{n_1}$ ways to choose this, each of which is equally likely.
2. If $\phi_T(\mathbf{x}, \mathbf{z}) = 0$ return to step (1).
3. Conduct experiment with treatment assignments, A.

Following standard practice in rerandomization (Morgan et al., 2012), we will focus on a criterion based on Mahalanobis distance, but incorporating a weighting term to express our desire for balance in the *target* distribution rather than in the source. We refer to this weighted Mahalanobis distance as $M(\frac{2}{n}(\mathbf{wx})^T\mathbf{Z})$, where $M(\cdot)$ is defined as:

$$\begin{aligned} M(U) &\stackrel{\text{def}}{=} (U)^T Cov(U)^{-1} (U) \\ &= \|B\|^2 \text{ where } B = UCov(U)^{-1/2}. \end{aligned} \tag{2.3}$$

Thus, the balance condition $M(\frac{2}{n}(\mathbf{x})^T\mathbf{Z}) < a^2$ is equivalent to truncating the square norm of B to be less than a. Note that this is a standardized measure of the difference in importance-weighted covariate-means between treatment and control, since

$$\frac{2}{n}\sum_{i:Z_i=1} w_i x_i - \frac{2}{n}\sum_{i:Z_i=-1} w_i x_i = \frac{2}{n}(\mathbf{wx})^T\mathbf{Z}.$$

Thus, rerandomization simply rejects designs with covariate imbalance larger than a pre-specified value.

The novelty in our proposed design is to reject samples based on imbalance in the the *target* distribution rather than based on imbalance in the *source* distribution. The standard in the rerandomization literature is to focus on balance in the *source* distribution, which in our setup implies assuming that the target distribution is equal to the source distribution. Therefore, importance weights in this case are all equal to one. We call this balancing condition *Source Balance*, which we denote by $\phi_S^\alpha(\mathbf{x}, \mathbf{Z})$.

We explain the intuition behind using Target Balance rather than Source Balance. An unbiased estimator for the source's ATE is: $\hat{\tau}_Y^S \stackrel{\text{def}}{=} \frac{1}{n_1}\sum_i A_i Y_i^1 - \frac{1}{n_0}\sum_i A_i Y_i^0$. There are existing results (Li et al., 2018; Harshaw et al., 2019) that can be applied to linear models to show variance reduction of $\hat{\tau}_Y^S$ with Source Balance defined by $\mathbf{x}$.

By defining new variables $\tilde{Y}_i^a = W_i \cdot Y_i^a$ for $a \in \{0, 1\}$, the importance weighted estimator can now be expressed in a similar form to $\hat{\tau}_Y^S$: $\hat{\tau}_Y^T = \frac{1}{n_1}\sum_i A_i \tilde{Y}_i^1 - \frac{1}{n_0}\sum_i A_i \tilde{Y}_i^0$.

We are no longer in a linear setting because $\tilde{Y}_i^a = W_i(\beta_a^T X_i + \mathcal{E}_a)$. But by defining $\tilde{X}_i = W_i X_i$ and $\tilde{\mathcal{E}} = W_i\mathcal{E}$ we have: $\tilde{Y}_i^a = \beta_a^T \tilde{X}_i + \tilde{\mathcal{E}}_a$, where $\mathbb{E}\left[\tilde{Y}^a | X\right]$ is a linear function of $\tilde{X}$, which can be considered as a feature-transformed X. (Li et al., 2018; Harshaw et al., 2019) can now be applied to estimate $\hat{\tau}_Y^T$ with Target Balance defined by $\tilde{\mathbf{x}} = \mathbf{wx}$.

---

[2]We show later in Lemma 34 and Lemma 35 that $Cov(\mathbf{x}^T\mathbf{Z})) = \frac{n}{n-1}\mathbf{x}^T Q\mathbf{x}$ where $Q = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T$. Eaton and Perlman (1973) discuss conditions for which $\mathbf{X}^T Q\mathbf{X}$ is non-singular with probability 1.

Let $\rho(\mathbf{x}, \mathbf{z}) \in \{0, 1\}$ be a function of $\mathbf{x}$ and $\mathbf{z}$ used in the re-randomization procedure. Note that since we re-sample $\mathbf{Z}$ until $\rho = 1$, only the distribution of $\mathbf{Z}$ is affected by the balance condition. Let $\mathbf{Z}_\rho$ denote the distribution of $\mathbf{Z}$ after being accepted by the balance condition $\rho = 1$.

2.4 Related Work

Our work relates to and ties together two distinct strands of research: (1) ex-post generalization of experimental results to population average effects and (2) ex-ante experimental design. We will discuss each in turn.

Generalization.

Within the literature on methods for generalization, work has generally focused on ex-post adjustments to experiments previously run.

The foundational work of Stuart et al. (2011) provides an approach based on propensity scores for generalizing the results of experimental interventions to target populations. Our work will leverage this general framework, but introduce methods for optimizing an experimental design to ensure effective generalization performance of resulting estimates. Hartman et al. (2015) similarly uses a combination of matching and weighting to generalize experimental results in-sample to a population average treatment effect on the treated. Other work has also considered weighting-based approaches to generalization (Buchanan et al., 2018).

Dehejia et al. (2019) shows how to use an outcome-modeling approach to extrapolate effects estimated in one population to a population. In contrast to Hartman et al. (2015) and Stuart et al. (2011), this approach relies on modeling the outcomes and then predicting effects in different locations rather than simply reweighting data observed in-sample.

Dahabreh et al. (2018) provides a variety of estimation methods to generalize to a target population, including doubly-robust methods. Rudolph and van der Laan

(2017), likewise, provides a doubly-robust targeted maximum likelihood estimator for transporting effects.

There has also been work focused particularly on identification in this setting. Dahabreh et al. (2019) defines a rigorous sampling framework for describing generalizability of experimental results and identifiability conditions through the g-formula. Pearl and Bareinboim (2011) lays out a general framework for determining identifiability of effects generalized to new populations through.

Miratrix et al. (2018) and Coppock et al. (2018) challenge the premise of the necessity for generalization due to the rarity of heterogeneous treatment effects. These studies specifically focused on survey experiments, however, and it isn't truly up for debate that many important objects of study have important heterogeneous components (Allcott, 2015; Vivalt, 2015; Dehejia et al., 2019).

Experimental design.

The standard practice for experimental design is blocking (Greevy et al., 2004), in which units are divided into clusters and then a fixed number of units within each cluster are assigned to treatment. This ensures balance on the cluster indicators within the sample. Higgins et al. (2016) provides a blocking scheme based on k-nearest-neighbors that can be calculated more efficiently than the "optimal" blocking of (Greevy et al., 2004).

Kallus (2018) takes an optimization approach to the problem of experimental design. This work optimizes treatment allocations based on in-sample measures of balance (particularly with respect to kernel means), showing how assumptions of smoothness are necessary to improve on simple Bernoulli randomization.

Rerandomization approaches simply draw allocations randomly until one is located which meets the pre-specified balance criteria. This is also the basis of our proposed method. Morgan et al. (2012) analyzes the rerandomization procedure of discarding randomized assignments that have more in-sample imbalance than a pre-specified

criteria in terms of Mahalanobis distance. Li et al. (2018) provides asymptotic results for rerandomization that does not rely on distributional assumptions on the covariates.

Harshaw et al. (2019) provides an efficient method for obtaining linear balance using a Gram-Schmidt walk. Their algorithm includes a robustness-balance tradeoff tuneable by a parameter in their algorithm, and provides useful tools for analyzing experimental design which we use in our theoretical analyses in Section 2.5.

All aforementioned work on experimental design places as its objective estimation of effects on the sample (i.e. it optimizes for the sample average treatment effect). This work departs by considering the alternative objective of prioritizing estimation on a target population (i.e. the population average treatment effect).

2.5 Analysis

In this section we will analyze the expectation and variance of our importance-weighted estimator in Eq. 2.2 with Target Balance in Definition 1.

Section 2.5.1 shows that using the importance-weighted estimator with Target Balance results in an unbiased estimator of the target's ATE (Theorem 4).

In Section 2.5.2 we analyze the variance. In Section 2.5.2.1, Corollary 3 shows that when the dimension of the covariates $d = 1$, for a finite sample size n, Target Balance reduces the variance. Moreover, among all reasonable balance criteria with rejection probability at most α (including Source Balance), Target Balance achieves the optimal variance reduction (Theorem 5). Section 2.5.2.2 shows that when $d \geq 1$, when the sample size is large, Target Balance reduces the variance (Theorem 6) and achieves a lower variance than Source Balance (Theorem 7).

2.5.1 Expectation

In this section we will show that our importance-weighted estimator in Eq. 2.2 is an unbiased estimator of the target's ATE with Target Balance:

Theorem 4. *Let $\hat{\tau}_Y^T$ be the importance-weighted estimator in Equation 2.2. When $n_0 = n_1 = n/2$:* $\mathbb{E}^S_{\mathbf{X},\mathbf{Y},\mathbf{Z}_{\phi_T}}\left[\hat{\tau}_Y^T\right] = \tau_Y^T$.

The proof makes use of the fact that the conditional distributions of Y given X in both the source and the target are the same ($p_S(Y|X) = p_T(Y|X)$), and therefore $\frac{p_T(X)}{p_S(X)} = \frac{p_T(X,Y)}{p_S(X,Y)}$.

2.5.2 Variance

In this section we analyze the variance. We use $\tilde{Y}_i^a$ and $\tilde{y}_i^a$ to denote $W_i Y_i^a$ and $w_i y_i^a$ for $a \in \{0, 1\}$.

2.5.2.1 Finite Sample Size Variance Reduction for $d = 1$

In this section we will show that when X is a 1-dimensional random variable and the sample size is finite, Target Balance reduces the variance compared to complete randomization. Moreover, among all symmetric balance conditions (defined below) with rejection probability at most α (including Source Balance), Target Balance achieves the optimal variance reduction. The variance can be decomposed into 2 terms (Lemma 5) where the second term does not depend on the balance. The first term is the variance of a 1d symmetric random variable, and Target Balance corresponds to truncating the tail, which results in the largest variance reduction (Theorem 5).

Let $\rho(\mathbf{x}, \mathbf{Z}) \in \{0, 1\}$ denote a function that depends on only $\mathbf{x}$ and $\mathbf{Z}$, and satisfies the symmetric condition $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$. This definition captures all reasonable balance conditions (including Source Balance) where $\rho = 1$ denotes acceptance and $\rho = 0$ denotes rejection. Note that the constant function $\rho(\mathbf{x}, \mathbf{Z}) = 1$ for all $\mathbf{x}, \mathbf{Z}$ also satisfies the criteria $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$, and $\rho = 1$ becomes the entire sample space. We proceed to compare Target Balance with any ρ satisfying the criteria above.

First we note that by the law of total variance:

Lemma 3. *For any function* $\rho(\mathbf{x}, \mathbf{Z}) \in \{0, 1\}$ *satisfying* $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$*:*

$$var^S_{\mathbf{X},\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y) = \mathbb{E}^S_{\mathbf{X}}\left[var^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{X})\right] + var^S_{\mathbf{X}}(\frac{1}{n}\sum_{i=1}^{n} W_i(\beta_1 - \beta_0)^T X_i)\ .$$

Note that the second term does not depend on ρ. Therefore we focus on analyzing the variance conditioned on $\mathbf{X} = \mathbf{x}$ in this section, and the result for $\text{var}^S_{\mathbf{X},\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y)$ easily follows from $\text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x})$.

Let $C_i = \frac{Y_i^1 + Y_i^0}{2}$, $c_i = \frac{y_i^1 + y_i^0}{2}$, $\beta = \frac{\beta_1 + \beta_0}{2}$, $\mathcal{E} = \frac{\mathcal{E}_1 + \mathcal{E}_0}{2}$ and $\sigma^2_{\mathcal{E}} = \text{var}(\mathcal{E})$. The variance of the importance weighted estimator can be written as

Lemma 4. *Let* $n_0 = n_1 = n/2$*. For any function* $\rho(\mathbf{x}, \mathbf{Z}) \in \{0, 1\}$ *satisfying* $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$*:*

$$var_{\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}, \mathbf{y}) = \frac{4}{n^2}\mathbb{E}_{\mathbf{Z}_\rho}\left[\left(\sum_{i=1}^{n} Z_i w_i c_i\right)^2 \middle| \mathbf{x}, \mathbf{y}\right].$$

Using the law of total variance and the fact that $W_i C_i = W_i X_i \beta + W_i \mathcal{E}$ and $\mathbb{E}[\mathcal{E}|\mathbf{x}] = 0$ we have:

Lemma 5. *Let* $n_0 = n_1 = n/2$*. For any function* $\rho(\mathbf{x}, \mathbf{Z}) \in \{0, 1\}$ *satisfying* $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$*:*

$$var^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) = \frac{4}{n^2}\beta^2\mathbb{E}_{\mathbf{Z}_\rho}\left[\left(\sum_{i=1}^{n} w_i x_i Z_i\right)^2 \middle| \mathbf{x}\right] + \frac{6}{n^2}\sigma^2_{\mathcal{E}}\sum_{i=1}^{n} w_i^2.$$

We note that the design affects only the first term in the above decomposition. Let $V \stackrel{\text{def}}{=} \frac{2}{n}\sum_i Z_i w_i x_i = \frac{2}{n}\tilde{\mathbf{x}}^T\mathbf{Z}$ and let $B := V\text{var}(V)^{-1/2}$. Recall that the Malahanobis distance $M(\frac{2}{n}(\mathbf{wx})^T\mathbf{Z}) = ||B||^2$. Re-randomization procedure corresponds to truncating B where B is a mean zero random variable (as Z_i's are random variables) that is symmetric about zero.

It is easy to show that the best way to truncate a symmetric random variable B to minimize the variance is to truncate the tail symmetrically $\|B\|^2 < a$ for some threshold a. Therefore Target Balance reduces the variance, and among all the balance conditions with rejection probability at most α (including Source Balance), Target Balance achieves the optimal variance reduction.

Theorem 5. *Let $n_0 = n_1 = n/2$ and $d = 1$. Let $\rho(\mathbf{x}, \mathbf{Z})$ be a function satisfying $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$ and $\mathbb{P}(\rho = 1|\mathbf{x}) \geq 1 - \alpha$. Then:*

$$var^S_{\mathbf{Y},\mathbf{Z}_{\phi^\alpha_T}}(\hat{\tau}^T_Y|\mathbf{x}) \leq var^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}).$$

Applying Theorem 5 with ρ being the constant function $\rho(\mathbf{x}, \mathbf{Z}) = 1$ for all $\mathbf{x}, \mathbf{Z}$, we have:

Corollary 3. *When $d = 1$ and $n_0 = n_1 = n/2$, using Target Balance reduces the variance compared to complete randomization:*

$$var^S_{\mathbf{Y},\mathbf{Z}_{\phi_T}}(\hat{\tau}^T_Y|\mathbf{x}) \leq var^S_{\mathbf{Y},\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x}).$$

2.5.2.2 Asymptotic Variance Reduction for $d \geq 1$

In this section we show that when the sample size is large, Target Balance reduces the variance and achieves a lower variance than Source Balance. We discuss the case of finite sample size in the appendix.

From Li et al. (2018), the importance weighted estimator can be decomposed into 2 components: part 1 is related to the covariates and part 2 is unrelated. Only part 1 is reduced by rerandomization while part 2 is unaffected. The covariates can be chosen to be the importance-weighted covariates (Target Balance) or the unweighted covariates (Source Balance). Since the importance-weighted covariates aligns better with the

importance-weighted outcomes, part 1 will be larger and therefore the reduction by re-randomization will be larger.

In this section we condition on $\mathbf{x}$ and $\mathbf{y}$ so the randomness only comes from $\mathbf{Z}$. Similar to Section 2.5.2.1, first we note that by the law of total variance:

Lemma 6. *For any function $\rho(\mathbf{x}, \mathbf{Z}) \in \{0, 1\}$ satisfying $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$:*

$$var^S_{\mathbf{X},\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y) = \mathbb{E}^S_{\mathbf{X},\mathbf{Y}} var_{\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{X}, \mathbf{Y}) + var^S_{\mathbf{X},\mathbf{Y}}(\sum_{i=1}^{n} W_i(Y_i^1 - Y_i^0)).$$

Since the second term does not depend on ρ, we focus on analyzing the variance conditioned on $\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}$ in this subsection. The result for $\text{var}^S_{\mathbf{X},\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y)$ follows from $\text{var}_{\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{X}, \mathbf{Y})$.

Conditioning on $\mathbf{x}$ and $\mathbf{y}$, Li et al. (2018) state that if the following conditions (Condition 1 in Li et al. (2018)) are satisfied, finite central limit theorem implies that $(\hat{\tau}^T_Y, \frac{2}{n}\tilde{\mathbf{x}}^T\mathbf{Z})$ approaches a normal distribution as n goes to infinity. Let $\text{avg}(\tilde{\mathbf{y}}^a), \text{avg}(\mathbf{x})$ and $\text{avg}(\tilde{\mathbf{x}})$ denote the average of the rows of $\tilde{\mathbf{y}}^a$ for $a \in \{0, 1\}, \mathbf{x}$ and $\tilde{\mathbf{x}}$ and let $\tilde{\tau}_i \stackrel{\text{def}}{=} \tilde{y}^1_i - \tilde{y}^0_i$.

Assumption 7. *As $n \to \infty$:*

- *The finite population variances and covariance $cov(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=1}^n (x_i - avg(\mathbf{x}))(x_i - avg(\mathbf{x}))^T/(n-1), cov(\tilde{\mathbf{x}}) \stackrel{\text{def}}{=} \sum_{i=1}^n (\tilde{x}_i - avg(\tilde{\mathbf{x}}))(\tilde{x}_i - avg(\tilde{\mathbf{x}}))^T/(n-1), cov(\tilde{\boldsymbol{\tau}}) \stackrel{\text{def}}{=} \sum_{i=1}^n (\tilde{\tau}_i - avg(\tilde{\boldsymbol{\tau}}))(\tilde{\tau}_i - avg(\tilde{\boldsymbol{\tau}}))^T/(n-1), cov(\tilde{\mathbf{y}}^a) \stackrel{\text{def}}{=} \sum_{i=1}^n (\tilde{y}^a - avg(\tilde{\mathbf{y}}^a))(\tilde{y}^a - avg(\tilde{\mathbf{y}}^a))^T/(n-1), cov(\tilde{y}^a, \mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=1}^n (\tilde{y}^a - avg(\tilde{\mathbf{y}}^a))(x_i - avg(\mathbf{x}))^T/(n-1), cov(\tilde{y}^a, \tilde{\mathbf{x}}) \stackrel{\text{def}}{=} \sum_{i=1}^n (\tilde{y}^a - avg(\tilde{\mathbf{y}}^a))(\tilde{x}_i - avg(\tilde{\mathbf{x}}))^T/(n-1)$ for treatment index $a \in \{0, 1\}$ have limiting values, and the limit of $cov(\mathbf{x})$ and $cov(\tilde{\mathbf{x}})$ is non-singular.*
- $\max_{1 \le i \le n} |\tilde{y}^a_i - avg(\tilde{\mathbf{y}})^a|^2/n \to 0$ *for* $a \in \{0, 1\}$, $\max_{1 \le i \le n} \|x_i - avg(\mathbf{x})\|_2^2/n \to 0$ *and* $\max_{1 \le i \le n} \|\tilde{x}_i - avg(\tilde{\mathbf{x}})\|_2^2/n \to 0$

We show in Lemma 35 that Assumption 7 is satisfied in our case where all variables are sampled i.i.d from distributions when Assumption 8 is satisfied. Therefore we assume Assumption 8 for the remaining of the section.

Assumption 8. *For a random variable $X \in R^d$, let $X(k)$ denote the k-th coordinate of X for $1 \leq k \leq d$. We assume that $Y^a, a \in \{0,1\}$, $X(j), 1 \leq j \leq d$ and $W = \frac{p_T(X)}{p_S(X)}$ have finite 8th moment according to the source distribution, and $Cov^S(\tilde{X})$ and $Cov^S(\tilde{X}) = Cov^S(\frac{p_T(X)}{p_S(X)}X)$ is non-singular.*

We apply Corollary 2 in Li et al. (2018) to give the expression for the asymptotic variance of $\hat{\tau}_Y^T$ under Mahalanobis balance condition. Let as-var denote the variance of the asymptotic sampling distribution of a sequence of random variables. Applying Corollary 2 in Li et al. (2018) to our case with covariates $\tilde{\mathbf{x}}$ and $\mathbf{x}$ and the weighted outcome $\tilde{y}$ directly yields the following result showing both Target Balance and Source Balance reduce the variance.

Theorem 6 (Shown in the proof of Corollary 2 in (Li et al., 2018))**.** *Suppose $n_0 = n_1 = n/2$. Let $\tau_{\mathbf{x},\mathbf{y}} \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{Z}}[\hat{\tau}_Y^T|\mathbf{x},\mathbf{y}]$. When Assumption 7 is satisfied, conditioning on $\mathbf{x}$ and $\mathbf{y}$:*

$$
\begin{aligned}
&\textit{as-var}_{\mathbf{Z}_{\phi_S}}\left(\sqrt{n}(\hat{\tau}_Y^T - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{x},\mathbf{y}\right) \\
&= \textit{as-var}_{\mathbf{Z}}(\sqrt{n}(\hat{\tau}_Y^T - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{x},\mathbf{y})(1-(1-v_{d,a})\lim_{n\to\infty} R^2_{\mathbf{x}}), \\
&\textit{as-var}_{\mathbf{Z}_{\phi_T}}\left(\sqrt{n}(\hat{\tau}_Y^T - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{x},\mathbf{y}\right) \\
&= \textit{as-var}_{\mathbf{Z}}(\sqrt{n}(\hat{\tau}_Y^T - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{x},\mathbf{y})(1-(1-v_{d,a})\lim_{n\to\infty} R^2_{\tilde{\mathbf{x}}}),
\end{aligned}
$$

where $R^2_{\tilde{\mathbf{x}}} = Corr(\hat{\tau}_Y^T, \frac{2}{n}\tilde{\mathbf{x}}^T\mathbf{Z})$, $R^2_{\mathbf{x}} = Corr(\hat{\tau}_Y^T, \frac{2}{n}\mathbf{x}^T\mathbf{Z})$ and $v_{d,a} = \frac{P(\chi^2_{d+2}\leq a)}{P(\chi^2_d \leq a)}$.

We now show that Target Balance has a smaller variance than Source Balance. We use the following equivalent expressions for $R^2_{\mathbf{x}}$ and $R^2_{\tilde{\mathbf{x}}}$. Let $Q = \frac{n}{n-1}\left(\mathbf{I}_d - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$

where $\mathbf{I}_d$ is an identity matrix of dimension d. Recall that $c_i = \frac{y_i^0 + y_i^1}{2}$. Let $\mathbf{c} \stackrel{\text{def}}{=} (c_1, \cdots, c_n)$ and $\tilde{\mathbf{c}} = \mathbf{wc}$. We will show that:

Lemma 7. *When $n_0 = n_1 = n/2$:*

$$R^2_{\mathbf{x}} = \sqrt{\frac{\|Q\tilde{\mathbf{c}}\|^2 - \min_{\hat{\beta}} \|Q\tilde{\mathbf{c}} - Q\mathbf{x}\hat{\beta}\|^2}{\|Q\tilde{\mathbf{c}}\|^2}}.$$
$$R^2_{\tilde{\mathbf{x}}} = \sqrt{\frac{\|Q\tilde{\mathbf{c}}\|^2 - \min_{\hat{\beta}} \|Q\tilde{\mathbf{c}} - Q\tilde{\mathbf{x}}\hat{\beta}\|^2}{\|Q\tilde{\mathbf{c}}\|^2}}.$$

Intuitively $R^2_{\tilde{\mathbf{x}}}$ and $R^2_{\mathbf{x}}$ describe how well $\tilde{\mathbf{c}}$ is described by a linear function of $\tilde{\mathbf{x}}$ and $\mathbf{x}$, respectively. Because of our model, a linear model in terms of $\tilde{\mathbf{x}} = \mathbf{wx}$ fits $\tilde{\mathbf{c}} = \mathbf{wc}$ better than a linear model in terms of $\mathbf{x}$. Therefore, $R^2_{\tilde{\mathbf{x}}}$ will be larger than $R^2_{\mathbf{x}}$ and using ϕ_T will result in a smaller variance than ϕ_S.

Therefore, asymptotically, with the same rejection probability, using Target Balance results in a smaller variance than Source Balance .

Theorem 7. *Suppose $n_0 = n_1 = n/2$ and Assumption 8 is true. Recall that $\mathbb{P}(\phi_S^a = 0|\mathbf{x})$ is the rejection probability when using threshold a in Source Balance, and $\mathbb{P}(\phi_T^a = 0|\mathbf{x})$ is the the rejection probability when using threshold a in Target Balance. With the same rejection threshold a:*

$$\lim_{n\to\infty} \mathbb{P}(\phi_S^a = 0|\mathbf{x}) = \lim_{n\to\infty} \mathbb{P}(\phi_T^a = 0|\mathbf{x}) \text{ for any } \mathbf{x},$$
$$\text{as-var}_{\mathbf{Z}_{\phi_T^a}}\left(\hat{\tau}_Y^T|\mathbf{X},\mathbf{Y}\right) \leq \text{as-var}_{\mathbf{Z}_{\phi_S^a}}\left(\hat{\tau}_Y^T|\mathbf{X},\mathbf{Y}\right) \text{ almost surely.}$$

2.6 Simulations

We perform simulations on the two following models:

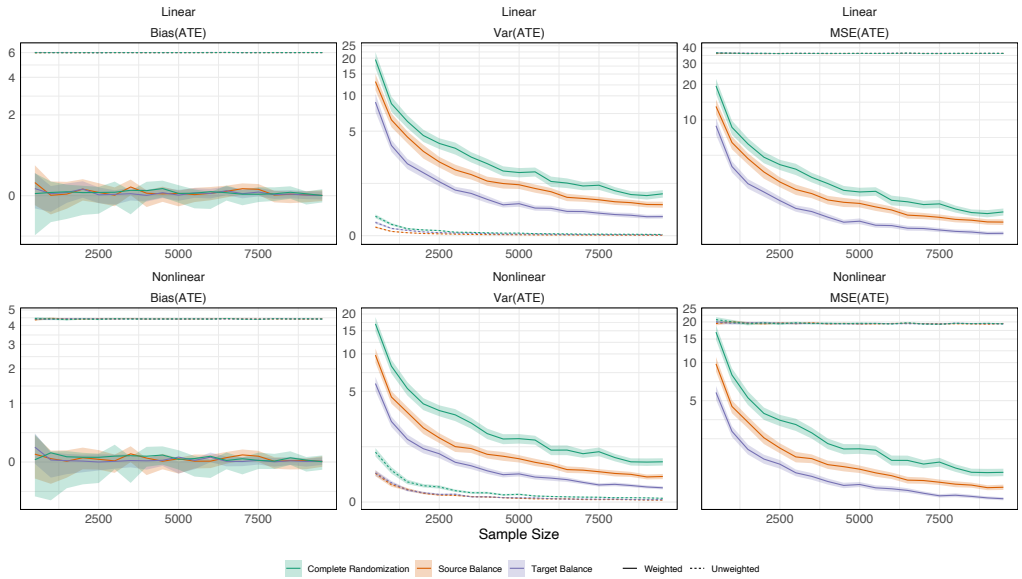


Figure 2.1: Bias, Variance and MSE as a function of the sample size. All unweighted estimators are biases because they measure the ATE of the source distribution. As there is no importance weight threshold, all weighted estimators are unbiased (Theorem 4) but the weighted estimator with Target Balance has the lowest variance. *The y axes are in log scale.*

Linear Model

$$Y^0 = X + Norm(0,1); Y^1 = 3X + Norm(0,1)$$

Nonlinear Model

$$Y^0 = X^T X + Norm(0,1); Y^1 = 2X^T X + Norm(0,1)$$

We use the following source and target distributions for X. In the source distribution, $X \sim \text{MultivariateNorm}(\mathbf{1}, \mathbf{I})$ where $\mathbf{I}$ is the identity matrix. In the target distribution, $X \sim \text{MultivariateNorm}(\mathbf{1} + \delta, \mathbf{I})$ where δ is a parameter that will be specified later.

We randomly choose an assignment such that $n_1 = n_0 = n/2$. To select the random assignment with the top balance, instead of choosing a fixed threshold α, we select the rejection probability $\alpha = 0.99$ as in Def. 1. To implement this, we draw $100/(1-\alpha)$

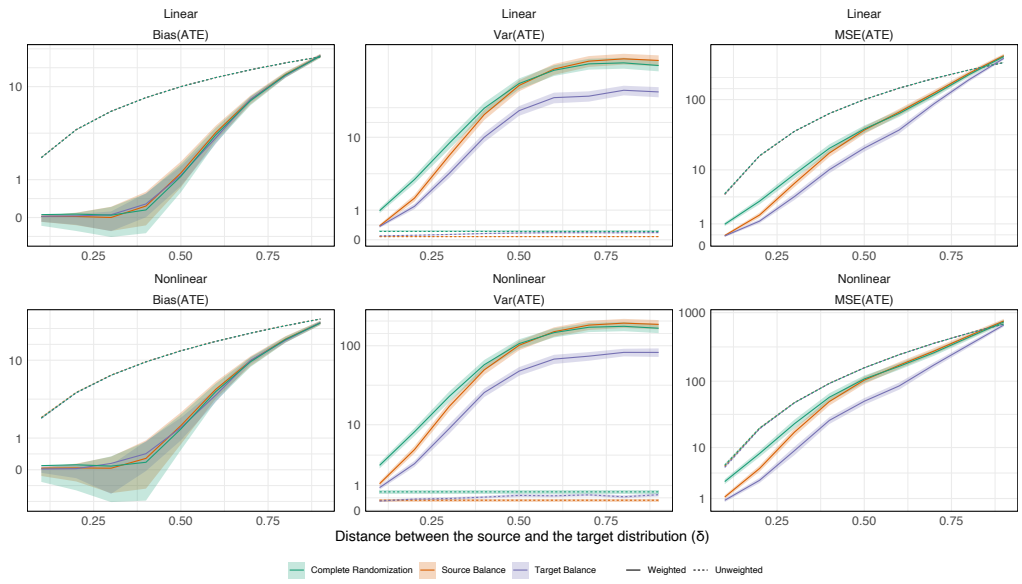


Figure 2.2: Bias, Variance and MSE as a function of the distance δ (defined in Section 2.6) between the source and the target distribution. Because of the importance weight threshold, the biases of the importance weighted methods increase as the δ increase. If the distance is too large, the bias of the importance weighted estimators is large, leading to high MSE. However when the distance is not too large, the weighted estimator with Target Balance has the lowest MSE. *The y axes are in log scale.*

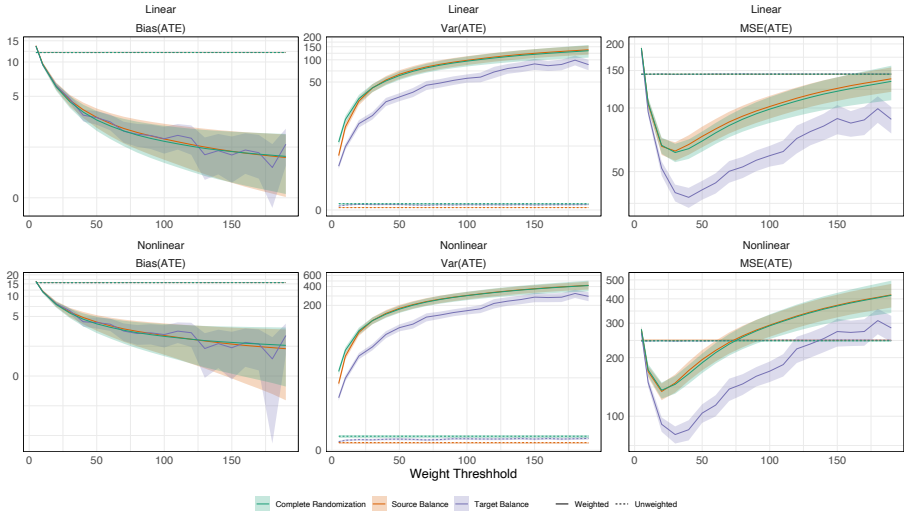


Figure 2.3: Bias, Variance and MSE as a function of importance weight threshold. As the threshold increases, the bias of the weighted methods decreases and the variance of the weighted methods increases. Therefore there is a threshold when the MSE is minimized. The weighted estimator with Target Balance has the lowest MSE for a reasonably good threshold. *The y axes are in log scale.*

assignments at random, calculate their Mahalanobis distance and pick one among the smallest 100 uniformly at random.

If the source and the target distributions are far away, importance weighting can induce large variance. We use the weight clipping technique, in which if the importance weight is larger than a threshold, it will be set to that threshold. It will induce bias but reduce variance, and therefore reduce mean square error (MSE).

We compare 6 methods (WE, CR), (WE, SB), (WE, TB), (UE, CR), (UE, SB) and (UE, TB) by combining the following 2 properties:

Weighted and Unweighted.

- *Weighted Estimator (WE).* We consider the importance weighted estimator in Eq. 2.2.
- *Unweighted Estimator (UE).* We consider the unweighted estimator which is equivalent to Eq.2.2 with all weights set to one.

Complete Randomization, Source Balance and Target Balance.

- *Complete Randomization (CR).* This is the randomized assignment without balancing.
- *Source Balance (SB).* This is the rerandomization algorithm seeking Source Balance.
- *Target Balance (TB).* This is the rerandomization algorithm seeking Target Balance as in Definition 1.

We study the MSE of our methods in relation to the 3 following parameters: the sample size n, the importance weights threshold and the distance δ. Recall that in the source distribution, $X \sim \text{MultivariateNorm}(\mathbf{1}, \mathbf{I})$ where $\mathbf{I}$ is the identity matrix and in the target distribution, $X \sim \text{MultivariateNorm}(\mathbf{1} + \delta, \mathbf{I})$.

Sample Size. In this experiment for both models we vary the sample size from 500 to 9500 with step size 500 and set the number of covariates to 10. For the linear model, $\delta = 0.3$. For the nonlinear model, $\delta = 0.2$. δ is chosen to be small enough so that we do not need weight clipping. For each sample size we repeat the experiment

500 times. There is no importance weight threshold. The results are shown and discussed in Figure 2.1.

Threshold. In this experiment for both models we vary the importance weight threshold from 5, then 10 to 190 with step size 10. We set the number of covariates to 10 and the sample size to be 1000 and $\delta = 0.6$. δ is chosen to be large enough so that weight clipping is necessary. For each threshold we repeat the experiment 500 times. The results are shown and discussed in Figure 2.3.

Distance δ. In this experiment for both models we vary δ from 0.1 to 0.9 with step size 0.1. We set the number of covariates to 10, the sample size to be 1000 and the importance weight threshold to be 40. From the weight threshold experiment, we know that if the weight threshold is too large, the variance is too high while if the weight threshold is too small, the bias will be too high. Therefore we pick the value 40 as a reasonable weight threshold. For each threshold we repeat the experiment 500 times. The results are shown and discussed in Figure 2.2. Across all simulations, Target Balance with the Weighted Estimator substantially reduces the MSE.

2.7 Conclusion

In this work, we've shown that a desire for generalizability should change the way experiments are designed and run. In particular, we argue that balance should be sought on the target population rather than the samples in which randomization will actually be performed. We present a method for designing an experiment along these lines, show theoretically that it is unbiased and more efficient than sample balancing.

CHAPTER 3

TOWARDS PRACTICAL MEAN BOUNDS FOR SMALL SAMPLES

3.1 Introduction

The chapter is based on the paper published in the International Conference on Machine Learning 2021 (Phan et al., 2021b). In this work, we revisit the classic statistical problem of defining a confidence interval on the mean μ of an unknown distribution with CDF F from an i.i.d. sample $\mathbf{X} = X_1, X_2, \ldots, X_n$, and the closely related problems of producing upper confidence bounds on the mean.

To produce a non-trivial UCB, one must make assumptions about F, such as finite variance, sub-Gaussianity, or that its support is contained on a known interval $[a, b]$. We adopt this last assumption, working with distributions whose support is known to fall in an interval $[a, b]$. For UCBs, we refer to two separate settings, the *one-ended support* setting, in which the distribution is known to fall in the interval $[-\infty, b]$, and the *two-ended support* setting, in which the distribution is known to fall in an interval $[a, b]$, where $a > -\infty$ and $b < \infty$.

A UCB has *guaranteed coverage* for a set of distributions $\mathcal{F}$ if, for all sample sizes $1 \leq n \leq \infty$, for all confidence levels $1 - \alpha \in (0, 1)$, and for all distributions $F \in \mathcal{F}$, the bound $\mu^{1-\alpha}_{\texttt{upper}}$ satisfies

$$Prob_F[\mu \leq \mu^{1-\alpha}_{\texttt{upper}}(X_1, X_2, ..., X_n)] \geq 1 - \alpha, \tag{3.1}$$

where μ is the mean of the unknown distribution F.

Among bounds with guaranteed coverage for distributions on an interval $[a, b]$, our interest is in bounds with good performance on *small sample sizes*. The reason is that, for 'large enough' sample sizes, excellent bounds and confidence intervals already exist. In particular, the confidence intervals based on Student's $t-$statistic (Student, 1908) are satisfactory in terms of coverage and accuracy for most practitioners, given that the sample size is greater than some threshold.[1]

The validity of the Student's t method depends upon the Gaussianity of the sample mean, which, strictly speaking does not hold for any finite sample size unless the original distribution itself is Gaussian. However, for many applications, the sample mean becomes close enough to Gaussian as the sample size grows (due to the effects described by the central limit theorem), that the resulting bounds hold with probabilities close to the confidence level. Such results vary depending upon the unknown distribution, but it is generally accepted that a large enough sample size can be defined to cover any distributions that might occur in a given situation.[2] The question is what to do when the sample size is smaller than such a threshold.

Establishing good confidence intervals on the mean for small samples is an important but often overlooked problem. The t-test is widely used in medical and social sciences. Small clinical trials (such as Phase 1 trials), where such tests could potentially be applied, occur frequently in practice (Institute of Medicine, 2001). In addition, there are several machine learning applications. The sample mean distribution of an importance-weighted estimator is skewed even when the sample size is much larger than 30, so tighter bounds with guarantees may be beneficial. Algorithms in Safe Reinforcement Learning (Thomas et al., 2015) use importance weights to estimate the

[1] An adequate sample size for the Student's t method depends upon the setting, but a common rule is $n > 30$.

[2] An example in which the sample mean is still visibly skewed (and hence inappropriate for use with Student's t) even after $n = 80$ samples is given for log-normal distributions in the supplementary material.

return of a policy and use confidence bounds to estimate the range of the mean. The UCB multi-armed bandit algorithm is designed using the Hoeffding bound - a tighter bound may lead to better performance with guarantees.

In the two-ended support setting, our bounds provide a new and better option for guaranteed coverage with small sample sizes.[3] At least one version of our bound is tighter (or as tight) for *every possible sample* than the bound by Anderson (Anderson, 1969a), which is arguably the best existing bound with guaranteed coverage for small sample sizes. In the limit as $a \to -\infty$, i.e., the one-ended support setting, this version of our bound is equivalent to Anderson.

It can be shown from Learned-Miller and Thomas (2019) that Anderson's UCB is less than or equal to Hoeffding's for *any* sample when $\alpha \leq 0.5$, and is strictly less than Hoeffding's when $\alpha \leq 0.5$ and $n \geq 3$. Therefore our bound is also less than or equal to Hoeffding's for *any* sample when $\alpha \leq 0.5$, and is strictly better than Hoeffding's inequality when $\alpha \leq 0.5$ and $n \geq 3$.

Below we review bounds with coverage guarantees, those that do *not* exhibit guaranteed coverage, and those for which the result is unknown.

3.1.1 Distribution free bounds with guaranteed coverage

Several bounds exist that have guaranteed coverage. These include Hoeffding's inequality (Hoeffding, 1963), Anderson's bound (Anderson, 1969a), and the bound due to Maurer and Pontil (2009).

Hoeffding's inequality. For a distribution F on $[a, b]$, Hoeffding's inequality (Hoeffding, 1963) provides a bound on the probability that the sample mean, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, will deviate from the mean by more than some amount $t \geq 0$:

---

[3]Code accompanying this chapter is available at `https://github.com/myphan9/small_sample_mean_bounds`.

$$\Pr\left(\mu - \bar{X}_n \leq t\right) \leq e^{-\frac{2nt^2}{(b-a)^2}}. \tag{3.2}$$

Defining α to be the right hand side of this inequality, solving for t as a function of α, and rewriting in terms of α rather than t, one obtains a $1-\alpha$ UCB on the mean of

$$b^{\alpha,\text{Hoeffding}}(\mathbf{X}) \stackrel{\text{def}}{=} \bar{X}_n + (b-a)\sqrt{\frac{\ln(1/\alpha)}{2n}}. \tag{3.3}$$

Maurer and Pontil. One limitation of Hoeffding's inequality is that the amount added to the sample mean to obtain the UCB scales with the range of the random variable over $\sqrt{n}$, which shrinks slowly as n increases.

Bennett's inequality (Bennett, 1962) considers both the sample mean and the sample variance and obtains a better dependence on the range of the random variable *when the variance is known.* Maurer and Pontil (2009) derived a UCB for the variance of a random variable, and suggest combining this with Bennet's inequality (via the union bound) to obtain the following $1-\alpha$ UCB on the mean:

$$b^{\alpha,\text{M\&P}}(\mathbf{X}) \stackrel{\text{def}}{=} \bar{X}_n + \frac{7(b-a)\ln(2/\alpha)}{3(n-1)} + \sqrt{\frac{2\hat{\sigma}^2\ln(2/\alpha)}{n}}.$$

Notice that Maurer and Pontil's UCB scales with the range $(b-a)$, divided by n (as opposed to the $\sqrt{n}$ of Hoeffding's). However, the $\sqrt{n}$ dependence is unavoidable to some extent: Maurer and Pontil's UCB scales with the sample standard deviation $\hat{\sigma}$ divided by $\sqrt{n}$. As a result, Maurer and Pontil's bound tends to be tighter than Hoeffding's when both n is large and the range of the random variable is large relative to the variance. Lastly, notice that Maurer and Pontil's bound requires $n \geq 2$ for the sample standard deviation to be defined.

Anderson's bound. Anderson (1969a)[4] introduces a bound by defining an 'envelope' of equal width that, with high probability, contains the true CDF. The upper and lower extremes of such an envelope define the CDFs with the minimum and maximum attainable means for distributions that fit within the envelope, and thus bound the mean with high probability.[5]

In practice, Anderson's bound tends to be significantly tighter than Maurer and Pontil's inequality unless the variance of the random variable is miniscule in comparison to the range of the random variable (and n is sufficiently large). However, neither Anderson's inequality nor Maurer and Pontil's inequality strictly dominates the other. That is, neither upper bound is strictly less than or equal to the other in all cases. However, Anderson's bound *does* dominate Hoeffding's inequality (Learned-Miller and Thomas, 2019).

Some authors have proposed specific envelopes for use with Anderson's technique (Diouf and Dufour, 2005; Learned-Miller and DeStefano, 2008; Romano and Wolf, 2000). However, none of these variations are shown to dominate Anderson's original bound. That is, while they give tighter intervals for some samples, they are looser for others.

Other bounds. Fienberg et al. (1977) proposed a bound for distributions on a discrete set of support points, but nothing prevents it, in theory, from being applied to an arbitrarily dense set of points on an interval such as $[0, 1]$. This bound has a number of appealing properties, and comes with a proof of guaranteed coverage. However, the main drawback is that it is currently computationally intractable, with a

[4] An easier to access and virtually equivalent version of Anderson's work can be found in (Anderson, 1969b).

[5] In his original paper, Anderson also suggests a large family of envelopes, each of which produces a distinct bound. Our simulation results in Section 3.5 are based on the equal-width envelope, but our theoretical results in Section 3.4 hold for all possible envelopes.

computation time that depends exponentially on the number of points in the support set, precluding many (if not most) practical applications.

In an independent concurrent work, Waudby-Smith and Ramdas (2021) proposed another confidence interval for the mean, which generalizes and improves upon Hoeffding's inequality.

3.1.2 Bounds that do not exhibit guaranteed coverage

Many bounds that are used in practice are known to violate Eq. (3.1) for certain distributions. These include the aforementioned Student's t method, and various bootstrap procedures, such as the bias-corrected and accelerated (BCa) bootstrap and the percentile bootstrap. See Efron and Tibshirani (1993) for details of these methods. A simple explanation of the failure of bootstrap methods for certain distributions is given by (Romano and Wolf, 2000, pages 757–758). Presumably if one wants guarantees of Eq. (3.1), one cannot use these methods (unless one has extra information about the unknown distribution).

3.1.3 Bounds conjectured to have guaranteed coverage

There are at least two known bounds that perform well in practice but for which no proofs of coverage are known. One of these, used in accounting procedures, is the so-called Stringer bound (Stringer, 1963). It is known to violate Eq. (1) for confidence levels $\alpha > 0.5$ (Pap and van Zuijlen, 1995), but its coverage for $\alpha < 0.5$ is unknown.

A little known bound by Gaffke (2005) gives remarkably tight bounds on the mean, but has eluded a proof of guaranteed coverage. This bound was recently rediscovered by Learned-Miller and Thomas (2019), who do an empirical study of its performance and provide a method for computing it efficiently.

We demonstrate in Section 3.4 that our bound dominates those of both Hoeffding and Anderson. **To our knowledge, this is the first bound that has been shown to dominate Anderson's bound.**

3.2 A Family of Confidence Bounds

In this section we define our new upper confidence bound. Let n be the sample size. We use bold-faced letters to denote a vector of size n and normal letters to denote a scalar. Uppercase letters denote random variables and lowercase letters denote values taken by them. For example, $X_i \in \mathcal{R}$ and $\mathbf{X} = (X_1, ..., X_n) \in \mathcal{R}^n$ are random variables. $x_i \in \mathcal{R}$ is a value of X_i, and $\mathbf{x} = (x_1, ..., x_n) \in \mathcal{R}^n$ is a value of $\mathbf{X}$. For a sample $\mathbf{x}$, we let $F(\mathbf{x}) \stackrel{\text{def}}{=} (F(x_1), \cdots, F(x_n)) \in [0,1]^n$.

Order statistics play a central role in our work. We denote random variable order statistics $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ and of a specific sample as $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$.

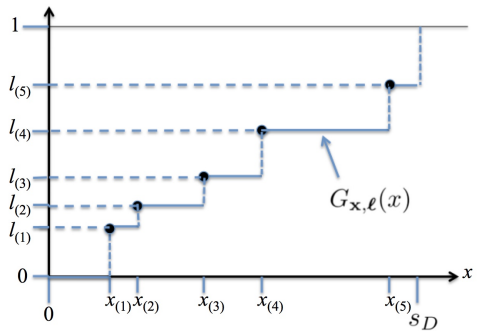
Given a sample $\mathbf{X} = \mathbf{x}$ of size n and a confidence level $1 - \alpha$, we would like to calculate a UCB for the mean. Let F be the CDF of X_i, i.e., the true distribution and $D \subset \mathcal{R}$ be the support of F. We assume that D has a finite upper bound. Given D and any function $T : D^n \to \mathcal{R}$ we will calculate an upper confidence bound $b^{\alpha}_{D,T}(\mathbf{x})$ for the mean of F.

We show in Lemma 8 that if D^+ is a superset of D with finite upper bound, then $b^{\alpha}_{D^+,T}(\mathbf{x}) \geq b^{\alpha}_{D,T}(\mathbf{x})$. Therefore we only need to know a superset of the support with finite upper bound to obtain a guaranteed bound.

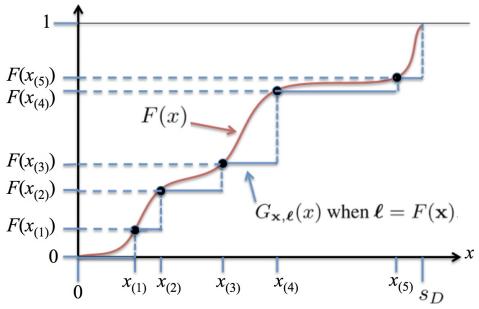
Let $s_D \stackrel{\text{def}}{=} \sup\{x : x \in D\}$. We next describe a method for pairing the sample $\mathbf{x}$ with another vector $\boldsymbol{\ell} \in [0,1]^n$ to produce a stairstep CDF function $G_{\mathbf{x},\boldsymbol{\ell}}$. Let $x_{(n+1)} \stackrel{\text{def}}{=} s_D$. Consider the step function $G_{\mathbf{x},\boldsymbol{\ell}} : \mathcal{R} \to [0,1]$ defined from $\boldsymbol{\ell}$ and $\mathbf{x}$ as follows (see Figure 3.1a):

$$G_{\mathbf{x},\boldsymbol{\ell}}(x) = \begin{cases} 0, & \text{if } x < x_{(1)} \\ \ell_{(i)}, & \text{if } x_{(i)} \leq x < x_{(i+1)} \\ 1, & \text{if } x \geq s_D. \end{cases} \tag{3.4}$$

In particular, when $\boldsymbol{\ell} = (1/n, \ldots, n/n)$, $G_{\mathbf{x},\boldsymbol{\ell}}$ becomes the empirical CDF. Also note that when $\boldsymbol{\ell} = F(\mathbf{x})$, $\forall x, G_{\mathbf{x},\boldsymbol{\ell}}(x) \leq F(x)$, as illustrated in Figure 3.1b.



(a) The stairstep function $G_{\mathbf{x},\boldsymbol{\ell}}$, which is a function of the sample $\mathbf{x}$ and a vector $\boldsymbol{\ell}$ of values between 0 and 1. When $\boldsymbol{\ell} = (1/n, \ldots, n/n)$, $G_{\mathbf{x},\boldsymbol{\ell}}$ becomes the empirical CDF.



(b) The CDF of a distribution F in red, with a random sample of five order statistics on the x-axis. The blue stairstep function shows the function $G_{\mathbf{x},\boldsymbol{\ell}}(x)$ when $\boldsymbol{\ell} = F(\mathbf{x})$. Notice that for all x, $G_{\mathbf{x},\boldsymbol{\ell}}(x) \leq F(x)$.

Figure 3.1

Following Learned-Miller and Thomas (2019), if we consider $G_{\mathbf{x},\boldsymbol{\ell}}$ to be a CDF, we can compute the mean of the resulting distribution as a function of two vectors $\mathbf{x}$ and $\boldsymbol{\ell}$ as

$$m_D(\mathbf{x}, \boldsymbol{\ell}) \stackrel{\text{def}}{=} \sum_{i=1}^{n+1} x_{(i)}(\ell_{(i)} - \ell_{(i-1)}) \tag{3.5}$$

$$= s_D - \sum_{i=1}^{n} \ell_{(i)}(x_{(i+1)} - x_{(i)}), \tag{3.6}$$

where $\ell_{(0)} \stackrel{\text{def}}{=} 0$, $\ell_{(n+1)} \stackrel{\text{def}}{=} 1$ and $x_{(n+1)} \stackrel{\text{def}}{=} s_D$. When s_D is finite, this is well-defined. Notice that this function is defined in terms of the *order statistics* of $\mathbf{x}$ and $\boldsymbol{\ell}$. Learned-Miller and Thomas (2019) refer to this as the *induced mean* for the sample $\mathbf{x}$ by the vector $\boldsymbol{\ell}$. Although we borrow the above terms from Learned-Miller and Thomas (2019), the bound we introduce below is a new class of bounds, and differs from the bounds discussed in their work.

An ordering on D^n. Next, we introduce a scalar-valued function T which we will use to define a total order on samples in D^n, and define a set of samples less

than or equal to another sample. In particular, for any function $T : \mathcal{R}^n \to \mathcal{R}$, let $\mathbb{S}_{D,T}(\mathbf{x}) = \{\mathbf{y} \in D^n | T(\mathbf{y}) \leq T(\mathbf{x})\}$.

The greatest induced mean for a given U. Let $\mathbf{U} = U_1, ..., U_n$ be a sample of size n from the continuous uniform distribution on $[0, 1]$, with $\mathbf{u} \stackrel{\text{def}}{=} (u_1, \cdots, u_n)$ being a particular sample of $\mathbf{U}$.

Now consider the random quantity

$$b_{D,T}(\mathbf{x}, \mathbf{U}) \stackrel{\text{def}}{=} \sup_{\mathbf{z} \in \mathbb{S}_{D,T}(\mathbf{x})} m_D(\mathbf{z}, \mathbf{U}), \tag{3.7}$$

which depends upon a fixed sample $\mathbf{x}$ (non-random) and also on the random variable $\mathbf{U}$.

Our upper confidence bound. Let $0 < p < 1$. Let $Q(p, Y)$ be the *quantile function* of the scalar random variable Y, i.e.,

$$Q(p, Y) \stackrel{\text{def}}{=} \inf\{y \in \mathbb{R} : F_Y(y) \geq p\}, \tag{3.8}$$

where $F_Y(y)$ is the CDF of Y. We define $b^{\alpha}_{D,T}(\mathbf{x})$ to be the $(1 - \alpha)$-quantile of the random quantity $b_{D,T}(\mathbf{x}, \mathbf{U})$.

Definition 2 (Upper confidence bound on the mean)**.** *Given a sample* $\mathbf{x}$ *and a confidence level* $1 - \alpha$*:*

$$b^{\alpha}_{D,T}(\mathbf{x}) \stackrel{\text{def}}{=} Q(1 - \alpha, b_{D,T}(\mathbf{x}, \mathbf{U})), \tag{3.9}$$

where $b_{D,T}(\mathbf{x}, \mathbf{U})$ *is defined in Eq. 3.7.*

To simplify notation, we drop the superscript and subscripts whenever clear. We show in Section 3.2.1 that this UCB has guaranteed coverage for all sample sizes n, for all confidence levels $0 < 1 - \alpha < 1$ and for all distributions F and support D where s_D is finite.

We show below that a bound computed from a superset $D^+ \supseteq D$ will be looser than or equal to a bound computed from the support D. Therefore it is enough to know a superset of the support D to obtain a bound with guaranteed coverage.

Lemma 8. *Let $D^+ \supseteq D$ where s_{D^+} is finite. For any sample $\mathbf{x}$:*

$$b_D^\alpha(\mathbf{x}) \leq b_{D^+}^\alpha(\mathbf{x}). \tag{3.10}$$

Proof. Since s_{D^+} is finite, $m_{D^+}(\mathbf{y}, \mathbf{u})$ is well-defined. Since $D \subseteq D^+$, for any $\mathbf{y}$ and $\mathbf{u}$, $m_D(\mathbf{y}, \mathbf{u}) \leq m_{D^+}(\mathbf{y}, \mathbf{u})$. Then

$$\sup_{\mathbf{y} \in \mathbb{S}_D(\mathbf{x})} m_D(\mathbf{y}, \mathbf{u}) \leq \sup_{\mathbf{y} \in \mathbb{S}_D(\mathbf{x})} m_{D^+}(\mathbf{y}, \mathbf{u}) \tag{3.11}$$

$$\leq \sup_{\mathbf{y} \in \mathbb{S}_{D^+}(\mathbf{x})} m_{D^+}(\mathbf{y}, \mathbf{u}), \tag{3.12}$$

where the last inequality is because $\mathbb{S}_D(\mathbf{x}) \subseteq \mathbb{S}_{D^+}(\mathbf{x})$.

Let $b_D(\mathbf{x}, \mathbf{U}) = \sup_{\mathbf{z} \in \mathbb{S}_D(\mathbf{x})} m_D(\mathbf{z}, \mathbf{U})$ and $b_{D^+}(\mathbf{x}, \mathbf{U}) = \sup_{\mathbf{z} \in \mathbb{S}_{D^+}(\mathbf{x})} m_{D^+}(\mathbf{z}, \mathbf{U})$. Then $b_D^\alpha(\mathbf{x})$ and $b_{D^+}^\alpha(\mathbf{x})$ are the $(1-\alpha)$-quantiles of $b_D(\mathbf{x}, \mathbf{U})$ and $b_{D^+}(\mathbf{x}, \mathbf{U})$. Since $b_D(\mathbf{x}, \mathbf{u}) \leq b_{D^+}(\mathbf{x}, \mathbf{u})$ for any $\mathbf{u}$, $b_D^\alpha(\mathbf{x}) \leq b_{D^+}^\alpha(\mathbf{x})$. □

In Section 3.2.1 we show that the bound has guaranteed coverage. In Section 3.3 we discuss how to efficiently compute the bound. In Section 3.4 we show that when T is a certain linear function, the bound is equal to or tighter than Anderson's for any sample. In addition, we show that when the support is known to be $\{0, 1\}$, our bound recovers the well-known Clopper-Pearson confidence bound for binomial distributions (Clopper and Pearson, 1934). In Section 3.5, we present simulations that show the consistent superiority of our bounds over previous bounds.

3.2.1 Guaranteed Coverage

In this section we show that our bound has guaranteed coverage in Theorem 8. We omit superscripts and subscripts if they are clear from context.

3.2.1.1 Preview of proof

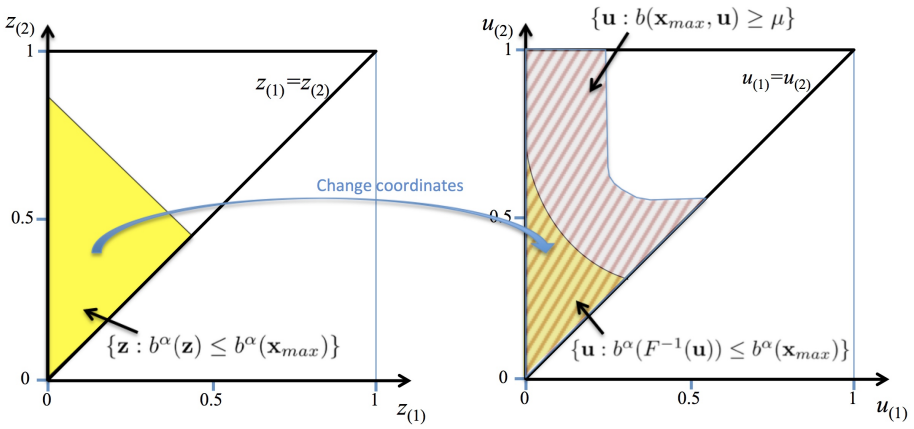


Figure 3.2: Illustrations of Section 3.2.1.1. **Left.** The yellow region shows samples of $\mathbf{z} = [z_{(1)}, z_{(2)}]$ such that $b^\alpha(\mathbf{z}) \leq b^\alpha(\mathbf{x}_{max})$. **Right.** The same yellow region, but in the coordinates $\mathbf{u} = F^{-1}(\mathbf{z})$. We will show that the yellow region is a subset of the striped, which contains $\mathbf{u}$ such that $b(\mathbf{x}_{max}, \mathbf{u}) \geq \mu$.

We explain the idea behind our bound at a high level using a special case. Note that our proof is more general than our special case, which makes assumptions such as the continuity of F to simplify the intuition.

Suppose that F is continuous. Then the *probability integral transform* $F_X(X)$ of X is uniformly distributed on $[0, 1]$ (Angus, 1994). Suppose there exists a sample $\mathbf{x}_\mu$ such that $b^\alpha(\mathbf{x}_\mu) = \mu$. Then the probability that a sample $\mathbf{Z}$ outputs $b^\alpha(\mathbf{Z}) < \mu$ is equal to the probability $\mathbf{Z}$ outputs $b^\alpha(\mathbf{Z}) < b^\alpha(\mathbf{x}_\mu)$ (the yellow region on the left of Fig. 3.2). This is the region where the bound fails, and we would like to show that the probability of this region is at most α.

Let $\mathbf{U} \stackrel{\text{def}}{=} F(\mathbf{Z})$ and $\mathbf{u} \stackrel{\text{def}}{=} F(\mathbf{z})$. Then U_i is uniformly distributed on $[0, 1]$. If F is invertible, we can transform the region $\{\mathbf{z} : b^\alpha(\mathbf{z}) < b^\alpha(\mathbf{x}_\mu)\}$ to $\{\mathbf{u} : b^\alpha(F^{-1}(\mathbf{u})) < b^\alpha(\mathbf{x}_\mu)\}$ where $F^{-1}(\mathbf{u}) \stackrel{\text{def}}{=} (F^{-1}(u_1), \ldots, F^{-1}(u_n))$ (the yellow region on the right of Fig. 3.2).

Through some calculations using the definition of function b, we can show that the yellow region $\{\mathbf{u} : b^\alpha(F^{-1}(\mathbf{u})) < b^\alpha(\mathbf{x}_\mu)\}$ is a subset of the striped region $\{\mathbf{u} : b(\mathbf{x}_\mu, \mathbf{u}) \geq \mu\}$.

Note that since $b^\alpha(\mathbf{x}_\mu) = \mu$, μ is equal to the $1-\alpha$ quantile of $b(\mathbf{x}_\mu, \mathbf{U})$. Therefore, by the definition of quantile, the probability of the striped region is at most α:

$$\mathbb{P}_{\mathbf{U}}(b(\mathbf{x}_\mu, \mathbf{U}) \geq \mu) \leq \alpha, \tag{3.13}$$

and thus the probability of the yellow region is at most α.

3.2.1.2 Main Result

In this section, we present some supporting lemmas and then the main result in Theorem 8. The proofs of the simpler lemmas have been deferred to the supplementary material.

Lemma 9. *Let X be a random variable with CDF F and $Y \stackrel{\text{def}}{=} F(X)$, known as the probability integral transform of X. Let U be a uniform random variable on $[0, 1]$. Then for any $0 \leq y \leq 1$,*

$$\mathbb{P}(Y \leq y) \leq \mathbb{P}(U \leq y). \tag{3.14}$$

If F is continuous, then Y is uniformly distributed on $[0, 1]$.

In the next lemma we show that the mean satisfies the following property. Let F and G be two CDF functions such that $F(x)$ is always larger than or equal to $G(x)$ for all x. Then the mean of F is smaller than the mean of G.

Lemma 10. *Let F and $G_{\mathbf{x},\boldsymbol{\ell}}$ be two CDF functions such that $\forall x \in \mathcal{R}$, $F(x) \geq G_{\mathbf{x},\boldsymbol{\ell}}(x)$. Let μ_F and μ_G denote the means of F and $G_{\mathbf{x},\boldsymbol{\ell}}$. Then*[6]

$$\mu_F \leq \mu_G. \tag{3.15}$$

For use in the next lemma, we define a partial order for the samples on D^n. Note that it is defined with respect to the *order statistics* of the sample, not the original components.

Definition 3 (Partial Order)**.** *For any two samples $\mathbf{z}$ and $\mathbf{y}$, we define $\mathbf{z} \preceq \mathbf{y}$ to indicate that $z_{(i)} \leq y_{(i)}, 1 \leq i \leq n$.*

Lemma 11. *Let $\mathbf{Z}$ be a random sample of size n from F. Let $\mathbf{U} = U_1, ..., U_n$ be a sample of size n from the continuous uniform distribution on $[0, 1]$. For any function $T : D^n \to R$ and any $\mathbf{x} \in D^n$:*

$$\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) \leq \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu). \tag{3.16}$$

Proof sketch. Let $\cup$ denote the union of events and $\{\}$ denote an event. Then for any $\mathbf{x} \in D^n$:

$$\begin{aligned}
\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) &= \mathbb{P}_{\mathbf{Z}}(\mathbf{Z} \in \mathbb{S}(\mathbf{x})) && (3.17)\\
&= \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{\mathbf{Z} = \mathbf{y}\}) && (3.18)\\
&\leq \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{\mathbf{Z} \preceq \mathbf{y}\}) && (3.19)\\
&\leq \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{F(\mathbf{Z}) \preceq F(\mathbf{y})\}) \text{ by monotone } F && (3.20)\\
&\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{\mathbf{U} \preceq F(\mathbf{y})\}). && (3.21)
\end{aligned}$$

[6]This is the only property required of the mean for the subsequent lemmas and theorems. Since quantiles of a distribution also satisfy this condition, this method could also be used to give UCBs for various quantiles.

The last step is by an extension of Lemma 9. Recall that $m_D(\mathbf{y}, \mathbf{u}) = s_D - \sum_{i=1}^{n} u_{(i)}(y_{(i+1)} - y_{(i)})$ where $\forall i, y_{(i+1)} - y_{(i)} \geq 0$. Therefore if $\mathbf{u} \preceq F(\mathbf{y})$ then $m_D(\mathbf{y}, \mathbf{u}) \geq m_D(\mathbf{y}, F(\mathbf{y}))$:

$$\mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{\mathbf{U} \preceq F(\mathbf{y})\}) \tag{3.22}$$

$$\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{m_D(\mathbf{y}, \mathbf{U}) \geq m_D(\mathbf{y}, F(\mathbf{y}))\}), \text{by Lemma 10} \tag{3.23}$$

$$\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{m_D(\mathbf{y}, \mathbf{U}) \geq \mu\}), \text{by Lemma 10} \tag{3.24}$$

$$\leq \mathbb{P}_{\mathbf{U}}(\sup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})} m_D(\mathbf{y}, \mathbf{U}) \geq \mu) \tag{3.25}$$

$$= \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu). \tag{3.26}$$

□

We include a more detailed version of the proof for the above lemma in the supplementary material.

Lemma 12. *Let $\mathbf{U} = U_1, ..., U_n$ be a sample of size n from the continuous uniform distribution on $[0, 1]$. Let $\mathbf{X}$ and $\mathbf{Z}$ denote i.i.d. samples of size n from F. For any function $T : D^n \to \mathcal{R}$ and any $\alpha \in (0, 1)$,*

$$\mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{U}}(b_{D,T}(\mathbf{X}, \mathbf{U}) \geq \mu) \leq \alpha\right) \leq \mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{X})) \leq \alpha\right).$$

Proof. From Lemma 11 for any sample $\mathbf{x}$,

$$\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) \leq \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu). \tag{3.27}$$

Therefore,

$$\mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{X})) \leq \alpha\right) \geq \mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{U}}(b(\mathbf{X}, \mathbf{U}) \geq \mu) \leq \alpha\right). \tag{3.28}$$

□

Lemma 13. *Let* $\mathbf{U} = U_1, ..., U_n$ *be a sample of size* n *from the continuous uniform distribution on* $[0, 1]$*. Let* $\mathbf{X}$ *be a random sample of size* n *from* F*. For any function* $T : D^n \to \mathcal{R}$ *and any* $\alpha \in (0, 1)$*,*

$$\mathbb{P}_{\mathbf{X}}(b^{\alpha}_{D,T}(\mathbf{X}) < \mu) \leq \mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{U}}(b_{D,T}(\mathbf{X}, \mathbf{U}) \geq \mu) \leq \alpha\right). \tag{3.29}$$

Proof. Because $b^{\alpha}(\mathbf{x})$ is the $1 - \alpha$ quantile of $b(\mathbf{x}, \mathbf{U})$, by the definition of quantile: $\mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \leq b^{\alpha}(\mathbf{x})) \geq 1 - \alpha$. Therefore $\mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq b^{\alpha}(\mathbf{x})) \leq \alpha$. If $b^{\alpha}(\mathbf{x}) < \mu$ then $\mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu) \leq \alpha$. Since $b^{\alpha}(\mathbf{x}) < \mu$ implies $\mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu) \leq \alpha$, we have

$$\mathbb{P}_{\mathbf{X}}(b^{\alpha}(\mathbf{X}) < \mu) \leq \mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{U}}(b(\mathbf{X}, \mathbf{U}) \geq \mu) \leq \alpha\right). \tag{3.30}$$

□

We now show that the bound has guaranteed coverage.

Theorem 8. *Let* $\mathbf{X}$ *be a random sample of size* n *from* F*. For any function* $T : D^n \to R$ *and for any* $\alpha \in (0, 1)$*:*

$$\mathbb{P}_{\mathbf{X}}(b^{\alpha}_{D,T}(\mathbf{X}) < \mu) \leq \alpha. \tag{3.31}$$

Proof. Let $\mathbf{Z}$ be a random sample of size n from F.

$$\begin{aligned}
\mathbb{P}_{\mathbf{X}}(b^{\alpha}(\mathbf{X}) < \mu) &\leq \mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{U}}(b(\mathbf{X}, \mathbf{U}) \geq \mu) \leq \alpha\right) \text{ by Lemma 13} && (3.32)\\
&\leq \mathbb{P}_{\mathbf{X}}\left(\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{X})) \leq \alpha\right) \text{ by Lemma 12} && (3.33)\\
&= \mathbb{P}\left(W \leq \alpha\right) \text{ where } W \stackrel{\text{def}}{=} \mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{X})) && (3.34)\\
&\leq \alpha \text{ by Lemma 9.} && (3.35)
\end{aligned}$$

□

3.3 Computation

In this section we present a Monte Carlo algorithm to compute the bound. First we note that since the bound only depends on $\mathbf{x}$ via the function $T(\mathbf{x})$, we can precompute a table of the bounds for each value of $T(\mathbf{x})$. We discuss how to adjust for the uncertainty in the Monte Carlo result in Appendix C.4.

Algorithm 1 Monte Carlo estimation of $m^{\alpha}_{D^+,T}(\mathbf{x})$ where $D^+ = [0,1]$. This pseudocode uses 1-based array indexing.

Input: A sample $\mathbf{x} \in D^n$, confidence parameter $1-\alpha < 1$, a function $T : [0,1]^n \to \mathcal{R}$ and Monte Carlo sampling parameter l.
Output: An estimation of $m^{\alpha}_{D^+,T}(\mathbf{x})$
$n \leftarrow length(\mathbf{x})$.
Create array **ms** to hold l floating point numbers, and initialize it to zero.
Create array **u** to hold n floating point numbers.
for $i \leftarrow 1$ to l **do**
  for $j \leftarrow 1$ to n **do**
    u[j] $\sim$ Uniform(0,1).
  end for
  Sort(**u**, ascending).
  Solve: $M = \max_{y_{(1)},\cdots,y_{(n)}} m(\mathbf{y},\mathbf{u})$ subject to:
    1) $T(\mathbf{y}) \le T(\mathbf{x})$.
    2) $\forall i : 1 \le i \le n, 0 \le y_{(i)} \le 1$.
    3) $y_{(1)} \le y_{(2)} \le ... \le y_{(n)}$.
  ms$[i] = M$.
end for
Sort(**ms**, ascending).
Return **ms**$[\lceil (1-\alpha)l \rceil]$.

Let the superset of the support D^+ be a closed interval with a finite upper bound. If m is a continuous function,

$$\sup_{\mathbf{y} \in \mathbf{S}_{D^+}(\mathbf{x})} m(\mathbf{y},\mathbf{u}) = \max_{\mathbf{y} \in \mathbf{S}_{D^+}(\mathbf{x})} m(\mathbf{y},\mathbf{u}). \tag{3.36}$$

Therefore $b_{D^+}(\mathbf{x},\mathbf{u})$ is the solution to

$$\max_{y_{(1)},\ldots,y_{(n)}} m(\mathbf{y},\mathbf{u}) \tag{3.37}$$

subject to:

- $T(\mathbf{y}) \leq T(\mathbf{x})$,
- $\forall i \in \{1, \ldots, n\}, y_{(i)} \in D^+$,
- $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$.

When D^+ is an interval and T is linear, this is a linear programming problem and can be solved efficiently.

We can compute the $1-\alpha$ quantile of a random variable M using Monte Carlo simulation, sampling M l times. Letting $m_{(1)} \leq ... \leq m_{(l)}$ be the sorted values, we output $m_{(\lceil (1-\alpha)l \rceil)}$ as an approximation of the $1-\alpha$ quantile.

Running time. Note that since the bound only depends on $\mathbf{x}$ via the function $T(\mathbf{x})$, we can precompute a table of the bounds for each value of $T(\mathbf{x})$ to save time. When T is linear, the algorithm needs to solve a linear programming problem with n variables and $2n$ constraints l times. For sample size $n = 50$, computing the bound for each sample $\mathbf{x} \in D^n$ takes just a few seconds using $l = 10{,}000$ Monte Carlo samples.

3.4 Relationships with Existing Bounds

In this section, we compare our bound to previous bounds including those of Clopper and Pearson, Hoeffding, and Anderson. Proofs omitted in this section can be found in the supplementary material.

3.4.1 Special Case: Bernoulli Distribution

When we know that $D = \{0, 1\}$, the distribution is Bernoulli. If we choose T to be the sample mean, our bound becomes the same as the Clopper-Pearson confidence bound for binomial distributions (Clopper and Pearson, 1934). See the supplementary material for details.

3.4.2 Comparisons with Anderson and Hoeffding

In this section we show that for any sample size n, any confidence level α and for any sample $\mathbf{x}$, our method produces a bound no larger than Anderson's bound (Theorem 9) and Hoeffding's bound (Theorem 10).

Note that if we only know an upper bound b of the support (1-ended support setting), we can set $D^+ = (-\infty, b]$ and our method is equal to Anderson's and dominates Hoeffding's. As the lower support bound increases (2-ended setting), our bound becomes tighter or remains constant, whereas Anderson's remains constant, as it does not incorporate information about a lower support. Thus, in some cases where our bound can benefit from a lower support, we are tighter than Anderson's. We discuss the details in Appendix C.5.3 (Theorem 14).

Anderson's bound constructs an upper bound for the mean by constructing a lower bound for the CDF. We defined a lower bound for the CDF as follows.

Definition 4 (Lower confidence bound for the CDF)**.** *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a sample of size n from the distribution on $\mathcal{D}^+$ with unknown CDF F. Let $\alpha \in (0, 1)$. Let $H_{\mathbf{X}} : \mathcal{R} \to [0, 1]$ be a function computed from the sample $\mathbf{X}$ such that for any CDF F,*

$$\mathbb{P}_{\mathbf{X}}(\ \forall x \in R, F(x) \geq H_{\mathbf{X}}(x)) \geq 1 - \alpha. \tag{3.38}$$

Then $H_{\mathbf{X}}$ is called a $(1 - \alpha)$ lower confidence bound for the CDF.

If there exists a CDF F such that

$$\mathbb{P}_{\mathbf{X}}(\ \forall x \in R, F(x) \geq H_{\mathbf{X}}(x)) = 1 - \alpha, \tag{3.39}$$

then $H_{\mathbf{X}}$ is called an exact *$(1 - \alpha)$ lower confidence bound for the CDF.*

In Figs. 3.1a and 3.1b, it is easy to see that if the stairstep function $G_{\mathbf{X},\boldsymbol{\ell}}$ is a lower confidence bound for the CDF then its induced mean $m(\mathbf{X}, \boldsymbol{\ell})$ is an upper confidence bound for μ.

Lemma 14. *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a sample of size n from a distribution with mean μ. Let $\boldsymbol{\ell} \in [0,1]^n$. If $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $(1-\alpha)$ lower confidence bound for the CDF then*

$$\mathbb{P}_{\mathbf{X}}(m(\mathbf{X}, \boldsymbol{\ell}) \geq \mu) \geq 1 - \alpha. \tag{3.40}$$

Let $U_{(i)}, 1 \leq i \leq n$ be the order statistics of the uniform distribution. Note that for any CDF F:

$$\mathbb{P}_{\mathbf{X}}(\ \forall x \in \mathcal{R}, F(x) \geq G_{\mathbf{X},\boldsymbol{\ell}}(x)) = \mathbb{P}_{\mathbf{X}}(\forall i : 1 \leq i \leq n, F(X_{(i)}) \geq \ell_{(i)}) \tag{3.41}$$

$$\geq \mathbb{P}_{\mathbf{U}}(\forall i : 1 \leq i \leq n, U_{(i)} \geq \ell_{(i)}) \text{ by Lemma 9,} \tag{3.42}$$

where Eq. 3.42 is an equality if F is the CDF of a continuous random variable. Therefore $G_{\mathbf{X},\boldsymbol{\ell}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF is equivalent to $\boldsymbol{\ell}$ satisfying:

$$\mathbb{P}_{\mathbf{U}}(\forall i : 1 \leq i \leq n, U_{(i)} \geq \ell_{(i)}) = 1 - \alpha. \tag{3.43}$$

Anderson (1969a) presents $b_{\boldsymbol{\ell}}^{\alpha,\text{Anderson}}(\mathbf{x}) = m_{D^+}(\mathbf{x}, \boldsymbol{\ell})$ as a UCB for μ where $\boldsymbol{\ell} \in [0,1]^n$ is a vector such that $G_{\mathbf{X},\boldsymbol{\ell}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF.

In one instance of Anderson's bound, $\boldsymbol{\ell} = \mathbf{u}^{And} \in [0,1]^n$ is defined as

$$u_i^{\text{And}} \stackrel{\text{def}}{=} \max\left\{0, i/n - \beta(n)\right\}. \tag{3.44}$$

Anderson identifies $\beta(n)$ as the one-sided Kolmogorov-Smirnov statistic such that $G_{\mathbf{X},\boldsymbol{\ell}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF when $\boldsymbol{\ell} = \mathbf{u}^{\text{And}}$. $\beta(n)$ can be computed by Monte Carlo simulation (Appendix C.1).

Learned-Miller and Thomas (2019) show that for any sample $\mathbf{x}$, a looser version of Anderson's bound is better than Hoeffding's:

Lemma 15 (from Theorem 2 from (Learned-Miller and Thomas, 2019))**.** *For any sample size n, for any sample value $\mathbf{x} \in D^n$, for all $\alpha \in (0, 0.5]$:*

$$b_{\boldsymbol{\ell}}^{\alpha,Anderson}(\mathbf{x}) \leq b^{\alpha,Hoeffding}(\mathbf{x}), \tag{3.45}$$

where $\boldsymbol{\ell}$ is defined[7] *as*

$$\ell_i \stackrel{\text{def}}{=} \max\left\{0, i/n - \sqrt{\ln(1/\alpha)/(2n)}\right\}. \tag{3.46}$$

When $\alpha \leq 0.5$, this definition of $\boldsymbol{\ell}$ satisfies $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $(1-\alpha)$ lower confidence bound for the CDF.

The inequality in Eq. 3.45 is strict for $n \geq 3$.

We show below that our bound is always equal to or tighter than Anderson's bound. Appendix C.5.3 provides a more detailed analysis showing that our bound is equal to Anderson's when the lower bound of the support is too small and can be tighter than Anderson's when the lower bound of the support is large enough.

Theorem 9. *Let $\boldsymbol{\ell} \in [0,1]^n$ be a vector satisfying $G_{\mathbf{X},\boldsymbol{\ell}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF.*

[7] Although Anderson's bound $b_{\boldsymbol{\ell}}^{\alpha,\text{Anderson}}(\mathbf{x})$ is only defined when $G_{\mathbf{X},\boldsymbol{\ell}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF, here we re-use the same notation for the case when $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $(1-\alpha)$ lower confidence bound for the CDF.

Let $D^+ = (-\infty, b]$. For any sample size n, for any sample value $\mathbf{x} \in D^n$, for all $\alpha \in (0,1)$, using $T(\mathbf{x}) = b_{\boldsymbol{\ell}}^{\alpha,Anderson}(\mathbf{x})$ yields

$$b_{D^+,T}^{\alpha}(\mathbf{x}) \leq b_{\boldsymbol{\ell}}^{\alpha,Anderson}(\mathbf{x}). \tag{3.47}$$

We explain briefly why this is true. First, from Figure 3.1b, we can see that if $G_{\mathbf{X},\boldsymbol{\ell}}$ is a lower confidence bound then $\forall i, F(X_{(i)}) \geq \ell_{(i)}$. Note that $G_{\mathbf{X},\boldsymbol{\ell}}$ must be a lower bound for all unknown CDFs F, so we can pick a continuous F where, according to Lemma 9, $U \stackrel{\text{def}}{=} F(X)$ is uniformly distributed on $[0,1]$. Therefore $\boldsymbol{\ell}$ satisfies for any CDF F':

$$\mathbb{P}_{\mathbf{X}}(\forall i, F'(X_{(i)}) \geq \ell_{(i)}) \geq (\mathbb{P}_{\mathbf{U}}(\forall i, U_{(i)} \geq \ell_{(i)}) \text{ by Lemma 9} \tag{3.48}$$

$$\geq 1 - \alpha, \tag{3.49}$$

where the $U_{(i)}$'s are the order statistics of the uniform distribution. Since $b(\mathbf{x}, \mathbf{U})$ is defined from linear functions of $\mathbf{U}$ with negative coefficients (Eq. 3.6), if $\forall i, U_{(i)} \geq \ell_{(i)}$ then $b(\mathbf{x}, \mathbf{U}) \leq b(\mathbf{x}, \boldsymbol{\ell})$. Therefore with probability at least $1 - \alpha$, $b(\mathbf{x}, \mathbf{U}) \leq b(\mathbf{x}, \boldsymbol{\ell})$. So $b(\mathbf{x}, \boldsymbol{\ell})$ is at least the $1 - \alpha$ quantile of $b(\mathbf{x}, \mathbf{U})$, which is the value of our bound. Therefore $b(\mathbf{x}, \boldsymbol{\ell})$ is at least the value of our bound.

Finally, if T is Anderson's bound, through some calculations we can show that $b_{D^+,T}(\mathbf{x}, \boldsymbol{\ell}) = m_{D^+}(\mathbf{x}, \boldsymbol{\ell})$, which is Anderson's bound. The result follows.

The comparison with Hoeffding's bound follows directly from Lemma 15 and Theorem 9:

Theorem 10. *Let $D^+ = (-\infty, b]$. For any sample size n, for any sample value $\mathbf{x} \in D^n$, for all $\alpha \in (0, 0.5]$, using $T(\mathbf{x}) = b_{\boldsymbol{\ell}}^{\alpha,Anderson}(\mathbf{x})$ where $\boldsymbol{\ell} = \mathbf{u}^{And}$ yields:*

$$b_{D^+,T}^{\alpha}(\mathbf{x}) \leq b^{\alpha,Hoeffding}(\mathbf{x}), \tag{3.50}$$

where the inequality is strict when $n \geq 3$.

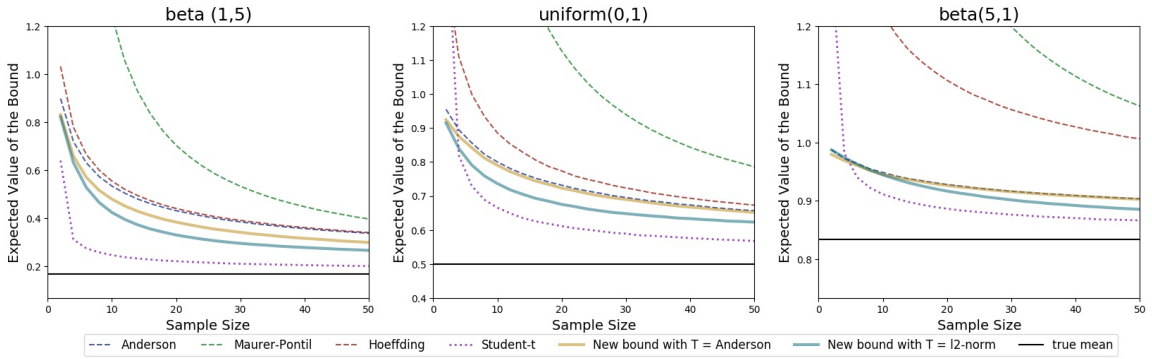


Figure 3.3: The expected value of the bounds for $\alpha = 0.05$ and $D^+ = [0, 1]$. For each sample size, we sample $\mathbf{X}$ 10,000 times, compute the bound for each sample, and take the average. Our new bound with T being Anderson's bound consistently has lower expected value than Anderson's (Theorem 9), Hoeffding's (Theorem 10) and Maurer and Pontil's. With T being the $l2$-norm, the bound is substantially tighter in these examples, and also has guaranteed coverage.

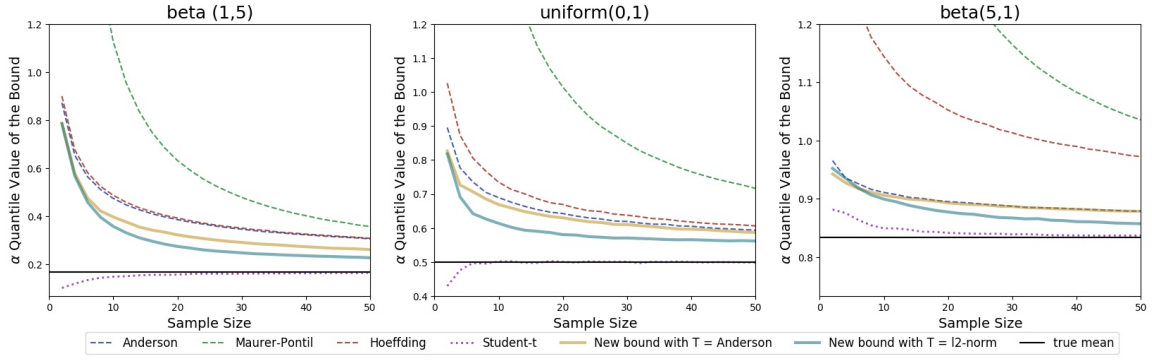


Figure 3.4: The α-quantile of the bound distribution for $\alpha = 0.05$ and $D^+ = [0, 1]$. For each sample size, we sample $\mathbf{X}$ 10,000 times, compute the bound for each sample, and take the α quantile. If the α-quantile is below the true mean, the bound does not have guaranteed coverage. For the $\texttt{uniform}(0, 1)$ and $\texttt{beta}(1, 5)$ distribution, when the sample size is small, Student-t does not have guarantee.

Diouf and Dufour (2005) present several instances of Anderson's bound with different $\boldsymbol{\ell}$ computed from the Anderson-Darling or the Eicker statistics (Theorem 4, 5 and Theorem 6 with constant ϵ).

Note that the result from Theorem 9 can be generalized for bounds $m(\mathbf{X}, \boldsymbol{\ell})$ constructed from a $(1 - \alpha)$ confidence lower bound $G_{\mathbf{X},\boldsymbol{\ell}}$ using Lemma 14. We show the general case in the supplementary material.

3.5 Simulations

We perform simulations to compare our bounds to Hoeffding's inequality, Anderson's bound, Maurer and Pontil's, and Student-t's bound (Student, 1908), the latter being

$$b^{\alpha,\text{Student}}(\mathbf{X}) \stackrel{\text{def}}{=} \bar{X}_n + \sqrt{\frac{\hat{\sigma}^2}{n}} t_{1-\alpha,n-1}. \tag{3.51}$$

We compute Anderson's bound with $\boldsymbol{\ell} = \mathbf{u}^{And}$ defined in Eq. 3.44 through Monte Carlo simulation (described in Appendix C.1). We use $\alpha = 0.05$, $D^+ = [0, 1]$ and $l = 10{,}000$ Monte Carlo samples. We consider two functions T:

1. Anderson: $T(\mathbf{x}) = b_{\boldsymbol{\ell}}^{\alpha,\text{Anderson}}(\mathbf{x})$, again with $\boldsymbol{\ell} = \mathbf{u}^{And}$. Because this T is linear in $\mathbf{x}$, it can be computed with the linear program in Eq. 3.36.

2. $l2$ norm: $T(\mathbf{x}) = (\sum_{i=1}^{n} x_i^2)/n$. In this case, T requires the optimization of a linear functional over a convex region, which results in a simple convex optimization problem.

We perform experiments on three distributions: $\texttt{beta}(1, 5)$ (skewed right), $\texttt{uniform}(0, 1)$ and $\texttt{beta}(5, 1)$ (skewed left). Their PDFs are included in the supplementary material for reference. Additional experiments are in the supplementary material.

In Figure 3.3 and Figure 3.4 we plot the expected value and the α-quantile value of the bounds as the sample size increases. Consistent with Theorem 9, our bound with T being Anderson's bound outperforms Anderson's bound. Our new bound performs better than Anderson's in distributions that are skewed right, and becomes similar to Anderson's in left-skewed distributions. Our bound outperforms Hoeffding and Maurer and Pontil's for all three distributions. Student-t fails (the error rate exceeds α) for $\texttt{beta}(1, 5)$ and $\texttt{uniform}(0, 1)$ when the sample size is small (Figure 3.4).

APPENDIX A

SUPPLEMENTARY MATERIAL: THOMPSON SAMPLING WITH APPROXIMATE INFERENCE

A.1 Proof of Theorem 1 and Corollary 1

First we will prove Theorem 1. Let $\Omega_i \subseteq \Omega$ denote the region where arm i is the best arm. Let $\Pi_{t,i}$ denote $\Pi_t(\Omega_i)$, the posterior probability that arm i is the best arm. For $r > 1$, We construct the pdf of Q_t's as follows:

$$q_t(m) = \begin{cases} \frac{1}{r}\pi_t(m), & \text{if } m_1 > m_2 \\ (1 - \Pi_{t,1}/r)\frac{\pi_t(m)}{1-\Pi_{t,1}}, & \text{if } m_1 \leq m_2 \text{ and } \Pi_{t,1} < 1. \\ (1 - 1/r)2, & \text{if } m_1 \leq m_2 \text{ and } \Pi_{t,1} = 1. \end{cases} \tag{A.1}$$

We will prove the theorem by the following steps:

- In Lemma 16 we show that Q_t's are valid distributions.
- In Lemma 17 we show that when $\alpha > 0$ the α-divergence between Q_t and Π_t can be arbitrarily small.
- In Lemma 18 we show that sampling from Q_t for $\Theta(T)$ time-steps will generate linear frequentist regret, and lower bound the regret.

Since the regret is linear, in Appendix A.1.4 we discuss the constant average regret per time-step as a function of ϵ and α. In Appendix A.1.5 we provide the Bayesian regret proof for Corollary 1.

Lemma 16. *Q_t defined in Eq. A.1 satisfies:*

$$\int q_t(m)dm = 1. \tag{A.2}$$

Lemma 17. *When $\alpha > 0$, for all $\epsilon > 0$, for all Π_t, there exists $r > 1$ such that when Q_t's are constructed from r as shown in Eq. A.1, $D_\alpha(\Pi_t, Q_t) < \epsilon$.*

Lemma 18. *The expected frequentist regret of the policy that constructs Q_t's as in Eq. A.1 and sample from Q_t for $T' = \Theta(T)$ time-steps is linear and the lower bound of the average regret per time-step is*

$$L = \begin{cases} c\Delta(1 - (1 - \epsilon\alpha(1-\alpha))^{\frac{1}{1-\alpha}}), & \text{when } \alpha > 1 \text{ and } 0 < \epsilon \\ c\Delta(1 - \frac{1}{e^\epsilon}), & \text{when } \alpha = 1 \text{ and } 0 < \epsilon \\ c\Delta(1 - (1 - \epsilon\alpha(1-\alpha))^{\frac{1}{1-\alpha}}), & \text{when } 0 < \alpha < 1 \text{ and } 0 < \epsilon \leq \frac{1}{\alpha(1-\alpha)} \end{cases}, \tag{A.3}$$

where $c = \frac{T'}{T}$ is $\Theta(1)$.

A.1.1 Proof of Lemma 16

Proof. If $\Pi_{t,1} < 1$:

$$\begin{aligned} \int q_t(m)dm &= \int_{\Omega_1} q_t(m)dm + \int_{\Omega_2} q_t(m)dm && \text{(A.4)} \\ &= \int_{\Omega_1} \frac{1}{r}\pi_t(m)dm + \int_{\Omega_2} \frac{1 - \Pi_{t,1}/r}{1 - \Pi_{t,1}}\pi_t(m)dm && \text{(A.5)} \\ &= \frac{1}{r}\Pi_{t,1} + \frac{1 - \Pi_{t,1}/r}{1 - \Pi_{t,1}}\Pi_{t,2} && \text{(A.6)} \\ &= \frac{1}{r}\Pi_{t,1} + \frac{1 - \Pi_{t,1}/r}{1 - \Pi_{t,1}}(1 - \Pi_{t,1}) && \text{(A.7)} \\ &= 1\ . && \text{(A.8)} \end{aligned}$$

If $\Pi_{t,1} = 1$:

$$\int q_t(m)dm = \int_{\Omega_1} q_t(m)dm + \int_{\Omega_2} q_t(m)dm \tag{A.9}$$

$$= \int_{\Omega_1} \frac{1}{r}\pi_t(m)dm + \int_{\Omega_2} 2(1-1/r)dm \tag{A.10}$$

$$= \frac{1}{r}\Pi_{t,1} + (1-1/r) \tag{A.11}$$

$$= 1\ . \tag{A.12}$$

□

A.1.2 Proof of Lemma 17

Proof. First we calculate the α-divergence between Π_t and Q_t constructed in Eq. A.1. We now upper bound $D_\alpha(\Pi_t, Q_t)$ when $\Pi_{t,1} < 1$ and $\Pi_{t,1} = 1$.

- $\Pi_{t,1} < 1$.

  Case 1: $\alpha > 0, \alpha \neq 1$**.** The following inequality is true by simple calculations when $0 < \alpha < 1$ or $\alpha > 1$:

$$\frac{\left(\frac{1-\Pi_{t,1}}{1-\frac{\Pi_{t,1}}{r}}\right)^{\alpha-1}}{\alpha(\alpha-1)} \leq \frac{r^{\alpha-1}}{\alpha(\alpha-1)}\ . \tag{A.13}$$

  We have:

$$D_\alpha(\Pi_t, Q_t) = \frac{1 - \int \left(\frac{\pi_t(m)}{q_t(m)}\right)^\alpha q_t(m)dm}{\alpha(1-\alpha)} \tag{A.14}$$

$$= \frac{1 - \int_{\Omega_1} \left(\frac{\pi_t(m)}{q_t(m)}\right)^\alpha q_t(m)dm - \int_{\Omega_2} \left(\frac{\pi_t(m)}{q_t(m)}\right)^\alpha q_t(m)dm}{\alpha(1-\alpha)} \tag{A.15}$$

$$= \frac{1 - \int_{\Omega_1} r^\alpha q_t(m)dm - \int_{\Omega_2} \left(\frac{1-\Pi_{t,1}}{1-\Pi_{t,1}/r}\right)^\alpha q_t(m)dm}{\alpha(1-\alpha)} \tag{A.16}$$

$$= \frac{1 - Q_t(\Omega_1)r^\alpha - Q_t(\Omega_2)\left(\frac{1-\Pi_{t,1}}{1-\Pi_{t,1}/r}\right)^\alpha}{\alpha(1-\alpha)} \tag{A.17}$$

$$= \frac{1 - \frac{\Pi_{t,1}}{r}r^\alpha - (1-\frac{\Pi_{t,1}}{r})\left(\frac{1-\Pi_{t,1}}{1-\Pi_{t,1}/r}\right)^\alpha}{\alpha(1-\alpha)} \tag{A.18}$$

$$= \frac{1}{\alpha(1-\alpha)}\left(1 - \Pi_{t,1}r^{-1+\alpha} - (1-\Pi_{t,1})^\alpha(1-\frac{\Pi_{t,1}}{r})^{1-\alpha}\right) \tag{A.19}$$

$$= \frac{\Pi_{t,1}r^{\alpha-1} + (1-\Pi_{t,1})\left(\frac{1-\Pi_{t,1}}{1-\frac{\Pi_{t,1}}{r}}\right)^{\alpha-1} - 1}{\alpha(\alpha-1)} \tag{A.20}$$

$$\leq \frac{1}{\alpha(\alpha-1)}\left(\Pi_{t,1}r^{\alpha-1} + (1-\Pi_{t,1})r^{\alpha-1} - 1\right) \text{ by Eq. A.13} \tag{A.21}$$

$$= \frac{1}{\alpha(\alpha-1)}\left(r^{-1+\alpha} - 1\right) \ . \tag{A.22}$$

Case 2: $\alpha = 1$**.**

$$D_\alpha(\Pi_t, Q_t) = \int \pi_t(m)\log\left(\frac{\pi_t(m)}{q_t(m)}\right)dm \tag{A.23}$$

$$= \int_{\Omega_1} \pi_t(m)\log\frac{\pi_t(m)}{q_t(m)}dm + \int_{\Omega_2} \pi_t(m)\log\frac{\pi_t(m)}{q_t(m)}\ dm \tag{A.24}$$

$$= \int_{\Omega_1} \pi_t(m)\log(r)dm + \int_{\Omega_2} \pi_t(m)\log\frac{1-\Pi_{t,1}}{1-\Pi_{t,1}/r}dm \tag{A.25}$$

$$= \Pi_{t,1}\log(r) + (1-\Pi_{t,1})\log\frac{1-\Pi_{t,1}}{1-\Pi_{t,1}/r}\ . \tag{A.26}$$

$$\leq \Pi_{t,1}\log(r) + (1-\Pi_{t,1})\log(r) \text{ because } r > 1 \tag{A.27}$$

$$\leq \log(r)\ . \tag{A.28}$$

- $\Pi_{t,1} = 1$. Since $\Pi_{t,2} = 0$, $\int_{\Omega_2} \pi_t(m)dm = 0$ and therefore $\pi_t(m) = 0$ almost everywhere on Ω_2 (Royden, 2010).

  Case 1: $\alpha > 0, \alpha \neq 1$**.**

$$
\begin{aligned}
D_\alpha(\Pi_t, Q_t) =& \frac{1 - \int \left(\frac{\pi_t(m)}{q_t(m)}\right)^\alpha q_t(m)dm}{\alpha(1-\alpha)} && \text{(A.29)}\\
=& \frac{1 - \int_{\Omega_1} \left(\frac{\pi_t(m)}{q_t(m)}\right)^\alpha q_t(m)dm - \int_{\Omega_2} \left(\frac{\pi_t(m)}{q_t(m)}\right)^\alpha q_t(m)dm}{\alpha(1-\alpha)} && \text{(A.30)}\\
=& \frac{1 - \int_{\Omega_1} r^\alpha q_t(m)dm - 0}{\alpha(1-\alpha)} && \text{(A.31)}\\
=& \frac{1 - Q_t(\Omega_1) r^\alpha}{\alpha(1-\alpha)} && \text{(A.32)}\\
=& \frac{1 - \frac{1}{r} r^\alpha}{\alpha(1-\alpha)} && \text{(A.33)}\\
=& \frac{1 - r^{\alpha-1}}{\alpha(1-\alpha)}. && \text{(A.34)}\\
& && \text{(A.35)}
\end{aligned}
$$

  Case 2: $\alpha = 1$**.**

$$
\begin{aligned}
D_\alpha(\Pi_t, Q_t) =& \int \pi_t(m) \log\left(\frac{\pi_t(m)}{q_t(m)}\right) dm && \text{(A.36)}\\
=& \int_{\Omega_1} \pi_t(m) \log \frac{\pi_t(m)}{q_t(m)} dm + \int_{\Omega_2} \pi_t(m) \log \frac{\pi_t(m)}{q_t(m)}\, dm && \text{(A.37)}\\
=& \int_{\Omega_1} \pi_t(m) \log(r) dm + 0 && \text{(A.38)}\\
=& \log(r)\ . && \text{(A.39)}
\end{aligned}
$$

For both cases when $\Pi_{t,1} = 1$ and $\Pi_{t,1} < 1$, $D_\alpha(\Pi_t, Q_t)$ is upper bounded by:

$$
\begin{cases}
\frac{1-r^{\alpha-1}}{\alpha(1-\alpha)}, & \text{if } 0 < \alpha < 1 \text{ or } \alpha > 1 \\
\log(r), & \text{if } \alpha = 1\ .
\end{cases}
\tag{A.40}
$$

Since $\lim_{r\to 1}\log(r) = 0$ and $\lim_{r\to 1}\frac{1-r^{-1+\alpha}}{\alpha(1-\alpha)} = 0$, for any $\epsilon > 0$, there exists $r > 1$ such that

$$D_\alpha(\Pi_t, Q_t) \leq \epsilon \ . \tag{A.41}$$

□

A.1.3 Proof of Lemma 18

Proof. We will now lower bound the regret as a function of ϵ.

For any posterior Π_t, since the approximate algorithm sampling from Q_t picks the optimal arm with probability at most $1/r$ it then picks arm 2 with probability at least $1 - 1/r$.

Since we sample from Q_t for T' time steps, the lower bound of the average expected regret per time step is :

$$L = \frac{T'}{T}(m_1^* - m_2^*)(1 - 1/r) = c\Delta(1 - 1/r) \ . \tag{A.42}$$

where $\Delta = m_1^* - m_2^*$ and $c = \frac{T'}{T}$ is $\Theta(1)$.

We calculate ϵ as a function of r from Eq. A.40:

$$\epsilon = \begin{cases} \frac{1-r^{-1+\alpha}}{\alpha(1-\alpha)}, & \text{if } \alpha \neq 1 \\ log(r), & \text{if } \alpha = 1 \ . \end{cases} \tag{A.43}$$

The functions are continuous and increasing when $r > 1$. Then by direct calculations when $r \to \infty$ and $r \to 1$, the domain of ϵ is:

$$0 < \epsilon \text{ when } \alpha \geq 1 \ . \tag{A.44}$$

$$0 < \epsilon < \frac{1}{\alpha(1-\alpha)} \text{ when } 0 < \alpha < 1 \ . \tag{A.45}$$

Then

$$r = \begin{cases} (1-\epsilon\alpha(1-\alpha))^{\frac{1}{-1+\alpha}} & \text{when } \alpha > 1 \text{ and } 0 < \epsilon \\ e^{\epsilon} & \text{when } \alpha = 1 \text{ and } 0 < \epsilon \\ (1-\epsilon\alpha(1-\alpha))^{\frac{1}{-1+\alpha}} & \text{when } 0 < \alpha < 1 \text{ and } 0 < \epsilon \leq \frac{1}{\alpha(1-\alpha)}. \end{cases} \tag{A.46}$$

Therefore we can calculate the lower bound of the regret per time-step as:

$$L = \begin{cases} c\Delta(1-(1-\epsilon\alpha(1-\alpha))^{\frac{1}{1-\alpha}}), & \text{when } \alpha > 1 \text{ and } 0 < \epsilon \\ c\Delta(1-\frac{1}{e^{\epsilon}}), & \text{when } \alpha = 1 \text{ and } 0 < \epsilon \\ c\Delta(1-(1-\epsilon\alpha(1-\alpha))^{\frac{1}{1-\alpha}}), & \text{when } 0 < \alpha < 1 \text{ and } 0 < \epsilon \leq \frac{1}{\alpha(1-\alpha)}. \end{cases} . \tag{A.47}$$

We plot the lower bound of the average regret per time step when $\Delta = 0.1$ as a function of ϵ in Fig A.1.

□

A.1.4 The Average Regret per Time-step

To understand how the constant average regret per time-step depends on ϵ and α, we plot in Figure A.1 the lower bound of the average regret per time-step in Lemma 18 as a function of ϵ in the following setting of the example constructed in the proof of Theorem 1. The algorithm samples from Q_t at $T/2$ time-steps and $\Delta = 0.1$. In this case the average regret per time step is upper bounded by $\Delta/2 = 0.05$. The formula and proof are detailed in Lemma 18 in Appendix A.1. When α is around 1, the lower bound, and therefore the average regret per time-step, converges the fastest to $\Delta/2$ as ϵ increases. When α is very large or close to 0, the lower bound grows slowly as ϵ increases.

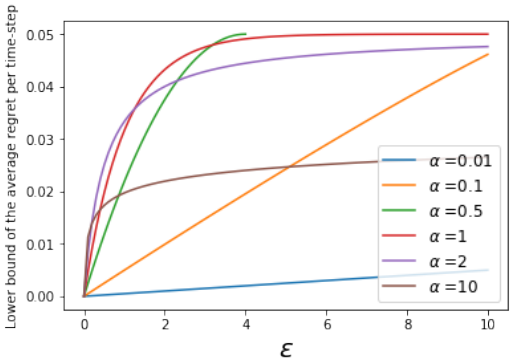


Figure A.1: Lower bound of regret per time-step as a function of ϵ when $m_1^* - m_2^* = 0.1$ and we sample from the approximation for $T/2$ time-steps in the example construted in the proof of Theorem 1. When α is around 1, the lower bound converges quickly as ϵ increases.

A.1.5 Proof of Corollary 1

Since $\mathbb{P}(M_1 > M_2) > 0$, there exist constants $\Delta > 0, \gamma > 0$ such that $\mathbb{P}(M_1 - M_2 \geq \Delta) = \gamma$. The probability that the assumption $m_1^* > m_2^*$ in Theorem 1 is satisfied is at least $\gamma > 0$. Therefore the expected regret over the prior is at least γ times the frequentist regret in Theorem 1, which is linear.

A.2 Proof of Theorem 2 and Corollary 2

First we will prove Theorem 2. Let $\Pi_{t,i}$ denote $\Pi_t(\Omega_i)$. Suppose that $\Pi_{t,2} > 0$ for all $t \geq 0$, we construct the pdf of Q_t's as follows:

$$q_t(m) = \begin{cases} 0, \text{ if } m_1 > m_2 \\ \frac{1}{\Pi_{t,2}} \pi_t(m), \text{ if } m_1 \leq m_2 \end{cases} \tag{A.48}$$

We will prove the theorem by the following steps:

- In Lemma 19 we show that Q_t's are valid distributions.
- In Lemma 20 we show that Q_t has linear frequentist regret, and calculate the constant average regret per time-step.

- In Lemma 21 we show that there exists a bad prior such that the α-divergence between Q_t and Π_t can be arbitrarily small.

In Appendix A.2.4 we discuss the prior-dependent error threshold ϵ that will cause linear regret. In Appendix A.2.5 we provide the Bayesian regret proof for Corollary 2.

Lemma 19. *Q_t constructed in Eq. A.48 satisfies:*

$$\int q_t(m)dm = 1. \tag{A.49}$$

Lemma 20. *Q_t constructed in Eq. A.48 chooses arm 2 at all time-steps. The average frequentist regret per time-step is $\Delta = m_1^* - m_2^*$.*

Lemma 21. *Let $\alpha < 1$, $M_1 - M_2$ and M_2 be independent and arm 2 be chosen at all time-steps before t.*

For any $\epsilon > 0$, there exists $0 < z < 1$ such that if $\Pi_{0,2} = z$ then $D_\alpha(\Pi_t, Q_t) < \epsilon$ where Q_t is constructed in Eq. A.48.

For any $0 < z < 1$, there exists $\epsilon > 0$ such that if $\Pi_{0,2} = z$ then $D_\alpha(\Pi_t, Q_t) < \epsilon$ where Q_t is constructed in Eq. A.48.

A.2.1 Proof of Lemma 19

Proof. Let $D = M_1 - M_2$ which is independent of M_2 by the assumption. Since the algorithm always picks arm 2, H_{t-1} only depends on M_2, and therefore H_{t-1} and D are independent.

Since D and H_{t-1} are independent, at all times t the posterior does not concentrate:

$$\Pi_{t,2} = \mathbb{P}(M_1 - M_2 < 0 | H_{t-1}) = \mathbb{P}(M_1 < M_2)\ . \tag{A.50}$$

Since $\Pi_{t,2} > 0$ for all t,$q_t(m)$ is well-defined. We have:

$$\int q_t(m)dm = \int_{\Omega_1} q_t(m)dm + \int_{\Omega_2} q_t(m)dm \tag{A.51}$$

$$= 0 + \int_{\Omega_2} \frac{1}{\Pi_{t,2}} \pi_t(m)dm \tag{A.52}$$

$$= \frac{1}{\Pi_{t,2}} \int_{\Omega_2} \pi_t(m)dm \tag{A.53}$$

$$= 1 \ . \tag{A.54}$$

□

A.2.2 Proof of Lemma 20

Proof. Under the approximate distribution, arm 2 is chosen with probability 1 at all times. Clearly this approximate distribution has linear regret, with $\Delta = m_1^* - m_2^*$ being the average regret per time-step. □

A.2.3 Proof of Lemma 21

Proof. From the proof of Lemma 19:

$$\Pi_{t,2} = \mathbb{P}(M_1 - M_2 < 0 | H_{t-1}) = \mathbb{P}(M_1 < M_2) \ . \tag{A.55}$$

For simplicity let

$$z := \mathbb{P}(M_1 < M_2) \ . \tag{A.56}$$

First we calculate the α-divergence between Π_t and Q_t constructed in Eq A.48.

- When $\alpha < 1, \alpha \neq 0$:

$$D_\alpha(\Pi_t, Q_t) = \frac{1 - \int \left(\frac{q_t(m)}{\pi_t(m)}\right)^{1-\alpha} \pi_t(m) dm}{\alpha(1-\alpha)} \tag{A.57}$$

$$= \frac{1 - \int_{\Omega_1} \left(\frac{q_t(m)}{\pi_t(m)}\right)^{1-\alpha} \pi_t(m) dm - \int_{\Omega_2} \left(\frac{q_t(m)}{\pi_t(m)}\right)^{1-\alpha} \pi_t(m) dm}{\alpha(1-\alpha)} \tag{A.58}$$

$$= \frac{1 - 0 - \int_{\Omega_2} \left(\frac{1}{\Pi_{t,2}}\right)^{1-\alpha} \pi_t(m) dm}{\alpha(1-\alpha)} \text{ since } \alpha < 1 \tag{A.59}$$

$$= \frac{1 - \left(\frac{1}{\Pi_{t,2}}\right)^{1-\alpha} \int_{\Omega_2} \pi_t(m) dm}{\alpha(1-\alpha)} \tag{A.60}$$

$$= \frac{1 - \left(\frac{1}{\Pi_{t,2}}\right)^{1-\alpha} \Pi_{t,2}}{\alpha(1-\alpha)} \tag{A.61}$$

$$= \frac{1 - (\Pi_{t,2})^{\alpha}}{\alpha(1-\alpha)} \tag{A.62}$$

$$= \frac{1 - z^{\alpha}}{\alpha(1-\alpha)} \,. \tag{A.63}$$

Since $\lim_{z \to 1} \frac{1-z^\alpha}{\alpha(1-\alpha)} = 0$, for any $\epsilon > 0$ there exists $0 < z < 1$ such that $D_\alpha(\Pi_t, Q_t) < \epsilon$.

For any $0 < z < 1$ there exists $\epsilon = \frac{1-z^\alpha}{\alpha(1-\alpha)} > 0$ such that $D_\alpha(\Pi_t, Q_t) < \epsilon$.

- When $\alpha = 0$:

$$D_\alpha(\Pi_t, Q_t) = \int q_t(m) \log \frac{q_t(m)}{\pi_t(m)} dm \tag{A.64}$$

$$= \int_{\Omega_1} q_t(m) \log \frac{q_t(m)}{\pi_t(m)} dm + \int_{\Omega_2} q_t(m) \log \frac{q_t(m)}{\pi_t(m)} dm \tag{A.65}$$

$$= \int_{\Omega_1} 0 \log(0) dm + \int_{\Omega_2} q_t(m) \log \frac{1}{\Pi_{t,2}} dm \tag{A.66}$$

$$= 0 + 1 \log \frac{1}{\Pi_{t,2}} = \log \frac{1}{\Pi_{t,2}} = \log \frac{1}{z} \,. \tag{A.67}$$

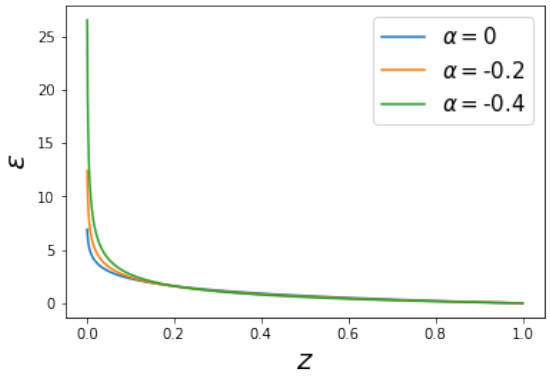
Since $\lim_{z \to 1} \log(1/z) = 0$, for any $\epsilon > 0$ there exists $0 < z < 1$ such that $D_0(\Pi_t, Q_t) < \epsilon$.

For any $z, 0 < z < 1$ there exists $\epsilon = \log \frac{1}{z} \geq 0$ such that $D_\alpha(\Pi_t, Q_t) < \epsilon$.

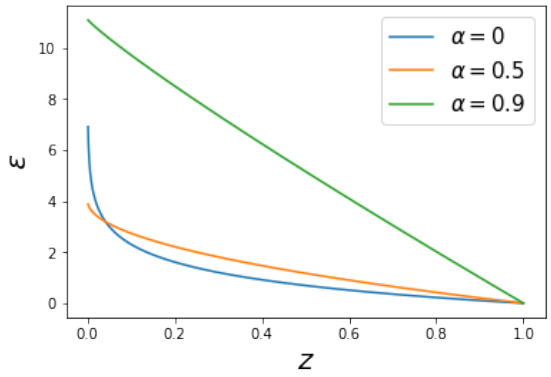
□

A.2.4 Prior-dependent Error Threshold for Linear Frequentist Regret

In the example constructed in the previous sections, the α-divergence between Π_t and Q_t can be calculated as: $\epsilon = \begin{cases} \frac{1-z^\alpha}{\alpha(1-\alpha)}, & \text{if } 0 < \alpha < 1 \text{ or } \alpha < 0 \\ \log \frac{1}{z}, & \text{if } \alpha = 0 \end{cases}$.



(a) $D_\alpha(\Pi_t, Q_t) = \epsilon$ as a function of z that can cause linear regret for some $\alpha \leq 0$.



(b) ϵ as a function of z that can cause linear regret for some $\alpha \in [0, 1)$.

Figure A.2: ϵ as a function z that makes the regret linear for different values of α for the example constructed in the proof of Theorem 2.

In Figure A.2, we show the values of ϵ as a function of z that will make the regret linear for different values of α. We can see that for both cases when $\alpha \leq 0$ and $0 \leq \alpha < 1$, and z is not too small, there is a threshold of ϵ for each value of z that makes the regret linear. For each value of z, if the error is smaller than the threshold we hypothesize that the regret might become sub-linear. However even if that is the case, it is not possible to calculate the exact threshold for more complicated priors. Therefore in Section 1.5.1 we propose an algorithm that is guaranteed to have sub-linear regret for any ϵ and any z when $\alpha \leq 0$.

A.2.5 Proof of Corollary 2

Since $\mathbb{P}(M_1 > M_2) > 0$, there exist constants $\Delta > 0, \gamma > 0$ such that the $\mathbb{P}(M_1 - M_2 \geq \Delta) = \gamma$. The probability that the assumption $m_1^* > m_2^*$ in Theorem 2 is satisfied is at least $\gamma > 0$. Therefore the expected regret over the prior is at least γ times the frequentist regret in Theorem 2, which is linear.

A.3 Proof of Lemma 2

To convert between $D_\alpha(\Pi_t, Q_t)$ and $D_\alpha(\overline{\Pi_t}, \overline{Q_t})$ we first prove the following lemma:

Lemma 22 (Jensen's Inequality)**.** [1] *Let $f : \mathcal{R}^2 \to \mathcal{R}$ be a convex function. Let $P : \mathcal{R}^k \to \mathcal{R}$ and $Q : \mathcal{R}^k \to \mathcal{R}$ be 2 functions. Let S is a subset of R^k, the domain of x and $|S|$ denote the volume of S. Then*

$$\frac{1}{|S|}\int_S f(P(x), Q(x))dx \geq f\left(\frac{1}{|S|}\int_S P(x)dx, \frac{1}{|S|}\int_S Q(x)dx\right) . \tag{A.68}$$

Proof. The multivariate Jensen's Inequality states that if X is a n-dimensional random vector and $f : \mathcal{R}^n \to \mathcal{R}$ is a convex function then

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X)) . \tag{A.69}$$

To use the multivariate Jensen's Inequality we define the 2-dimensional random vector $X : S \to \mathcal{R}^2$ by $X(x) := (P(x), Q(x))$ and a probability distribution over S with the density $d(x) = \frac{1}{|S|}$ for all $x \in S$:

Then the left-hand side of Eq. A.68 becomes $\mathbb{E}(f(X))$, while the right-hand side becomes $f(\mathbb{E}(X))$, and Eq. A.68 follows from the multivariate Jensen's Inequality. □

Now we will prove Lemma 2.

[1] We thank Huy Le for providing the proof of Lemma 22.

Proof of Lemma 2. Since $D_\alpha(p,q)$ is convex (Cichocki and Amari, 2010), the following functions:

$$f(p,q) = q \log \frac{q}{p}, \tag{A.70}$$

$$f(p,q) = p \log \frac{p}{q}, \tag{A.71}$$

$$f(p,q) = \frac{p^\alpha q^{1-\alpha}}{\alpha(\alpha-1)} \tag{A.72}$$

are convex, and we can apply Lemma 22:

- When $\alpha = 0$:

$$\begin{aligned}
D_\alpha(\Pi_t, Q_t) &= \int q_t(m) \log \frac{q_t(m)}{\pi_t(m)} dm && \text{(A.73)}\\
&= \sum_i \int_{\Omega_i} q_t(m) \log \frac{q_t(m)}{\pi_t(m)} dm && \text{(A.74)}\\
&\geq \sum_i |\Omega_i| \frac{1}{|\Omega_i|} \int_{\Omega_i} q_t(m) dm \log \frac{\frac{1}{|\Omega_i|}\int_{\Omega_i} q_t(m) dm}{\frac{1}{|\Omega_i|}\int_{\Omega_i} \pi_t(m) dm} \text{ by Lemma 22} && \text{(A.75)}\\
&= \sum_i Q_{t,i} \log \frac{Q_{t,i}}{\Pi_{t,i}} && \text{(A.76)}\\
&= D_\alpha(\overline{\Pi_t}, \overline{Q_t}) \ . && \text{(A.77)}
\end{aligned}$$

- When $\alpha = 1$:

$$D_\alpha(\Pi_t, Q_t) = \int \pi_t(m) \log \frac{\pi_t(m)}{q_t(m)} dm \tag{A.78}$$

$$= \sum_i \int_{\Omega_i} \pi_t(m) \log \frac{\pi_t(m)}{q_t(m)} dm \tag{A.79}$$

$$\geq \sum_i |\Omega_i| \frac{1}{|\Omega_i|} \int_{\Omega_i} \pi_t(m) dm \log \frac{\frac{1}{|\Omega_i|}\int_{\Omega_i} \pi_t(m) dm}{\frac{1}{|\Omega_i|}\int_{\Omega_i} q_t(m) dm} \text{ by Lemma 22} \tag{A.80}$$

$$= \sum_i \Pi_{t,i} \log \frac{\Pi_{t,i}}{Q_{t,i}} \tag{A.81}$$

$$= D_\alpha(\overline{\Pi_t}, \overline{Q_t}) \ . \tag{A.82}$$

- When $\alpha \neq 0, \alpha \neq 1$:

$$D_\alpha(\Pi_t, Q_t) = \int \frac{\pi(x)^\alpha q(x)^{1-\alpha} - 1}{-\alpha(1-\alpha)} dx \tag{A.83}$$

$$= \frac{-1}{\alpha(\alpha-1)} + \sum_i \int_{\Omega_i} \frac{\pi(x)^\alpha q(x)^{1-\alpha}}{\alpha(\alpha-1)} dx \tag{A.84}$$

$$\geq \frac{-1}{\alpha(\alpha-1)} + \sum_i |\Omega_i| \frac{(\frac{\Pi_{t,i}}{|\Omega_i|})^\alpha (\frac{Q_{t,i}}{|\Omega_i|})^{1-\alpha}}{\alpha(\alpha-1)} \text{ by Lemma 22} \tag{A.85}$$

$$= \frac{-1}{\alpha(\alpha-1)} + \sum_i \frac{\Pi_{t,i}^\alpha Q_{t,i}^{1-\alpha}}{\alpha(\alpha-1)} \tag{A.86}$$

$$= D_\alpha(\overline{\Pi_t}, \overline{Q_t}) \ . \tag{A.87}$$

□

A.4 Proof of Theorem 3

We will prove that the frequentist regret is sub-linear for any m^*. If the algorithm has sub-linear frequentist regret for all values $M = m^*$, the Bayesian regret (which is the expected value over M) will also be sub-linear.

Without loss of generalization, let arm 1 be the best arm. From Lemma 1, since $\sum_{t=1}^{\infty} p_t = \infty$, we have for all arms i, $\sum_{t=1}^{\infty} P(A_t = i|H_{t-1}) = \infty$ and therefore with probability 1:

$$\lim_{t\to\infty} \Pi_{t,1} = \lim_{t\to\infty} \mathbb{P}(A^* = 1|H_{t-1}) = 1 \ , \tag{A.88}$$

which means that the posterior probability that arm 1 is the best arm converges to 1.

We will prove the theorem by proving the following steps:

- In Lemma 23 we show that if the probability that the posterior chooses the best arm tends to 1, then the probability that the approximation chooses the best arm also tends to 1
- In Lemma 24 and Lemma 25 we show that if the probability that the approximation chooses the best arm also tends to 1 almost surely, then it has sub-linear regret with probability 1. Therefore it has sub-linear regret in expectation over the history.

Lemma 23. *Let $\alpha \leq 0$ and arm 1 be the true best arm. Let $\Omega_i = \{m|m_i = max(m_1, ..., m_k)\}$ be the region where arm i is the best arm. If the posterior probability that arm 1 is the best arm converges to 1:*

$$\lim_{t\to\infty} \Pi_{t,1} = 1 \tag{A.89}$$

and for all $t \geq 0$:

$$D_\alpha(\Pi_t, Q_t) < \epsilon, \tag{A.90}$$

then the sequence $\{Q_{t,1}\}_t$ where $Q_{t,1} = \int_{\Omega_1} q_t(m)dm$ converges and

$$\lim_{t\to\infty} Q_{t,1} = 1 \ . \tag{A.91}$$

Next we show that if the approximate distribution concentrates, then the probability that it chooses the wrong arm decreases as T goes to infinity.

Lemma 24. *If*

$$\lim_{t\to\infty} Q_{t,1} = 1 \tag{A.92}$$

then

$$\lim_{T\to\infty} \frac{\sum_{t=1}^{T}(1-Q_{t,1})}{T} = 0\ . \tag{A.93}$$

From Lemma 23 and Lemma 24, since $\lim_{t\to\infty}\Pi_{t,1} = 1$ with probability 1, we have $\lim_{T\to\infty}\frac{\sum_{t=1}^{T}(1-Q_{t,1})}{T} = 0$ with probability 1. We will now show that the expected regret is sub-linear:

Lemma 25. *Let $p_t = o(1)$ be such that $\sum_{t=1}^{\infty} p_t = \infty$. For any number of arms k, any prior Π_0 and any error threshold $\epsilon > 0$, the following algorithm has $o(T)$ regret: at every time-step t,*

- *with probability $1 - p_t$, sample from an approximate posterior Q_t such that $\lim_{T\to\infty}\frac{\sum_{t=1}^{T}(1-Q_{t,1})}{T} = 0$ with probability 1, and*
- *with probability p_t, sample an arm uniformly at random.*

A.4.1 Proof of Lemma 23

Proof. Let $Q_{t,i} = \int_{\Omega_i} q_t(m)dm$ and $\Pi_{t,i} = \int_{\Omega_i} \pi_t(m)dm$. Then

$$\lim_{t\to\infty} \Pi_{t,1} = 1 \tag{A.94}$$

and we want to show that $\{Q_{t,1}\}_t$ converges and

$$\lim_{t\to\infty} Q_{t,1} = 1\ . \tag{A.95}$$

Since $D_\alpha(\overline{\Pi_t}, \overline{Q_t}) < \epsilon$ and $\lim \Pi_{t,1} = 1$ we want to show that $\limsup Q_{t,1} = 1$. By contradiction, assume that:

$$\limsup Q_{t,1} = c < 1 \ . \tag{A.96}$$

Then there exists a sub-sequence of $\{Q_{t,1}\}_t$, denoting $Q_{t_1,1}, Q_{t_2,1}, ..., Q_{t_n,1}, ..$ such that

$$\lim_{n\to\infty} Q_{t_n,1} = c \ . \tag{A.97}$$

which implies

$$0 < 1 - c = \lim_{n\to\infty} \sum_{i=2}^{k} Q_{t_n,i} \leq \sum_{i=2}^{k} \limsup_{n\to\infty} Q_{t_n,i}. \tag{A.98}$$

Therefore there exists $j \in [2, k]$ such that:

$$\limsup_{n\to\infty} Q_{t_n,j} = d > 0 \ . \tag{A.99}$$

Then there exists a sub-sequence of $\{Q_{t_n,j}\}_n$, denoting $Q_{t_{n_1},j}, Q_{t_{n_2},j}, ..., Q_{t_{n_m},j}, ..$ such that

$$\lim_{m\to\infty} Q_{t_{n_m},j} = d \ . \tag{A.100}$$

We consider the 2 cases:

- When $\alpha = 0$:

$$D_\alpha(\overline{\Pi}_t, \overline{Q}_t) = \sum_{i=1}^{k} Q_{t,i} \log \frac{Q_{t,i}}{\Pi_{t,i}} \ . \tag{A.101}$$

By Jensen's inequality we have:

$$\sum_i a_i \log \frac{b_i}{a_i} \leq (\sum_i a_i) \log \left(\frac{\sum_i a_i \frac{b_i}{a_i}}{\sum_i a_i} \right) \Longleftrightarrow \sum_i a_i \log \frac{a_i}{b_i} \geq (\sum_i a_i) \log \left(\frac{\sum_i a_i}{\sum_i b_i} \right). \tag{A.102}$$

Then we have:

$$\begin{aligned}
\epsilon &= \lim_{m\to\infty} D_\alpha(\overline{\Pi}_{t_{n_m}}, \overline{Q}_{t_{n_m}}) && \text{(A.103)}\\
&= \lim_{m\to\infty} \sum_{i:i\neq j, 1\leq i\leq k} Q_{t_{n_m},i} \log \frac{Q_{t_{n_m},i}}{\Pi_{t_{n_m},i}} + \lim_{m\to\infty} Q_{t_{n_m},j} \log \frac{Q_{t_{n_m},j}}{\Pi_{t_{n_m},j}} && \text{(A.104)}\\
&\geq \lim_{m\to\infty} (\sum_{i:i\neq j, 1\leq i\leq k} Q_{t_{n_m},i}) \log \frac{\sum_{i:i\neq j,1\leq i\leq k} Q_{t_{n_m},i}}{\sum_{i:i\neq j,1\leq i\leq k} \Pi_{t_{n_m},i}} + \lim_{m\to\infty} Q_{t_{n_m},j} \log \frac{Q_{t_{n_m},j}}{\Pi_{t_{n_m},j}} \\
& && \text{by Eq. A.102} \quad \text{(A.105)}\\
&= \lim_{m\to\infty} (1 - Q_{t_{n_m},j}) \log \frac{1 - Q_{t_{n_m},j}}{1 - \Pi_{t_{n_m},j}} + \lim_{m\to\infty} Q_{t_{n_m},j} \log \frac{Q_{t_{n_m},j}}{\Pi_{t_{n_m},j}} && \text{(A.106)}\\
&= (1-d) \log \frac{1-d}{1} + d \log \frac{d}{0} && \text{(A.107)}\\
&= \infty \text{ since } d > 0, && \text{(A.108)}
\end{aligned}$$

which is a contradiction. Therefore $c = 1$.

- When $\alpha < 0$:

$$D_\alpha(\overline{\Pi}_t, \overline{Q}_t) = \frac{\sum_{i=1}^k \Pi_{t,i}^\alpha Q_{t,i}^{1-\alpha} - 1}{\alpha(\alpha - 1)}. \tag{A.109}$$

Then we have:

$$\epsilon = \lim_{m\to\infty} D_\alpha(\overline{\Pi}_{t_{n_m}}, \overline{Q}_{t_{n_m}}) \tag{A.110}$$

$$= \frac{\lim_{m\to\infty} \sum_{i:i\neq j, 1\leq i\leq k} \Pi^{\alpha}_{t_{n_m},i} Q^{1-\alpha}_{t_{n_m},i} + \lim_{m\to\infty} \Pi^{\alpha}_{t_{n_m},j} Q^{1-\alpha}_{t_{n_m},j} - 1}{\alpha(\alpha-1)} \tag{A.111}$$

$$\geq \frac{0 + \frac{d^{1-\alpha}}{(0)^{-\alpha}} - 1}{\alpha(\alpha-1)} \tag{A.112}$$

$$= \infty, \text{ since } d > 0 \text{ and } \alpha < 0, \tag{A.113}$$

which is a contradiction. Therefore $c = 1$.

Similarly we will show that:

$$\liminf Q_{t,1} = 1 \ . \tag{A.114}$$

By contradiction, assume that:

$$\liminf Q_{t,1} = c' < 1 \ . \tag{A.115}$$

Then there exists a sub-sequence of $\{Q_{t,1}\}_t$, denoting $Q_{t_1,1}, Q_{t_2,1}, ..., Q_{t_{n'},1}, ..$ such that

$$\lim_{n\to\infty} Q_{t_{n'},1} = c' \ . \tag{A.116}$$

Using the same argument following Eq. A.97 we will have $c' = 1$. Since $\liminf Q_{t,1} = \limsup Q_{t,1} = 1$, we have that $\{Q_{t,1}\}_t$ converges and

$$\lim Q_{t,1} = 1 \ . \tag{A.117}$$

□

A.4.2 Proof for Lemma 24

For simplicity let x_t denote $1 - Q_{t,1}$. We want to show that if a sequence $\{x_t\}$ satisfies $x_t \geq 0 \ \forall t$ and:

$$\lim_{t\to\infty} x_t = 0, \tag{A.118}$$

then

$$\lim_{T\to\infty} S_T = 0, \tag{A.119}$$

where $S_T = \frac{\sum_{t=1}^{T} x_t}{T}$.

Since $\lim_{t\to\infty} x_t = 0$ and $x_t \geq 0 \ \forall t$, for any $\epsilon > 0$ there exists T_0 such that for all $t > T_0$:

$$x_t < \frac{\epsilon}{2} \, . \tag{A.120}$$

Then for all $T > T_0$:

$$S_T = \frac{x_1 + ... + x_{T_0}}{T} + \frac{x_{T_0+1} + ... + x_T}{T} \tag{A.121}$$

$$\leq \frac{x_1 + ... + x_{T_0}}{T} + \frac{\frac{\epsilon}{2}T}{T} \tag{A.122}$$

$$\leq \frac{x_1 + ... + x_{T_0}}{T} + \frac{\epsilon}{2} \, . \tag{A.123}$$

Choose T_1 large enough such that $\frac{x_1+...+x_{T_0}}{T_1} < \frac{\epsilon}{2}$. Let $T_2 = \max(T_0, T_1)$. Then for all $T > T_2$:

$$S_T = \frac{x_1 + ... + x_{T_0}}{T} + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \, . \tag{A.124}$$

Therefore for any $\epsilon > 0$, there exists T_2 such that for all $T > T_2$, $S_T < \epsilon$. Since $S_T \geq 0 \ \forall T$, we have:

$$\lim_{T \to \infty} S_T = 0 \ . \tag{A.125}$$

A.4.3 Proof of Lemma 25

Without loss of generalization, let arm 1 be the true best arm. Let $\Delta = m_1^* - \max(m_2^*, ..., m_k^*)$ be the gap between the highest mean m_1^* and the next highest mean of the arms.

Since $p_t = o(1)$, $\sum_{t=1}^{T} p_t$ is $o(T)$. Therefore the regret from the uniform sampling steps is $o(T)$.

Since $1 - Q_{t,1}$ is the probability of choosing a sub-optimal arm by sampling from Q_t, the regret of sampling from Q_t is upper bounded by:

$$\mathbb{E} \sum_{t=1}^{T} \Delta(1 - Q_{t,1}) \ . \tag{A.126}$$

Since $\lim_{T \to \infty} \frac{\sum_{t=1}^{T}(1-Q_{t,1})}{T} = 0$ with probability 1, we have

$$\lim_{T \to \infty} \frac{\sum_{t=1}^{T} \Delta(1 - Q_{t,1})}{T} = 0 \tag{A.127}$$

with probability 1. Therefore

$$\lim_{T \to \infty} \mathbb{E} \frac{\sum_{t=1}^{T} \Delta(1 - Q_{t,1})}{T} = 0, \tag{A.128}$$

which means that the regret of sampling from Q_t is sub-linear. Since both the expected regrets of the uniform sampling steps and of sampling from Q_t are sub-linear, the total expected regret is sub-linear.

A.5 Ensemble Sampling and Uniform Exploration

To the best of our knowledge, (Lu and Van Roy, 2017) is the only work that provides a theoretical analysis of Thompson sampling when the sampling step is approximate. Lu and Van Roy (2017) propose an approximate sampling method called Ensemble sampling where they maintain a set of $\mathcal{M}$ models to approximate the posterior, and analyze its regret for linear contextual bandits. When the model is a k-armed bandit, the regret bound is as follow:

Lemma 26 (implied by Lu and Van Roy (2017))**.** *Let π^{TS} and π^{ES} denote the exact Thompson sampling and Ensemble sampling policies. Let $\Delta = max(m_1^*, ..., m_k^*) - \min(m_1^*, ..., m_k^*)$. For all $\epsilon > 0$, if*

$$\mathcal{M} \geq \frac{2k}{\epsilon^2} \log \frac{2kT}{\epsilon^2 \delta}, \tag{A.129}$$

then

$$\text{Regret}(T, \pi^{ES}) \leq \text{Regret}(T, \pi^{TS}) + \epsilon \Delta T + \delta \Delta T. \tag{A.130}$$

(Lu and Van Roy, 2017) prove the regret bound by only using the following property of the Ensemble sampling method: at time t, with probability $1-\delta$, Ensemble sampling satisfies the following constraint:

$$\text{KL}(\overline{Q}_t, \overline{\Pi}_t) < \epsilon^2, \tag{A.131}$$

where ϵ is a constant if $\mathcal{M}$ is $\Theta(\log(T))$. If ϵ is a constant the regret will be linear because of the term $\epsilon \Delta T$.

At time t, with probability $1 - \delta$, $\text{KL}(\overline{Q}_t, \overline{\Pi}_t) < \epsilon^2$. The first 2 terms in the right hand side of Eq. A.130 comes from the time-steps when $\text{KL}(\overline{Q}_t, \overline{\Pi}_t) < \epsilon^2$, and the last term comes from the other case with probability δ.

Theorem 3 shows that applying an uniform sampling step will make the posterior concentrate and the regret sub-linear.

So if we want to maintain a small number of models $M = \Theta(\log(T))$ and achieve sub-linear regret, we can apply Theorem 3 as follow. First we choose δ to be small such that the last term in Eq. A.130 $\delta\Delta T$ is $o(T)$. Then we apply the uniform sampling step as shown in Theorem 3, so that the first 2 terms in the right hand side of Eq. A.130 become sub-linear. We can then achieve sub-linear regret with Ensemble sampling with a $\Theta(\log T)$ number of models.

A.6 Posterior Calculation

In our simulations, when both the prior and the reward distributions are Gaussian, we calculate the true posterior using the following closed-form solution.

Let the posterior at time t be multivariate Gaussian distribution $\text{Norm}(\mu_t, \Sigma_t)$ where μ_t is a $k \times 1$ vector and Σ_t is a $k \times k$ covariance matrix. Let the reward distribution of arm i be $\text{Norm}(m_i^*, \sigma^2)$ where σ is known and m_i^*'s are unknown.

Let $A_t \in \{0, 1\}^k$ be a 0/1 vector where $A_t(i) = 1$ if arm i is chosen at time t, and 0 otherwise. Let $r_t \in \mathcal{R}$ be the reward of the arm chosen at time t.

Then the posterior at time $t + 1$ is $\text{Norm}(\mu_{t+1}, \Sigma_{t+1})$ where:

$$\Sigma_{t+1} = (\Sigma_t^{-1} + A_t A_t^T/\sigma^2)^{-1} \tag{A.132}$$

$$\mu_{t+1} = \Sigma_{t+1}(\Sigma_t^{-1}\mu_t + A_t r_t/\sigma^2) \ . \tag{A.133}$$

APPENDIX B

SUPPLEMENTARY MATERIAL: DESIGNING TRANSPORTABLE EXPERIMENTS UNDER S-ADMISSABILITY

In Section B.1 we discuss the variance reduction for $d \geq 1$ when the sample size is finite. In Section B.2 we show the proofs of Section 2.5.1. In Section B.3 we show the proofs of Section 2.5.2.1. In Section B.4 we show the proofs of Section 2.5.2.2. In Section B.5 we show the proofs of Appendix B.1.

For a random variable R with value r, we write the expectation, variance and covariance conditioning on r as a short-hand for conditioning on $R = r$. On the other hand, the expectation, variance and covariance conditioning on R are functions of R and therefore are random variables. For example,$\mathbb{E}[\hat{\tau}_Y^T|\mathbf{X}, \mathbf{Y}]$ is a function of $\mathbf{X}$ and $\mathbf{Y}$, $\mathbb{E}[\hat{\tau}_Y^T|\mathbf{X}, \mathbf{y}] = \mathbb{E}[\hat{\tau}_Y^T|\mathbf{X}, \mathbf{Y} = \mathbf{y}]$ is a function of $\mathbf{X}$, while $\mathbb{E}[\hat{\tau}_Y^T|\mathbf{x}, \mathbf{y}] = \mathbb{E}[\hat{\tau}_Y^T|\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}]$ is a value.

Conditioning on $\mathbf{x}$ and $\mathbf{y}$, the randomness only comes from $\mathbf{Z}$. Therefore $\text{var}_{\mathbf{Z}_\rho}(.|\mathbf{x}, \mathbf{y})$, $Cov_{\mathbf{Z}_\rho}(.|\mathbf{x}, \mathbf{y})$ and $\mathbb{E}_{\mathbf{Z}_\rho}(.|\mathbf{x}, \mathbf{y})$ can be written as $\text{var}_{\mathbf{Z}}(.|\mathbf{x}, \mathbf{y}, \rho = 1), Cov_{\mathbf{Z}}(.|\mathbf{x}, \mathbf{y}, \rho = 1)$ and $\mathbb{E}_{\mathbf{Z}}(.|\mathbf{x}, \mathbf{y}, \rho = 1)$ respectively. We use both notations in the proofs.

For a random variable R, we use $Cov(R)^{-1/2}$ to denote the Cholesky square root of $Cov(R)^{-1}$.

We restate the model and some notations here for convenience. Let the model be:

$$Y_i^1 = X_i^T \beta_1 + \mathcal{E}_i^1. \qquad Y_i^0 = X_i^T \beta_0 + \mathcal{E}_i^0. \tag{B.1}$$

Let ϵ_i^1 and ϵ_i^0 be the values taken by random variables $\mathcal{E}_i^1$ and $\mathcal{E}_i^0$. Let $C_i = \frac{Y_i^0+Y_i^1}{2}$, $\tilde{C}_i = W_i C_i$, $\mathbf{C} \stackrel{\text{def}}{=} (C_1, \cdots, C_n)$ and $\tilde{\mathbf{C}} = (\tilde{C}_1, \cdots, \tilde{C}_n)$. Let $c_i, \tilde{c}_i, \mathbf{c}$ and $\tilde{\mathbf{c}}$ be the values taken by $C_i, \tilde{C}_i, \mathbf{C}$ and $\tilde{\mathbf{C}}$. Then

$$C_i = X_i^T\beta + \mathcal{E}_i \qquad\qquad c_i = x_i^T\beta + \epsilon_i, \tag{B.2}$$

$$\tilde{C}_i = \tilde{X}_i^T\beta + \tilde{\mathcal{E}}_i \qquad\qquad \tilde{c}_i = \tilde{x}_i^T\beta + \tilde{\epsilon}_i, \tag{B.3}$$

where $\beta = \frac{\beta_1+\beta_0}{2}, \mathcal{E}_i = \frac{\mathcal{E}_i^1+\mathcal{E}_i^0}{2}, \tilde{X}_i = W_i X_i$ and $\tilde{\mathcal{E}}_i = W_i\mathcal{E}_i$. Let ϵ_i and $\tilde{\epsilon}_i = w_i\epsilon_i$ be the value taken by $\mathcal{E}_i$ and $\tilde{\mathcal{E}}_i$. Let $\tilde{\boldsymbol{\mathcal{E}}} = (\tilde{\mathcal{E}}_1, \cdots, \tilde{\mathcal{E}}_n)$.

B.1 Additional Results: Finite Sample Size Variance Reduction for $d \geq 1$

In this section we discuss the finite sample case when X is a multivariate random variable, which is a generalization of the result in Section 2.5.2.1 when $d = 1$. We show that when the sample size is finite, if β points to all directions with equal probability, then a balance condition which also consider the target population and is similar to Target Balance achieves the optimal variance reduction in expectation over β. The proofs are in Appendix B.5.

We will use the variance decomposition in the matrix form similar to (Harshaw et al., 2019) and provide intuition about the effect of balancing on the variance. The following lemma is the general case when $d \geq 1$ of Lemma 5 in Section 2.5.2.1.

Lemma 27. *For any function $\rho(\mathbf{x}, \mathbf{Z}) \in \{0, 1\}$ satisfying $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$:*

$$var^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}_Y^T|\mathbf{x}) = \beta^T Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})\beta + \frac{6}{n^2}\sigma^2_{\mathcal{E}}\sum_{i=1}^n w_i^2, \tag{B.4}$$

for $V \stackrel{\text{def}}{=} \frac{2}{n}(\mathbf{w} \cdot \mathbf{x})^T\mathbf{Z} = \frac{2}{n}\tilde{\mathbf{x}}^T\mathbf{Z}$.

Since the design affects only the first term in the above expression, we focus on the the random variable V. V is now a d-dimensional vector and β is unknown.

To understand the first term, we use the same decomposition of $\beta^T Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})\beta$ as in (Harshaw et al., 2019), which we restate for completeness here. Let $\mathbf{e}_1, ..., \mathbf{e}_n$ and $\lambda_1, .., \lambda_n$ be the normalized eigenvectors and corresponding eigenvalues of matrix $Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})$. Since $Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})$ is symmetric, the eigenvectors form an orthonormal basis so we can write β as a linear combination of $\mathbf{e}_1, .., \mathbf{e}_n$ and get:

$$\beta = \|\beta\| \sum_{i=1}^{n} \eta_i \mathbf{e}_i, \tag{B.5}$$

where $\eta_i = \langle \beta, \mathbf{e}_i \rangle / \|\beta\|$ is the coefficient that captures the alignment of the weighted outcome β with respect to the eigenvector $\mathbf{e}_i$. Therefore:

$$\beta^T Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})\beta = \|\beta\|^2 \sum_{i=1}^{n} \eta_i^2 \lambda_i. \tag{B.6}$$

In the worst case, β can align with the eigenvector of $Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})$ with the largest eigenvalue. Therefore a good design is one with ρ that minimize the largest eigenvalue of $Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})$. We leave this for future works. In this work we consider the average case direction - when β with norm $\|\beta\| = l$ can point in any direction with equal probability. In that case, we have

Lemma 28.

$$\mathbb{E}_{\|\beta\|=l} \beta^T Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})\beta = \frac{l^2}{2} \pi \, Trace(Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})). \tag{B.7}$$

We can then ask for the balance event Ω which results in minimizing the trace of $Cov_{\mathbf{Z}}(V|\mathbf{x}, \Omega)$, which is shown in the following lemma. Note that when $d = 1$, the trace of $Cov_{\mathbf{Z}}(V|\mathbf{x}, \Omega)$ is the variance $\text{var}_{\mathbf{Z}}(V|\mathbf{x}, \Omega)$, and this result is the general case of minimizing the variance of a 1-dimensional random variable in Section 2.5.2.1.

Lemma 29. *Let $U \in \mathcal{R}^d$ be a discrete random variable such that $\mathbb{E}[U] = 0$. Let $u_\alpha \stackrel{\text{def}}{=} \min\{u \in \mathcal{R} : \mathbb{P}(\|U\|^2 < u) \geq 1 - \alpha$.*

Let Ω be an event such that $\mathbb{P}(\Omega) \geq 1 - \alpha$ and $\mathbb{E}[U|\Omega] = 0$. Then:

$$Trace(Cov(U|\|U\|^2 < u_\alpha) \leq Trace(Cov(U|\Omega)) \tag{B.8}$$

It follows from Lemma 27, Lemma 28 and Lemma 29 that we can minimize $\mathbb{E}_\beta \text{var}^S_{\mathbf{Y},\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x}, \Omega)$ by defining the following balance condition:

Definition 5 (Alternate Target Balance)**.** *With a rejection threshold α, define the balance condition*

$$\phi'^{\alpha}_T = \begin{cases} 1, & \text{if } \|V\|^2 < a, \\ 0, & \text{otherwise.} \end{cases}$$

where a be such that $\mathbb{P}(\phi'^{\alpha}_T = 1|\mathbf{x}) = 1 - \alpha$.

Recall that Target Balance use the condition $\|B\|^2 < a$ where $B = VCov_{\mathbf{Z}}(V)^{-1/2}$ is the normalized random variable of V. Note since that $V = \frac{2}{n}\tilde{\mathbf{x}}^T\mathbf{Z}$, Alternate Target Balance also considers the target population in the design phase. However Alternate Target Balance is not invariant under linear transformations of the covariates x_i's while Target Balance is.

We have the following Theorem which is a generalization of Theorem 5 in Section 2.5.2.1.

Theorem 11. *Let $\|\beta\| = l$ and β points in any direction with equal probability and $n_0 = n_1 = n/2$.*

Let $\rho(\mathbf{X}, \mathbf{Z})$ be a function satisfying $\rho(\mathbf{X}, \mathbf{Z}) = \rho(\mathbf{X}, -\mathbf{Z})$ and $\mathbb{P}(\rho = 1|\mathbf{x}) \geq 1 - \alpha$. Then

$$\mathbb{E}_\beta var^S_{\mathbf{Y},\mathbf{Z}_{\phi'^{\alpha}_T}}(\hat{\tau}^T_Y|\mathbf{x}) \leq \mathbb{E}_\beta var^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}). \tag{B.9}$$

Similar to Section 2.5.2.1, applying Theorem 5 with ρ being the constant function $\rho(\mathbf{x}, \mathbf{Z}) = 1$ for all $\mathbf{x}, \mathbf{Z}$, we have:

Corollary 4. *Let $\|\beta\| = l$ and β points in any direction with equal probability. When $n_0 = n_1 = n/2$, using Alternate Target Balance reduces the variance compared to complete randomization in expectation over β.*

$$\mathbb{E}_\beta var^S_{\mathbf{Z}_{\phi'_T}, \mathbf{Y}}(\hat{\tau}^T_Y|\mathbf{x}) \leq \mathbb{E}_\beta var^S_{\mathbf{Z}, \mathbf{Y}}(\hat{\tau}^T_Y|\mathbf{x}) \tag{B.10}$$

Recall that the first term in the decomposition in Lemma 27 is equal to:

$$\beta^T Cov_{\mathbf{Z}_\rho}(V|\mathbf{x})\beta = \gamma^T Cov_{\mathbf{Z}_\rho}(B|\mathbf{x})\gamma = \gamma^T Cov_{\mathbf{Z}}(B|\mathbf{x}, \rho = 1)\gamma. \tag{B.11}$$

where $\gamma = \beta^T Cov_{\mathbf{Z}}(V)^{1/2}$ and $B = VCov_{\mathbf{Z}}(V)^{-1/2}$.

When the sample size is large, B converges to a standard normal distribution. Recall that Target Balance is equal to truncating $\|B\|^2 < a$. So $Cov_{\mathbf{Z}_{\phi_T}}(B|\mathbf{x})$ is the covariance of a standard normal random variable B truncated by $\|B\|^2 < a$. From Theorem 3.1 in Morgan et al. (2012) when B is a standard normal distribution, $Cov(B|\mathbf{x}, \phi_T = 1) = vCov(B|\mathbf{x})$ for some $v < 1$, so the variance is reduced. However we do not need to go through this analysis because Li et al. (2018) already has variance reduction results for the case when the sample size is large. In Section 2.5.2.2 we use the result from Li et al. (2018) directly to show that Target Balance achieves a smaller variance than Source Balance.

B.2 Proofs of Section 2.5.1

In this section we prove Theorem 4. We made use of the following lemma from Morgan et al. (2012):

Lemma 30 (from the proof of Theorem 2.1 in Morgan et al. (2012))**.** *Let* $\mathbf{A} \stackrel{\text{def}}{=} (A_1, ..., A_n)^T \in \mathcal{R}^n$*. Let* $n_1 = n_0 = n/2$*. For any function* $\rho(\mathbf{x}, \mathbf{A}) \in \{0, 1\}$ *satisfying* $\rho(\mathbf{x}, \mathbf{A}) = \rho(\mathbf{x}, 1 - \mathbf{A})$*:*

$$\mathbb{E}^S_{\mathbf{A}}[A_i|\mathbf{x}, \mathbf{y}, \rho = 1] = \frac{1}{2}. \tag{B.12}$$

We also prove the following lemma in order to prove Theorem 4:

Lemma 31. *For any function* $\rho(\mathbf{x}, \mathbf{A}) \in \{0, 1\}$ *satisfying* $\rho(\mathbf{x}, \mathbf{A}) = \rho(\mathbf{x}, 1 - \mathbf{A})$*:*

$$\mathbb{E}_{\mathbf{A}|\rho=1}[\hat{\tau}^T_Y|\mathbf{X}, \mathbf{Y}] = \frac{1}{n}\sum_{i=1}^{n} W_i(Y_i^1 - Y_i^0). \tag{B.13}$$

$$\mathbb{E}^S_{\mathbf{Y},\mathbf{A}|\rho=1}[\hat{\tau}^T_Y|\mathbf{X}] = \frac{1}{n}\sum_{i=1}^{n} W_i(\beta_1 - \beta_0)^T X_i. \tag{B.14}$$

Proof. From Lemma 30, $\mathbb{E}[A_i|\mathbf{X}, \mathbf{Y}, \rho = 1] = \mathbb{E}[A_i|\mathbf{X}, \rho = 1] = \frac{1}{2}$. Therefore:

$$\mathbb{E}_{\mathbf{A}|\rho=1}[\hat{\tau}^T_Y|\mathbf{X}, \mathbf{Y}] \tag{B.15}$$

$$= \frac{1}{n_1}\sum_{i=1}^{n} \mathbb{E}_{\mathbf{A}}\left[W_i A_i Y_i^1 \middle| \mathbf{X}, \mathbf{Y}, \rho = 1\right] - \frac{1}{n_0}\sum_{i=1}^{n} \mathbb{E}_{\mathbf{A}}\left[W_i(1 - A_i)Y_i^0 \middle| \mathbf{X}, \mathbf{Y}, \rho = 1\right] \tag{B.16}$$

$$= \frac{1}{n_1}\sum_{i=1}^{n} W_i Y_i^1 \mathbb{E}_{\mathbf{A}}\left[A_i \middle| \mathbf{X}, \mathbf{Y}, \rho = 1\right] - \frac{1}{n_0}\sum_{i=1}^{n} W_i Y_i^0 \mathbb{E}_{\mathbf{A}}\left[1 - A_i \middle| \mathbf{X}, \mathbf{Y}, \rho = 1\right] \tag{B.17}$$

$$= \frac{1}{n}\sum_{i=1}^{n} W_i(Y_i^1 - Y_i^0)\,. \tag{B.18}$$

$$\mathbb{E}^S_{\mathbf{A}|\rho=1,\mathbf{Y}}[\hat{\tau}^T_Y|\mathbf{X}] \tag{B.19}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\mathbb{E}_{\mathbf{A}}[\hat{\tau}^T_Y|\mathbf{X}, \mathbf{Y}, \rho = 1]|\mathbf{X}\right] \tag{B.20}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\frac{1}{n}\sum_{i=1}^{n} W_i(Y_i^1 - Y_i^0)|\mathbf{X}\right] \tag{B.21}$$

$$= \frac{1}{n}\sum_{i=1}^{n} W_i(\beta_1 - \beta_0)^T X_i\,. \tag{B.22}$$

□

Proof of Theorem 4. Let D_S and D_T be the supports of the source and target distributions. Since $p_T(X) > 0 \to p_S(X) > 0$ and $p_T(Y|X) = p_S(Y|X)$, we have $D_T \subseteq D_S$. Using Lemma 31:

$$\mathbb{E}^S_{\mathbf{X},\mathbf{Y},\mathbf{Z}_{\phi_T}}\left[\hat{\tau}^T_Y\right] \tag{B.23}$$

$$=\mathbb{E}^S_{\mathbf{X},\mathbf{Y}}\mathbb{E}_{\mathbf{A}_{\phi_T}}[\hat{\tau}^T_Y|\mathbf{X},\mathbf{Y}] \tag{B.24}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}^S_{\mathbf{X},\mathbf{Y}}\left[W_i(Y_i^1 - Y_i^0)\right] \tag{B.25}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{(x,y)\in D_S}\left(\frac{p_T(x)}{p_S(x)}(y^1 - y^0)\right)p_S(x,y)dxy \tag{B.26}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{(x,y)\in D_S}\left(\frac{p_T(y|x)p_T(x)}{p_S(y|x)p_S(x)}(y^1 - y^0)\right)p_S(x,y)dxy \text{ because } p_T(y|x) = p_S(y|x) \tag{B.27}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{(x,y)\in D_S}\left(\frac{p_T(y,x)}{p_S(y,x)}(y^1 - y^0)\right)p_S(x,y)dxy \tag{B.28}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{(x,y)\in D_S}p_T(x,y)(y^1 - y^0)dxy \tag{B.29}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{(x,y)\in D_T}p_T(x,y)(y^1 - y^0)dxy \text{ because } D_T \subseteq D_S \tag{B.30}$$

$$=\tau^T_Y. \tag{B.31}$$

□

B.3 Proofs of Section 2.5.2.1

In this section we prove Lemma 3, Lemma 4, Lemma 5, Theorem 5 and Corollary 3. Note that the results in this section are the special case when $d = 1$ of the results in Section B.1. Lemma 4 is a special case when $d = 1$ of Lemma 37. Lemma 5 is a special case of Lemma 27 and Theorem 5 is a special case of Theorem 11. However in this section we state the full proofs for the case $d = 1$ so that the readers do not need

to read the proofs of Section B.1 in order to understand Section 2.5.2.1 in the main paper.

Proof of Lemma 3. By law of total variance:

$$\text{var}^S_{\mathbf{Z}_\rho,\mathbf{X},\mathbf{Y}}(\hat{\tau}^T_Y) = \mathbb{E}^S_{\mathbf{X}}\text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{X}) + \text{var}^S_{\mathbf{X}}\left(\mathbb{E}^S_{\mathbf{Y},\mathbf{Z}_\rho}[\hat{\tau}^T_Y|\mathbf{X}]\right). \tag{B.32}$$

Since $\rho(\mathbf{x},\mathbf{Z}) = \rho(\mathbf{x},-\mathbf{Z})$, from Lemma 31:

$$\mathbb{E}^S_{\mathbf{Y},\mathbf{Z}_\rho}[\hat{\tau}^T_Y|\mathbf{X}] = \frac{1}{n}\sum_{i=1}^{n} W_i(\beta_1-\beta_0)^T X_i. \tag{B.33}$$

Therefore:

$$\text{var}^S_{\mathbf{X}}\left(\mathbb{E}^S_{\mathbf{Y},\mathbf{Z}_\rho}[\hat{\tau}^T_Y|\mathbf{X}]\right) = \text{var}^S_{\mathbf{X}}\left(\frac{1}{n}\sum_{i=1}^{n} W_i(\beta_1-\beta_0)^T X_i)\right). \tag{B.34}$$

□

Proof of Lemma 4. By definition:

$$\text{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{y},\rho=1) = \mathbb{E}_{\mathbf{Z}}\left[(\hat{\tau}^T_Y - \mathbb{E}_{\mathbf{Z}}[\hat{\tau}^T_Y|\mathbf{x},\mathbf{y},\rho=1])^2|\mathbf{x},\mathbf{y},\rho=1\right]. \tag{B.35}$$

From Lemma 31

$$\mathbb{E}_{\mathbf{Z}}[\hat{\tau}^T_Y|\mathbf{x},\mathbf{y},\rho=1] = \frac{1}{n}\left(\sum_{i=1}^{n} w_i y_i^1 - \sum_{i=1}^{n} w_i y_i^0\right). \tag{B.36}$$

On the other hand conditioning on $\mathbf{X}=\mathbf{x}$ and $\mathbf{Y}=\mathbf{y}$ and let y_i^* denote the observed outcome of sample i:

$$\hat{\tau}^T_Y = \frac{2}{n}(\sum_{Z_i=1} w_i y_i^* - \sum_{Z_i=-1} w_i y_i^*) \tag{B.37}$$

$$= \frac{2}{n}\sum_{i=1}^{n} w_i A_i y_i^1 - \frac{2}{n}\sum_{i=1}^{n} w_i(1-A_i)y_i^0. \tag{B.38}$$

Therefore:

$$\text{var}_{\mathbf{Z}}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{y},\rho=1) \tag{B.39}$$

$$= \mathbb{E}_{\mathbf{Z}}\left[\left(\frac{2}{n}(\sum_{i=1}^{n} w_i A_i y_i^1 - \sum_{i=1}^{n} w_i(1-A_i)y_i^0) - \frac{1}{n}\sum_{i=1}^{n} w_i(y_i^1 - y_i^0)\right)^2 \middle| \mathbf{x},\mathbf{y},\rho=1\right] \tag{B.40}$$

$$= \mathbb{E}_{\mathbf{Z}}\left[\left(\frac{1}{n}(\sum_{i=1}^{n} w_i(2A_i-1)y_i^1 + \frac{1}{n}\sum_{i=1}^{n} w_i(2A_i-1)y_i^0)\right)^2 \middle| \mathbf{x},\mathbf{y},\rho=1\right] \tag{B.41}$$

$$= \frac{4}{n^2}\mathbb{E}_{\mathbf{Z}}\left[\left(\sum_{i=1}^{n} w_i Z_i \frac{y_i^1 + y_i^0}{2}\right)^2 \middle| \mathbf{x},\mathbf{y},\rho=1\right] \tag{B.42}$$

$$= \frac{4}{n^2}\mathbb{E}_{\mathbf{Z}}\left[\left(\sum_{i=1}^{n} Z_i w_i c_i\right)^2 \middle| \mathbf{x},\mathbf{y},\rho=1\right], \tag{B.43}$$

where $c_i = \frac{y_i^1+y_i^0}{2}$. □

Proof of Lemma 5. By law of total variance:

$$\text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}_Y^T|\mathbf{x}) = \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}_\rho}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y})|\mathbf{x}\right] + \text{var}^S_{\mathbf{Y}}(\mathbb{E}_{\mathbf{Z}_\rho}[\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y}]|\mathbf{x}) \tag{B.44}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}_\rho}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y})|\mathbf{x}\right] + \text{var}^S_{\mathbf{Y}}\left(\frac{1}{n}\sum_{i=1}^{n} w_i(Y_i^1 - Y_i^0)|\mathbf{x}\right) \tag{B.45}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}_\rho}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y})|\mathbf{x}\right] + \frac{1}{n^2}\sum_{i=1}^{n} w_i^2 \text{var}(\mathcal{E}_i^1 - \mathcal{E}_i^0) \tag{B.46}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}_\rho}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y})|\mathbf{x}\right] + \frac{2}{n^2}\sigma^2_{\mathcal{E}}\sum_{i=1}^{n} w_i^2\,. \tag{B.47}$$

Recall that $\tilde{C}_i = \beta\tilde{X}_i + \tilde{\mathcal{E}}_i$. From Lemma 4:

$$
\begin{aligned}
&\mathrm{var}_{\mathbf{Z}}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y},\rho=1)\\
&=\frac{4}{n^2}\mathbb{E}_{\mathbf{Z}}\left[\left(\sum_{i=1}^{n}Z_i\tilde{C}_i\right)^2\middle|\mathbf{x},\mathbf{Y},\rho=1\right]\\
&=\frac{4}{n^2}\mathbb{E}_{\mathbf{Z}}\left[\left(\mathbf{Z}^T\tilde{\mathbf{C}}\right)^2\middle|\mathbf{x},\mathbf{Y},\rho=1\right]\\
&=\frac{4}{n^2}\mathbb{E}_{\mathbf{Z}}\left[\left(\mathbf{Z}^T\beta\tilde{\mathbf{x}}+\mathbf{Z}^T\tilde{\boldsymbol{\mathcal{E}}}\right)^2\middle|\mathbf{x},\mathbf{Y},\rho=1\right]\\
&=\frac{4}{n^2}\beta^2\mathbb{E}_{\mathbf{Z}}\left[\left(\mathbf{Z}^T\tilde{\mathbf{x}}\right)^2\middle|\mathbf{x},\rho=1\right]+\frac{4}{n^2}\mathbb{E}_{\mathbf{Z}}\left[\left(\mathbf{Z}^T\tilde{\boldsymbol{\mathcal{E}}}\right)^2\middle|\mathbf{x},\mathbf{Y},\rho=1\right]\\
&\qquad\qquad+\frac{4}{n^2}2\mathbb{E}_{\mathbf{Z}}\left[\tilde{\mathbf{x}}^T\mathbf{Z}\mathbf{Z}^T\tilde{\boldsymbol{\mathcal{E}}}\middle|\mathbf{x},\mathbf{Y},\rho=1\,.\right] \qquad \text{(B.48)}
\end{aligned}
$$

Now we consider $\mathbb{E}_{\mathbf{Y}}^S\left[\mathrm{var}_{\mathbf{Z}}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right]$. The third term in Eq. B.48 becomes:

$$
\begin{aligned}
\frac{4}{n^2}2\mathbb{E}_{\mathbf{Y}}^S\left[\mathbb{E}_{\mathbf{Z}}\left[\tilde{\mathbf{x}}^T\mathbf{Z}\mathbf{Z}^T\tilde{\boldsymbol{\mathcal{E}}}\middle|\mathbf{x},\mathbf{Y},\rho=1\right]\middle|\mathbf{x}\right]&=\frac{8}{n^2}\mathbb{E}_{\mathbf{Z}}\left[\tilde{\mathbf{x}}^T\mathbf{Z}\mathbf{Z}^T\middle|\mathbf{x},\rho=1\right]\mathbb{E}_{\mathbf{Y}}^S[\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{x}] \qquad \text{(B.49)}\\
&=0 \text{ because } \mathbb{E}_{\mathbf{Y}}^S[\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{x}]=\mathbf{0}. \qquad \text{(B.50)}
\end{aligned}
$$

The second term in Eq. B.48 becomes:

$$\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\mathbb{E}_{\mathbf{Z}}\left[\left(\mathbf{Z}^T\tilde{\boldsymbol{\mathcal{E}}}\right)^2\Big|\mathbf{x},\mathbf{Y},\rho=1\right]\Big|\mathbf{x}\right] \tag{B.51}$$

$$=\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\mathbb{E}_{\mathbf{Z}}\left[\left(\sum_{i=1}^n Z_i w_i\mathcal{E}_i\right)^2\Big|\mathbf{x},\mathbf{Y},\rho=1\right]\Big|\mathbf{x}\right] \tag{B.52}$$

$$=\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^n (Z_i w_i\mathcal{E}_i)^2\big|\mathbf{x},\mathbf{Y},\rho=1\right]\Big|\mathbf{x}\right] \tag{B.53}$$

$$+\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\mathbb{E}_{\mathbf{Z}}\left[\sum_{i\neq j}(Z_i w_i\mathcal{E}_i)(Z_j w_j\mathcal{E}_j)\big|\mathbf{x},\mathbf{Y},\rho=1\right]\Big|\mathbf{x}\right] \tag{B.54}$$

$$=\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^n (Z_i w_i\mathcal{E}_i)^2\big|\mathbf{x},\mathbf{Y},\rho=1\right]\Big|\mathbf{x}\right] \tag{B.55}$$

$$+\frac{4}{n^2}\sum_{i\neq j}\mathbb{E}_{\mathbf{Z}}[Z_i Z_j|\mathbf{x},\rho=1]w_i w_j\mathbb{E}_{\mathbf{Y}}^S\left[\mathcal{E}_i\mathcal{E}_j|\mathbf{x}\right] \tag{B.56}$$

$$=\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^n (Z_i w_i\mathcal{E}_i)^2\big|\mathbf{x},\mathbf{Y},\rho=1\right]\Big|\mathbf{x}\right]+0 \tag{B.57}$$

$$\text{because } \mathbb{E}_{\mathbf{Y}}^S[\mathcal{E}_i\mathcal{E}_j|\mathbf{x}]=\mathbb{E}_{\mathbf{Y}}^S[\mathcal{E}_i|\mathbf{x}]\mathbb{E}_{\mathbf{Y}}^S[\mathcal{E}_j|\mathbf{x}]=0 \tag{B.58}$$

$$=\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\sum_{i=1}^n (w_i\mathcal{E}_i)^2]\big|\mathbf{x}\right]\ \text{ because } Z_i^2=1 \tag{B.59}$$

$$=\frac{4}{n^2}\sigma_{\mathcal{E}}^2\sum_{i=1}^n w_i^2 \tag{B.60}$$

The first term in Eq. B.48 becomes:

$$\frac{4}{n^2}\mathbb{E}_{\mathbf{Y}}^S\left[\beta^2\mathbb{E}_{\mathbf{Z}}\left[\left(\mathbf{Z}^T\tilde{\mathbf{x}}\right)^2\Big|\mathbf{x},\rho=1\right]\Big|\mathbf{x}\right]=\frac{4}{n^2}\beta^2\mathbb{E}_{\mathbf{Z}}\left[\left(\mathbf{Z}^T\tilde{\mathbf{x}}\right)^2\Big|\mathbf{x},\rho=1\right] \tag{B.61}$$

$$=\frac{4}{n^2}\beta^2\mathbb{E}_{\mathbf{Z}}\left[(\sum_{i=1}^n Z_i w_i x_i)^2\big|\mathbf{x},\rho=1\right]. \tag{B.62}$$

Putting all 3 terms together:

$$\mathbb{E}_{\mathbf{Y}}^S\left[\mathrm{var}_{\mathbf{Z}}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right]=\frac{4}{n^2}\beta^2\mathbb{E}_{\mathbf{Z}}\left[(\sum_{i=1}^n Z_i w_i x_i)^2\big|\mathbf{x},\rho=1\right]+\frac{4}{n^2}\sigma_{\mathcal{E}}^2\sum_{i=1}^n w_i^2\,. \tag{B.63}$$

Therefore:

$$\mathrm{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) = \mathbb{E}^S_{\mathbf{Y}}\left[\mathrm{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right] + \frac{2}{n^2}\sigma^2_{\mathcal{E}}\sum_{i=1}^n w_i^2 \tag{B.64}$$

$$= \frac{4}{n^2}\beta^2\mathbb{E}_{\mathbf{Z}}\left[(\sum_{i=1}^n Z_i w_i x_i)^2\big|\mathbf{x},\rho=1\right] + \frac{6}{n^2}\sigma^2_{\mathcal{E}}\sum_{i=1}^n w_i^2\,. \tag{B.65}$$

□

In order to prove Theorem 5, we will show that for a random variable U with $\mathbb{E}[U] = 0$, among events Ω preserve the expectation $\mathbb{E}[U|\Omega] = 0$, truncating the tail results in the smallest variance. Note that if $\rho(\mathbf{x},\mathbf{Z}) = \rho(\mathbf{x},-\mathbf{Z})$ it follows from Lemma 30 that $\mathbb{E}[\frac{2}{n}\tilde{\mathbf{x}}^T\mathbf{Z}|\rho=1] = \mathbb{E}[\frac{2}{n}\tilde{\mathbf{x}}^T\mathbf{Z}] = 0$.

In order to prove Theorem 5 we show how to minimize the variance of a random variable:

Lemma 32. *Let $U \in \mathcal{R}$ be a discrete random variable such that $\mathbb{E}[U] = 0$. Let $u_\alpha \stackrel{\text{def}}{=} \min\{u \in \mathcal{R} : \mathbb{P}(U^2 < u) \geq 1-\alpha\}$. Let Ω be an event such that $\mathbb{P}(\Omega) \geq 1-\alpha$ and $\mathbb{E}[U|\Omega] = 0$. Then:*

$$\mathbb{E}(U^2|U^2 < u_\alpha) \leq \mathbb{E}(U^2|\Omega) \tag{B.66}$$

Proof. Let $p(u)$ be the pmf of U. Define $f(u)$ as follow:

$$f(u) = p(u)\mathbf{1}(u \in \Omega). \tag{B.67}$$

then:

$$p(u|\Omega) = \frac{p(u)\mathbf{1}(u\in\Omega)}{\mathbb{P}(\Omega)} = \frac{f(u)}{1-\alpha}\,. \tag{B.68}$$

Therefore:

$$\mathbb{E}[U^2|\Omega] = \sum_u u^2 \frac{f(u)}{1-\alpha} \ . \tag{B.69}$$

We want to minimize $\mathbb{E}(U^2|\Omega)$:

$$\sum_u u^2 \frac{f(u)}{1-\alpha} \tag{B.70}$$

subject to:

$$f(u) \in \{0, p(u)\} \ \forall u \tag{B.71}$$

$$\mathbb{P}(\Omega) = \sum_u f(u) \geq 1 - \alpha. \tag{B.72}$$

This can be done by maximize $f(u)$ so that $f(u) = p(u)$ for the smallest u^2, which is equal to set Ω to be the event $U^2 < u_\alpha$. □

Proof of Theorem 5. Let $V := \frac{2}{n}\sum_i w_i x_i Z_i$ and $B = V\text{var}(V)^{-1/2}$. From Lemma 5:

$$\text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) = \beta^2 \mathbb{E}_{\mathbf{Z}}\left[V^2 \middle| \mathbf{x}, \rho = 1\right] + \frac{6}{n^2}\sigma^2_{\mathcal{E}} \sum_{i=1}^n w_i^2. \tag{B.73}$$

$$= \beta^2 \text{var}(V) \mathbb{E}_{\mathbf{Z}}\left[B^2 \middle| \mathbf{x}, \rho = 1\right] + \frac{6}{n^2}\sigma^2_{\mathcal{E}} \sum_{i=1}^n w_i^2. \tag{B.74}$$

Since $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$, from Lemma 30 we have $\mathbb{E}_{\mathbf{Z}}[B|\mathbf{x}, \rho = 1] = 0$, which satisfies the criteria in Lemma 32.

Let $\eta := 1 - \mathbb{P}(\rho = 1|\mathbf{x})$. Then $\eta \leq \alpha$. Let b_η be such that $\mathbb{P}(B^2 < b_\eta|\mathbf{x}) = 1 - \eta$ and b_α be such that $\mathbb{P}(B^2 < b_\alpha|\mathbf{x}) = 1 - \alpha$. From Lemma 32:

$$\mathbb{E}_{\mathbf{Z}}\left[B^2 \middle| \mathbf{x}, \rho = 1\right] \geq \mathbb{E}_{\mathbf{Z}}\left[B^2 \middle| \mathbf{x}, B^2 < b_\eta\right] \tag{B.75}$$

$$\geq \mathbb{E}_{\mathbf{Z}}\left[B^2 \middle| \mathbf{x}, B^2 < b_\alpha\right] \text{ because } b_\eta \geq b_\alpha \tag{B.76}$$

$$\geq \mathbb{E}_{\mathbf{Z}}\left[B^2 \middle| \mathbf{x}, \phi^\alpha_T = 1\right] \tag{B.77}$$

□

Proof of Corollary 3. Let ρ being the constant function $\rho(\mathbf{x}, \mathbf{Z}) = 1$ for all $\mathbf{x}, \mathbf{Z}$. Then:

$$\text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) = \text{var}^S_{\mathbf{Y},\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x}). \tag{B.78}$$

From Theorem 5 we have:

$$\text{var}^S_{\mathbf{Y},\mathbf{Z}_{\phi^\alpha_T}}(\hat{\tau}^T_Y|\mathbf{x}) \leq \text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) = \text{var}^S_{\mathbf{Y},\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x}) \tag{B.79}$$

□

B.4 Discussion on Section 2.5.2.2

Proof of Lemma 6. By law of total variance:

$$\text{var}^S_{\mathbf{X},\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y) = \mathbb{E}^S_{\mathbf{X},\mathbf{Y}}\left[\text{var}_{\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{X},\mathbf{Y})\right] + \text{var}^S_{\mathbf{X},\mathbf{Y}}\left(\mathbb{E}_{\mathbf{Z}_\rho}[\hat{\tau}^T_Y|\mathbf{X},\mathbf{Y}]\right) \tag{B.80}$$

Since $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$, from Lemma 31:

$$\mathbb{E}_{\mathbf{Z}}[\hat{\tau}^T_Y|\mathbf{X},\mathbf{Y},\rho = 1] = \frac{1}{n}\sum_{i=1}^{n} W_i(Y_i^1 - Y_i^0) \tag{B.81}$$

Therefore:

$$\text{var}^S_{\mathbf{X},\mathbf{Y}}\left(\mathbb{E}_{\mathbf{Z}}[\hat{\tau}^T_Y|\mathbf{X},\mathbf{Y},\rho = 1]\right) = \text{var}^S_{\mathbf{X},\mathbf{Y}}\left(\frac{1}{n}\sum_{i=1}^{n} W_i(Y_i^1 - Y_i^0)\right) \tag{B.82}$$

□

We now prove Lemma 7. We use the following result in Harshaw et al. (2019) to prove Lemma 7.

Lemma 33 (Lemma A1 in Harshaw et al. (2019))**.** *Let y_i^* denote the observed outcome of sample i:*

$$\frac{2}{n}(\sum_{i:z_i=1} y_i^* - \sum_{i:z_i=-1} y_i^*) - \frac{1}{n}\sum_{i=1}^{n}(y_i^1 - y_i^0) = \frac{2}{n}\mathbf{c}^T\mathbf{z} \tag{B.83}$$

where $c_i = \frac{y_i^1+y_i^0}{2}$ and $\mathbf{c} \stackrel{\text{def}}{=} (c_1, \cdots, c_n)$.

We will also use the following lemmas:

Lemma 34. *Let $Q \stackrel{\text{def}}{=} \frac{n-1}{n}\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]$. Let $\mathbf{I}_n$ denote the $n \times n$ identity matrix and $\mathbf{1}$ denote the n dimensional vector of 1. Then:*

$$Q = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T. \tag{B.84}$$

$$Q = Q^T \tag{B.85}$$

$$Q = Q^2 = Q^TQ = QQ^T. \tag{B.86}$$

Let $\mathbf{s} \in \mathcal{R}^{n\times d}$ be a matrix. Then

$$Q\mathbf{s} = \mathbf{s} - avg(\mathbf{s})$$

where $avg(\mathbf{s}) \in \mathcal{R}^d$ is the average of rows of $\mathbf{s}$.

Proof. First we will show that:

$$\mathbb{E}[\mathbf{Z}\mathbf{Z}^T] = \frac{n}{n-1}\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \tag{B.87}$$

by showing that $\mathbb{E}[Z_i^2] = 1$ and $\mathbb{E}[Z_iZ_j] = -\frac{1}{n-1}$ when $i \neq j$. First we have that $\mathbb{E}[Z_i^2] = 1$ because $Z_i^2 = 1$. Since there are exactly $n/2$ samples with value $Z_i = 1$ and $n/2$ samples with values $Z_i = -1$, note that $(\sum_{i=1}^n Z_i)^2 = 0$ and:

$$\mathbb{E}[(\sum_{i=1}^{n} Z_i)^2] = \mathbb{E}[\sum_{i=1}^{n} Z_i^2] + \sum_{i\neq j}\mathbb{E}[Z_iZ_j]\ . \tag{B.88}$$

Since all pairs (i, j) where $i \neq j$ have equal roles and there are $n(n-1)$ such pairs:

$$\mathbb{E}[Z_i Z_j] = \frac{\mathbb{E}[(\sum_{i=1}^n Z_i)^2] - \mathbb{E}[\sum_{i=1}^n Z_i^2]}{n(n-1)} \tag{B.89}$$

$$= \frac{0 - n}{n(n-1)} \tag{B.90}$$

$$= \frac{-1}{n-1} \tag{B.91}$$

Since Q is symmetric, $Q = Q^T$. We will show that $Q = Q^2$:

$$Q^2 = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \tag{B.92}$$

$$= \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{I}_n - \frac{1}{n}\mathbf{I}_n\mathbf{1}\mathbf{1}^T + \frac{1}{n^2}\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T \tag{B.93}$$

$$= \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T = Q \tag{B.94}$$

Since $Q = Q^T$, we have $Q = Q^2 = QQ^T = Q^TQ$. For the last property:

$$Q\mathbf{s} = \mathbf{I}_n\mathbf{s} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{s} = \mathbf{s} - \text{avg}(\mathbf{s})$$

because $\mathbf{I}_n\mathbf{s} = \mathbf{s}$ and $\frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{s} = \text{avg}(\mathbf{s})$ □

Lemma 35. *For any 2 random variable $X = (X(1), \cdots, X(d)) \in R^d$ and $Y = (Y(1), \cdots, Y(d') \in R^{d'}$ such that $X(j)$ and $Y(k)$ have finite 4th moment for any $j, 1 \leq j \leq d$ and for any $k, 1 \leq k \leq d'$:*

$$\frac{1}{n}\lim_{n\to\infty}(Q\mathbf{X})^TQ\mathbf{Y} = cov(X, Y). \tag{B.95}$$

As a result, if Assumption 8 is satisfied then Assumption 7 is satisfied.

Proof.

$$\frac{1}{n}\lim_{n\to\infty}(Q\mathbf{X})^TQ\mathbf{Y} \tag{B.96}$$

$$=\frac{1}{n}\lim_{n\to\infty}\sum_{i=1}^{n}(X_i-\frac{1}{n}\mathbf{X}^T\mathbf{1})(Y_i-\frac{1}{n}\mathbf{Y}^T\mathbf{1})^T \tag{B.97}$$

$$=\frac{1}{n}\lim_{n\to\infty}\sum_{i=1}^{n}(X_iY_i^T-\frac{1}{n}\mathbf{X}^T\mathbf{1}Y_i^T-\frac{1}{n}X_i\mathbf{1}^T\mathbf{Y}+\frac{1}{n^2}\mathbf{X}^T\mathbf{1}\mathbf{1}^T\mathbf{Y}) \tag{B.98}$$

$$=\frac{1}{n}\lim_{n\to\infty}\sum_{i=1}^{n}X_iY_i^T-\lim_{n\to\infty}\frac{1}{n}\mathbf{X}^T\mathbf{1}(\sum_{i=1}^{n}Y_i^T)-\lim_{n\to\infty}\frac{1}{n}(\sum_{i=1}^{n}X_i)\mathbf{1}^T\mathbf{Y} \tag{B.99}$$

$$+\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}^T\mathbf{1}\mathbf{1}^T\mathbf{Y} \tag{B.100}$$

$$=\frac{1}{n}\lim_{n\to\infty}\sum_{i=1}^{n}X_iY_i^T-\lim_{n\to\infty}\frac{1}{n^2}(\sum_{i=1}^{n}X_i)(\sum_{i=1}^{n}Y_i^T) \tag{B.101}$$

$$=\mathbb{E}[X_iY_i^T]-\mathbb{E}[X_i]\mathbb{E}[Y_i]^T \text{ almost surely.} \tag{B.102}$$

The last equality is true if $X(j)Y(k)$, $X(j)$ and $Y(k)$ have finite mean and variances, so we can apply strong LLN. $\mathbb{E}[X(i)Y(k)]\leq\sqrt{\mathbb{E}[X(i)^2]\mathbb{E}[Y(k)^2]}$, $\mathbb{E}[X(j)^2Y(K)^2]\leq\sqrt{\mathbb{E}[X(j)^4]\mathbb{E}[Y(k)^4]}$, so $X(j)Y(k)$ has finite mean and variance if $X(j)$ and $Y(k)$ have finite 4th moment.

We now show that Assumption 7 is satisfied. Since $Y^a, a\in\{0,1\}$, $X(j), 1\leq j\leq d$ and W have finite 8th moment, $\tilde{Y}^a, a\in\{0,1\}$ and $\tilde{X}(j), 1\leq j\leq d$ have finite 4th moment using Cauchy-Schwartz inequality.

The first condition in Assumption 7 is satisfied with probability 1 since the sample covariances converge to the distributions' covariance. In the second condition, $\max_{1\leq i\leq n}|\tilde{Y}_i^a-\text{avg}(\tilde{\mathbf{Y}})^a|^2/n$, $\max_{1\leq i\leq n}\|X_i-\text{avg}(\mathbf{X})\|_2^2/n$, $\max_{1\leq i\leq n}\left\|\tilde{X}_i-\text{avg}(\tilde{\mathbf{X}})\right\|_2^2/n$ can be upper bounded by $\sum_{i=1}^{n}|\tilde{Y}_i^a-\text{avg}(\tilde{\mathbf{Y}})^a|^2/n$, $\sum_{i=1}^{n}\|X_i-\text{avg}(\mathbf{X})\|_2^2/n$ and $\sum_{i=1}^{n}\left\|\tilde{X}_i-\text{avg}(\tilde{\mathbf{X}})\right\|_2^2/n$, which converge almost surely.

$cov(\mathbf{x})$ and $cov(\tilde{\mathbf{x}})$ converges almost surely to $Cov^S(X)$ and $Cov^S(\tilde{X})$, which are non-singular by assumption. □

Lemma 36. *For any 2 random variable $X = (X(1), \cdots, X(d)) \in R^d$ and $Y \in R$ such that $X(j)$ and Y have finite 4th moment for any $j, 1 \leq j \leq d$, with probability 1:*

$$\frac{1}{n} \lim_{n\to\infty} \min_{\tilde{\beta}} \left\| Q\mathbf{Y} - Q\mathbf{X}\tilde{\beta} \right\|^2 = \frac{1}{n} \min_{\tilde{\beta}} var \left[\left\| Y_i - X_i\tilde{\beta} \right\| \right] \tag{B.103}$$

Proof. Let $\hat{\beta}_n = (\mathbf{X}^T Q \mathbf{X})^{-1}(\mathbf{X}^T Q \mathbf{Y})$. Then $\hat{\beta}_n$ is the solution to:

$$\min_{\tilde{\beta}} \left\| Q\mathbf{Y} - Q\mathbf{X}\tilde{\beta} \right\|^2 \tag{B.104}$$

Let $\beta = (\text{var}[X_i])^{-1} cov[X_i, Y_i]$. Then β is the solution to:

$$\min_{\tilde{\beta}} \text{var} \left[\left\| Y_i - X_i\tilde{\beta} \right\| \right] \tag{B.105}$$

We have:

$$\lim_{n\to\infty} \hat{\beta}_n = \lim_{n\to\infty} (\mathbf{X}^T Q \mathbf{X})^{-1}(\mathbf{X}^T Q \mathbf{Y}) \tag{B.106}$$

$$= (\text{var}[X_i])^{-1} cov[X_i, Y_i] \text{by Lemma 35} \tag{B.107}$$

$$= \beta \tag{B.108}$$

We have:

$$\frac{1}{n} \lim_{n\to\infty} \left\| Q\mathbf{Y} - Q\mathbf{X}\hat{\beta}_n \right\|^2 \tag{B.109}$$

$$= \frac{1}{n} \lim_{n\to\infty} \|Q\mathbf{Y} - Q\mathbf{X}\beta\|^2 + 2\frac{1}{n} \lim_{n\to\infty} (Q\mathbf{Y} - Q\mathbf{X}\beta)^T (Q\mathbf{X}\beta - Q\mathbf{X}\hat{\beta}_n) \tag{B.110}$$

$$+ \frac{1}{n} \lim_{n\to\infty} \left\| Q\mathbf{X}\beta - Q\mathbf{X}\hat{\beta}_n \right\|^2 \tag{B.111}$$

We consider the 1st term in Eq. B.111. By strong LLN,

$$\frac{1}{n} \lim_{n\to\infty} \|Q\mathbf{Y} - Q\mathbf{X}\beta\|^2 \tag{B.112}$$

$$= \text{var} \|Y_i - X_i\beta\| \text{ by Lemma 35 if } Y_i - X_i\tilde{\beta} \text{ have finite 4th moment} \tag{B.113}$$

$Y_i - X_i\beta$ have finite 4th moment if $\mathbb{E}[(Y_i - X_i\beta)^4]$ is finite, which is true since $X(j)$ and Y have finite 4th moment.

We will show that the 2nd term in Eq. B.111 is 0. Since $\frac{1}{n}\lim_{n\to\infty}(Q\mathbf{Y} - Q\mathbf{X}\beta)^T Q\mathbf{X} = cov(Y_i - \beta^T X_i^T, X_i)$ almost surely from Lemma 35, which is finite and $\lim_{n\to\infty}(\beta - \hat{\beta}_n) = 0$ is finite,

$$\frac{1}{n}\lim_{n\to\infty}(Q\mathbf{Y} - Q\mathbf{X}\beta)^T(Q\mathbf{X}\beta - Q\mathbf{X}\hat{\beta}_n) \tag{B.114}$$

$$= \frac{1}{n}\lim_{n\to\infty}(Q\mathbf{Y} - Q\mathbf{X}\beta)^T Q\mathbf{X}(\beta - \hat{\beta}_n) \tag{B.115}$$

$$= \frac{1}{n}\lim_{n\to\infty}(Q\mathbf{Y} - Q\mathbf{X}\beta)^T Q\mathbf{X}\lim_{n\to\infty}(\beta - \hat{\beta}_n) \text{ almost surely} \tag{B.116}$$

$$= 0 \tag{B.117}$$

We will show that the 3rd term in Eq. B.111 is 0:

$$\frac{1}{n}\lim_{n\to\infty}\left\|Q\mathbf{X}\beta - Q\mathbf{X}\hat{\beta}_n\right\|^2 \tag{B.118}$$

$$= \frac{1}{n}\lim_{n\to\infty}(\beta - \hat{\beta}_n)^T\lim_{n\to\infty}\mathbf{X}^T Q\mathbf{X}\lim_{n\to\infty}(\beta - \hat{\beta}_n) \text{ because } \lim_{n\to\infty}\mathbf{X}^T Q\mathbf{X} \text{ is finite} \tag{B.119}$$

$$= 0. \tag{B.120}$$

Therefore:

$$\frac{1}{n}\lim_{n\to\infty}\left\|Q\mathbf{Y} - Q\mathbf{X}\hat{\beta}_n\right\|^2 = \operatorname{var}\|Y_i - X_i\beta\|^2 \tag{B.121}$$

$$= \min_{\tilde{\beta}}\operatorname{var}\left[\left\|Y_i - X_i\tilde{\beta}\right\|\right]. \tag{B.122}$$

□

Now we show the proofs of the main results.

Proof of Lemma 7. For any matrix $\mathbf{s} \in \mathcal{R}^{n \times d}$ we will compute $R^2_{\mathbf{s}} \stackrel{\text{def}}{=} Corr(\hat{\tau}^T_Y, \frac{2}{n}\mathbf{Z}^T\mathbf{s})$ where for any $Y \in \mathcal{R}, X \in \mathcal{R}^d$, $Corr(Y, X)$ is defined as:

$$Corr(Y, X) = Corr(Y, X^T\beta^{*)} \tag{B.123}$$

$$= \frac{Cov(Y, X^T\beta^*)}{\sqrt{\text{var}(Y)}\sqrt{\text{var}(X^T\beta^*)}} \tag{B.124}$$

where $\beta^* = \arg\min_{\hat{\beta}} \mathbb{E}\|Y - X^T\hat{\beta}\|^2$. Substituting $\mathbf{s} = \mathbf{x}$ and $\mathbf{s} = \tilde{\mathbf{x}}$ will give us $R^2_{\mathbf{x}}$ and $R^2_{\tilde{\mathbf{x}}}$.

Recall that $\tau^T_Y(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{Z}}[\hat{\tau}^T_Y|\mathbf{x}, \mathbf{y}] = \frac{1}{n}\sum_{i=1}^n (y^1_i - y^0_i)$. From Lemma 33, we have:

$$\hat{\tau}^T_Y = \frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{c}} + \tau^T_Y(\mathbf{x}, \mathbf{y})$$

.

We note that conditioning on $\mathbf{y}$, $\tau^T_Y(\mathbf{x}, \mathbf{y})$ is a constant independent of $\mathbf{Z}$. Let $Q \stackrel{\text{def}}{=} \frac{n-1}{n}\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]$ and note that $Q = Q^T$ and $Q = Q^2$. First, let us compute $\beta^* = \arg\min_{\hat{\beta}} \mathbb{E}_{\mathbf{Z}}\|\hat{\tau}^T_Y - \frac{2}{n}\mathbf{Z}^T\mathbf{s}\hat{\beta}\|^2$. We have

$$\beta^* = \arg\min_{\hat{\beta}} \mathbb{E}_{\mathbf{Z}} \|\hat{\tau}_Y^T - \frac{2}{n}\mathbf{Z}^T\mathbf{s}\hat{\beta}\|^2 \tag{B.125}$$

$$= \arg\min_{\hat{\beta}} \mathbb{E}_{\mathbf{Z}} \|\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{c}} + \frac{1}{n}\tau_Y^T(\mathbf{x},\mathbf{y}) - \frac{2}{n}\mathbf{Z}^T\mathbf{s}\hat{\beta}\|^2 \tag{B.126}$$

$$= \arg\min_{\hat{\beta}} \mathbb{E}_{\mathbf{Z}} \|\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{c}} - \frac{2}{n}\mathbf{Z}^T\mathbf{s}\hat{\beta}\|^2 \tag{B.127}$$

$$+ 2\frac{1}{n}\tau_Y^T(\mathbf{x},\mathbf{y})\mathbb{E}\left[\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{c}} - \frac{2}{n}\mathbf{Z}^T\mathbf{s}\hat{\beta}\right] + \left(\frac{1}{n}\tau_Y^T(\mathbf{x},\mathbf{y})\right)^2 \tag{B.128}$$

$$= \arg\min_{\hat{\beta}} \mathbb{E}_{\mathbf{Z}} \|\mathbf{Z}^T\tilde{\mathbf{c}} - \mathbf{Z}^T\mathbf{s}\hat{\beta}\|^2 \text{ because } \mathbb{E}[\mathbf{Z}|\mathbf{x},\mathbf{y}] = 0 \text{ from Lemma 30} \tag{B.129}$$

$$= \arg\min_{\hat{\beta}} (\tilde{\mathbf{c}} - \mathbf{s}\hat{\beta})^T\mathbb{E}[\mathbf{Z}\mathbf{Z}^T](\tilde{\mathbf{c}} - \mathbf{s}\hat{\beta}) \tag{B.130}$$

$$= \arg\min_{\hat{\beta}} (\tilde{\mathbf{c}} - \mathbf{s}\hat{\beta})^T Q(\tilde{\mathbf{c}} - \mathbf{s}\hat{\beta}) \tag{B.131}$$

$$= \arg\min_{\hat{\beta}} (\tilde{\mathbf{c}} - \mathbf{s}\hat{\beta})^T Q^T Q(\tilde{\mathbf{c}} - \mathbf{s}\hat{\beta}) \tag{B.132}$$

$$= \arg\min_{\hat{\beta}} \|Q\tilde{\mathbf{c}} - Q\mathbf{s}\hat{\beta}\|^2. \tag{B.133}$$

Using the fact that $Q = Q^TQ$, we have $\beta^* = (\mathbf{s}^TQ\mathbf{s})^{-1}\mathbf{s}^TQ\tilde{\mathbf{c}}$. By definition, we have

$$Corr(\hat{\tau}_Y^T, \frac{2}{n}\mathbf{Z}^T\mathbf{s}) = \frac{\mathbb{E}_{\mathbf{Z}}\left[\hat{\tau}_Y^T\frac{2}{n}\mathbf{Z}^T\mathbf{s}\beta^*\right] - \mathbb{E}_{\mathbf{Z}}\left[\hat{\tau}_Y^T\right]\mathbb{E}_{\mathbf{Z}}\left[\frac{2}{n}\mathbf{Z}^T\mathbf{s}\beta^*\right]}{\sqrt{\text{var}_{\mathbf{Z}}(\hat{\tau}_Y^T)\text{var}_{\mathbf{Z}}(\frac{2}{n}\mathbf{Z}^T\mathbf{s}\beta^*)}} \tag{B.134}$$

$$= \frac{\mathbb{E}_{\mathbf{Z}}\left[\hat{\tau}_Y^T\mathbf{Z}^T\mathbf{s}\beta^*\right]}{\sqrt{\text{var}_{\mathbf{Z}}(\hat{\tau}_Y^T)\text{var}_{\mathbf{Z}}(\mathbf{Z}^T\mathbf{s}\beta^*)}} \text{ because } \mathbb{E}[\mathbf{Z}] = 0 \tag{B.135}$$

$$= \frac{\mathbb{E}_{\mathbf{Z}}\left[\left(\frac{2}{n}\tilde{\mathbf{c}}^T\mathbf{Z} + \tau_Y^T(\mathbf{x},\mathbf{y})\right)\mathbf{Z}^T\mathbf{s}\beta^*\right]}{\sqrt{\text{var}_{\mathbf{Z}}\left(\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{c}} + \tau_Y^T(\mathbf{x},\mathbf{y})\right)\text{var}_{\mathbf{Z}}(\mathbf{Z}^T\mathbf{s}\beta^*)}} \tag{B.136}$$

$$= \frac{\mathbb{E}_{\mathbf{Z}}\left[\left(\frac{2}{n}\tilde{\mathbf{c}}^T\mathbf{Z}\right)\mathbf{Z}^T\mathbf{s}\beta^*\right]}{\sqrt{\text{var}_{\mathbf{Z}}\left(\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{c}}\right)\text{var}_{\mathbf{Z}}(\mathbf{Z}^T\mathbf{s}\beta^*)}} \tag{B.137}$$

$$= \frac{\mathbb{E}_{\mathbf{Z}}\left[\tilde{\mathbf{c}}^T\mathbf{Z}\mathbf{Z}^T\mathbf{s}\beta^*\right]}{\sqrt{\text{var}_{\mathbf{Z}}\left(\mathbf{Z}^T\tilde{\mathbf{c}}\right)\text{var}_{\mathbf{Z}}(\mathbf{Z}^T\mathbf{s}\beta^*)}} \tag{B.138}$$

For the numerator we have:

$$\mathbb{E}_{\mathbf{Z}}\left[\tilde{\mathbf{c}}^T\mathbf{Z}\mathbf{Z}^T\mathbf{s}\beta^*\right] = \tilde{\mathbf{c}}^TQ\mathbf{s}\beta^* \tag{B.139}$$

$$= \frac{n}{n-1}\tilde{\mathbf{c}}^TQ\mathbf{s}(\mathbf{s}^TQ\mathbf{s})^{-1}\mathbf{s}^TQ\tilde{\mathbf{c}} \tag{B.140}$$

$$= \frac{n}{n-1}\tilde{\mathbf{c}}^TQ\mathbf{s}(\mathbf{s}^TQ\mathbf{s})^{-1}\mathbf{s}^TQ\mathbf{s}(\mathbf{s}^TQ\mathbf{s})^{-1}\mathbf{s}^TQ\tilde{\mathbf{c}} \tag{B.141}$$

$$= \frac{n}{n-1}\left(\tilde{\mathbf{c}}^TQ\mathbf{s}(\mathbf{s}^TQ\mathbf{s})^{-1}\mathbf{s}^TQ\right)\left(Q\mathbf{s}(\mathbf{s}^TQ\mathbf{s})^{-1}\mathbf{s}^TQ\tilde{\mathbf{c}}\right) \tag{B.142}$$

$$= \frac{n}{n-1}(\beta^{*T}\mathbf{s}^TQ)(Q\mathbf{s}\beta^*) \tag{B.143}$$

$$= \frac{n}{n-1}\|Q\mathbf{s}\beta^*\|^2 \tag{B.144}$$

Since $\beta^* = \arg\min_{\hat{\beta}}\|Q\tilde{\mathbf{c}} - Q\mathbf{s}\hat{\beta}\|^2$, $Q\tilde{\mathbf{c}} - Q\mathbf{s}\beta^*$ and $Q\mathbf{s}\beta^*$ are orthogonal, and therefore $\|Q\mathbf{s}\beta^*\|^2 = \|Q\tilde{\mathbf{c}}\|^2 - \|Q\tilde{\mathbf{c}} - Q\mathbf{s}\beta^*\|^2$.

For the denominator, since $\mathbb{E}[\mathbf{Z}] = 0$ we have:

$$\text{var}_{\mathbf{Z}}\left(\mathbf{Z}^T\tilde{\mathbf{c}}\right)\text{var}_{\mathbf{Z}}(\mathbf{Z}^T\mathbf{s}\beta^*) = \mathbb{E}_{\mathbf{Z}}[\tilde{\mathbf{c}}^T\mathbf{Z}\mathbf{Z}^T\tilde{\mathbf{c}}]\mathbb{E}_{\mathbf{Z}}[\beta^{*T}\mathbf{s}^T\mathbf{Z}\mathbf{Z}^T\mathbf{s}\beta^*] \tag{B.145}$$

$$= \frac{n^2}{(n-1)^2}(\tilde{\mathbf{c}}^TQ\tilde{\mathbf{c}})(\beta^{*T}\mathbf{s}^TQ\mathbf{s}\beta^*) \tag{B.146}$$

$$= \frac{n^2}{(n-1)^2}(\tilde{\mathbf{c}}^TQ^TQ\tilde{\mathbf{c}})(\beta^{*T}\mathbf{s}^TQ^TQ\mathbf{s}\beta^*) \tag{B.147}$$

$$= \frac{n^2}{(n-1)^2}\|Q\tilde{\mathbf{c}}\|^2\|Q\mathbf{s}\beta^*\|^2 \tag{B.148}$$

Putting the numerator and denominator together we have:

$$R^2_{\mathbf{s}} = Corr(\hat{\tau}_Y^T, \frac{2}{n}\mathbf{Z}^T\mathbf{s}) \tag{B.149}$$

$$= \frac{\|Q\mathbf{s}\beta^*\|^2}{\|Q\tilde{\mathbf{c}}\|\|Q\mathbf{s}\beta^*\|} \tag{B.150}$$

$$= \frac{\|Q\mathbf{s}\beta^*\|}{\|Q\tilde{\mathbf{c}}\|} \tag{B.151}$$

$$= \frac{\sqrt{\|Q\tilde{\mathbf{c}}\|^2 - \|Q\tilde{\mathbf{c}} - Q\mathbf{s}\beta^*\|^2}}{\|Q\tilde{\mathbf{c}}\|} \tag{B.152}$$

Substituting $\mathbf{s} = \mathbf{x}$ and $\mathbf{s} = \tilde{\mathbf{x}}$ gives us the expression for $R^2_{\mathbf{x}}$ and $R^2_{\tilde{\mathbf{x}}}$. □

Proof of Theorem 7. We have

$$\tilde{C} = \tilde{X}^T\beta + \tilde{\mathcal{E}} \tag{B.153}$$

where $C = \frac{Y^0+Y^1}{2}$, $\mathcal{E} = \frac{\mathcal{E}_0+\mathcal{E}_1}{2}$, $\beta = \frac{\beta_0+\beta_1}{2}$, $\tilde{C} = \frac{p_T(X)}{p_S(X)}C, \tilde{X} = \frac{p_T(X)}{p_S(X)}X$ and $\tilde{\mathcal{E}} = \frac{p_T(X)}{p_S(X)}\mathcal{E}$. Since $Y^a, a \in \{0,1\}$, $X(j), 1 \leq j \leq d$ and W have finite 8th moment, $\tilde{C}$ and $\tilde{X}$ have finite 4th moment using Cauchy-Schwartz inequality. Let $S \in \mathcal{R}^d$ be a random variable independent of $\mathcal{E}_i$ and $S(j), 1 \leq j \leq d$ have finite 4th moment. Let $\mathbf{S} \in \mathcal{R}^{n\times d}$ be n samples $S_1, \cdots, S_n$ of S. By the definition of R^2,

$$R^2_{\mathbf{S}} = \frac{\|Q\tilde{\mathbf{C}}\|^2 - \min_{\hat{\beta}} \|Q\tilde{\mathbf{C}} - Q\mathbf{S}\hat{\beta}\|^2}{\|Q\tilde{\mathbf{C}}\|^2}. \tag{B.154}$$

We will show that $\lim_{n\to\infty} R^2_{\tilde{\mathbf{X}}} \geq \lim_{n\to\infty} R^2_{\mathbf{S}}$ almost surely for any S. We have:

$$\frac{1}{n}\lim_{n\to\infty} \min_{\hat{\beta}} \|Q\tilde{\mathbf{C}} - Q\mathbf{S}\hat{\beta}\|^2 \tag{B.155}$$

$$= \min_{\hat{\beta}} \operatorname{var}(\tilde{C} - S^T\hat{\beta}) \text{ almost surely by Lemma 36} \tag{B.156}$$

$$= \min_{\hat{\beta}} \mathbb{E}[(\tilde{C} - S^T\hat{\beta})^2] - \left(\mathbb{E}[\tilde{C} - S^T\hat{\beta}]\right)^2 \tag{B.157}$$

$$= \min_{\hat{\beta}} \mathbb{E}[\tilde{X}^T\beta - S^T\hat{\beta}]^2 + \mathbb{E}[\tilde{\mathcal{E}}^2] - \left(\mathbb{E}[\tilde{X}^T\beta - S^T\hat{\beta}]\right)^2 \tag{B.158}$$

$$\text{because } \mathbb{E}[\tilde{\mathcal{E}}] = 0 \text{ and } \mathcal{E} \text{ is independent of } \tilde{X} \text{ and } S \tag{B.159}$$

$$= \min_{\hat{\beta}} \operatorname{var}(\tilde{X}^T\beta - S^T\hat{\beta}) + \mathbb{E}[\tilde{\mathcal{E}}^2] \tag{B.160}$$

$$\geq \mathbb{E}[\tilde{\mathcal{E}}^2]. \tag{B.161}$$

When $S = \tilde{X}$, this is minimized, therefore:

$$\lim_{n\to\infty} R^2_{\tilde{\mathbf{X}}} \geq \lim_{n\to\infty} R^2_{\mathbf{S}} \text{ almost surely.} \tag{B.162}$$

Substituting $\mathbf{S} = \mathbf{X}$:

$$\lim_{n\to\infty} R^2_{\tilde{\mathbf{X}}} \geq \lim_{n\to\infty} R^2_{\mathbf{X}} \text{ almost surely.} \tag{B.163}$$

Recall that when Assumption 7 is satisfied,

$$\text{as-var}_{\mathbf{Z}}\left(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{x},\mathbf{y}, M\left(\frac{2}{n}\mathbf{Z}^T\mathbf{s}\right) \leq a\right) \tag{B.164}$$

$$= \text{as-var}_{\mathbf{Z}}(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{x},\mathbf{y})(1 - (1 - v_{d,a}) \lim_{n\to\infty} R^2_{\mathbf{s}}), \tag{B.165}$$

where as-var is the variance of the asymptotic sampling distribution.

We have:

$$\text{as-var}_{\mathbf{Z}}\left(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{X},\mathbf{Y}, \phi^a_S = 1\right) \tag{B.166}$$

$$= \text{as-var}_{\mathbf{Z}}\left(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{X},\mathbf{Y}, M\left(\frac{2}{n}\mathbf{Z}^T\mathbf{X}\right) \leq a\right) \tag{B.167}$$

$$= \text{as-var}_{\mathbf{Z}}(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{X},\mathbf{Y})(1 - (1 - v_{d,a}) \lim_{n\to\infty} R^2_{\mathbf{X}}) \tag{B.168}$$

$$\geq \text{as-var}_{\mathbf{Z}}(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{X},\mathbf{Y})(1 - (1 - v_{d,a}) \lim_{n\to\infty} R^2_{\tilde{\mathbf{X}}}) \text{ almost surely} \tag{B.169}$$

$$= \text{as-var}_{\mathbf{Z}}\left(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{X},\mathbf{Y}, M\left(\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{X}}\right) \leq a\right) \tag{B.170}$$

$$= \text{as-var}_{\mathbf{Z}}\left(\sqrt{n}(\hat{\tau}^T_Y - \tau_{\mathbf{x},\mathbf{y}})|\mathbf{X},\mathbf{Y}, \phi^a_T = 1\right). \tag{B.171}$$

And therefore:

$$\text{as-var}_{\mathbf{Z}}\left(\hat{\tau}^T_Y|\mathbf{X},\mathbf{Y}, \phi^a_S = 1\right) \geq \text{as-var}_{\mathbf{Z}}\left(\hat{\tau}^T_Y|\mathbf{X},\mathbf{Y}, \phi^a_T = 1\right). \tag{B.172}$$

Now we analyze the rejection probability. Let $U \in \mathcal{R}^d$ be a standard multivariate random variable. We have:

$$\lim_{n\to\infty} \mathbb{P}(\phi_S^a = 0|\mathbf{x}) = \lim_{n\to\infty} \mathbb{P}(M\left(\frac{2}{n}\mathbf{Z}^T\mathbf{x}\right) \leq a) \tag{B.173}$$

$$= \lim_{n\to\infty} \mathbb{P}(\|B_S\|^2 < a) \text{ where } B_S = \frac{2}{n}\mathbf{Z}^T\mathbf{x}Cov(\frac{2}{n}\mathbf{Z}^T\mathbf{x})^{-1/2} \tag{B.174}$$

$$= \mathbb{P}(\|U\|^2 < a), \tag{B.175}$$

because B_S converges in distribution to U by finite central limit theorem. Similarly we have:

$$\lim_{n\to\infty} \mathbb{P}(\phi_T^a = 0|\mathbf{x}) = \lim_{n\to\infty} \mathbb{P}(M\left(\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{x}}\right) \leq a) \tag{B.176}$$

$$= \lim_{n\to\infty} \mathbb{P}(\|B_T\|^2 < a) \text{ where } B_T \stackrel{\text{def}}{=} \frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{x}}Cov(\frac{2}{n}\mathbf{Z}^T\tilde{\mathbf{x}})^{-1/2} \tag{B.177}$$

$$= \mathbb{P}(\|U\|^2 < a), \tag{B.178}$$

because B_T converges in distribution to U by finite central limit theorem.

Therefore $lim_{n\to\infty}\mathbb{P}(\phi_S^a = 0|\mathbf{x}) = \lim_{n\to\infty} \mathbb{P}(\phi_T^a = 0|\mathbf{x})$. Asymptotically, with the same rejection probability, using Target Balance results in a smaller variance than Source Balance . □

B.5 Proofs of Section B.1

In this Section we present the proof of Lemma 27, Lemma 28, Lemma 29, Theorem 11 and Corollary 4.

In order to prove Lemma 27, we first prove the following lemma.

Lemma 37 (minor changes to Lemma 1 in Harshaw et al. (2019))**.** *Let* $\tilde{\epsilon}_i = \tilde{c}_i - \beta^T\tilde{x}_i$ *and* $\tilde{\boldsymbol{\epsilon}} = (\tilde{\epsilon}_1, \cdots, \tilde{\epsilon}_n)$*. For any function* $\rho(\mathbf{x}, \mathbf{Z}) \in \{0, 1\}$ *satisfying* $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$*:*

$$\frac{n^2}{4} var_{\mathbf{Z}}(\hat{\tau}_Y^T|\mathbf{x}, \mathbf{y}, \rho = 1) \tag{B.179}$$

$$= Cov(\tilde{\mathbf{c}}^T\mathbf{Z}|\rho = 1) \tag{B.180}$$

$$= \beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z}|\rho = 1]\beta + Cov(\tilde{\boldsymbol{\epsilon}}^T\mathbf{Z}|\rho = 1) + 2\beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z}, \tilde{\boldsymbol{\epsilon}}^T\mathbf{Z}|\rho = 1) \tag{B.181}$$

Proof of Lemma 37. By definition:

$$\text{var}_{\mathbf{Z}}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{y},\rho=1)=\mathbb{E}_{\mathbf{Z}}\left[(\hat{\tau}_Y^T-\mathbb{E}_{\mathbf{Z}}[\hat{\tau}_Y^T|\mathbf{x},\mathbf{y},\rho=1])^2|\mathbf{x},\mathbf{y},\rho=1\right] \tag{B.182}$$

We have:

$$\mathbb{E}_{\mathbf{Z}}[\hat{\tau}_Y^T|\mathbf{x},\mathbf{y},\rho=1] \tag{B.183}$$

$$=\frac{2}{n}\mathbb{E}_{\mathbf{Z}}\left[\sum_{Z_i=1}w_iy_i^*-\sum_{Z_i=-1}w_iy_i^*\middle|\rho=1\right] \tag{B.184}$$

$$=\frac{2}{n}\mathbb{E}\left[\sum_{i=1}^n A_iw_iy_i^1-\sum_{i=1}^n(1-A_i)w_iy_i^0\middle|\rho=1\right] \tag{B.185}$$

$$=\frac{2}{n}\left(\sum_{i=1}^n\mathbb{E}[A_i|\rho=1]w_iy_i^1-\sum_{i=1}^n\mathbb{E}[1-A_i|\rho=1]w_iy_i^0\right) \tag{B.186}$$

$$=\frac{1}{n}\left(\sum_{i=1}^n w_iy_i^1-\sum_{i=1}^n w_iy_i^0\right)\text{ because }\mathbb{E}[A_i|\rho=1]=1/2\text{ by Lemma 30.} \tag{B.187}$$

Therefore using Lemma 33:

$$\text{var}_{\mathbf{Z}}(\hat{\tau}_Y^T|\mathbf{x},\mathbf{y},\rho=1) \tag{B.188}$$

$$=\mathbb{E}_{\mathbf{Z}}\left[\left(\frac{2}{n}(\sum_{Z_i=1}w_iy_i^*-\sum_{Z_i=-1}w_iy_i^*)-\frac{1}{n}\sum_{i=1}^n w_i(y_i^1-y_i^0)\right)^2\middle|\mathbf{x},\mathbf{y},\rho=1\right] \tag{B.189}$$

$$=\frac{4}{n^2}\mathbb{E}[\tilde{\mathbf{c}}^T\mathbf{Z}\mathbf{Z}^T\tilde{\mathbf{c}}|\mathbf{x},\mathbf{y},\rho=1] \tag{B.190}$$

$$=\frac{4}{n^2}Cov(\tilde{\mathbf{c}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1)\text{ because }\mathbb{E}[\tilde{\mathbf{c}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1]=0\text{ from Lemma 30} \tag{B.191}$$

$$=\frac{4}{n^2}Cov((\tilde{\mathbf{x}}\beta+\tilde{\boldsymbol{\epsilon}})^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1) \tag{B.192}$$

$$=\beta^TCov(\tilde{\mathbf{x}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1)\beta \tag{B.193}$$

$$+Cov(\tilde{\boldsymbol{\epsilon}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1)+2\beta^TCov(\tilde{\mathbf{x}}^T\mathbf{Z},\tilde{\boldsymbol{\epsilon}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1). \tag{B.194}$$

□

Proof of Lemma 27. By law of total variance:

$$\text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) = \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right] + \text{var}^S_{\mathbf{Y}}(\mathbb{E}_{\mathbf{Z}}[\hat{\tau}^T_Y|\mathbf{x},\mathbf{Y},\rho=1]|\mathbf{x}) \quad \text{(B.195)}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right] + \text{var}^S_{\mathbf{Y}}\left(\frac{1}{n}\sum_{i=1}^n w_i(Y_i^1 - Y_i^0)|\mathbf{x}\right) \quad \text{(B.196)}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right] + \frac{1}{n^2}\sum_{i=1}^n w_i^2 \text{var}(\mathcal{E}_1 - \mathcal{E}_0) \quad \text{(B.197)}$$

$$= \mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right] + \frac{2}{n^2}\sigma^2_{\mathcal{E}}\sum_{i=1}^n w_i^2. \quad \text{(B.198)}$$

From Lemma 37 we can analyze the term inside the expectation in the first term of Eq. B.198:

$$\frac{n^2}{4}\text{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{y},\rho=1) \quad \text{(B.199)}$$

$$= \beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1]\beta + Cov(\tilde{\boldsymbol{\epsilon}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1) + 2\beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z},\tilde{\boldsymbol{\epsilon}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1) \quad \text{(B.200)}$$

$$= \beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1]\beta + \tilde{\boldsymbol{\epsilon}}^T Cov(\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1)\tilde{\boldsymbol{\epsilon}} + 2\beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z},\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1)\tilde{\boldsymbol{\epsilon}}. \quad \text{(B.201)}$$

Recall that $Y_i^1 = \beta_1^T X_i + \mathcal{E}_i^1$ and $Y_i^0 = \beta_1^T X_i + \mathcal{E}_i^0$. Let $\mathcal{E}_i = \frac{\mathcal{E}_i^1 + \mathcal{E}_i^0}{2}$ and $\tilde{\boldsymbol{\mathcal{E}}} = (\mathcal{E}_1, \cdots, \mathcal{E}_n)$. Since $\tilde{\boldsymbol{\epsilon}}$ is the value of $\tilde{\boldsymbol{\mathcal{E}}}$ we can analyze the first term of Eq. B.198:

$$\frac{n^2}{4}\mathbb{E}^S_{\mathbf{Y}}\left[\text{var}_{\mathbf{Z}}(\hat{\tau}^T_Y|\mathbf{x},\mathbf{Y},\rho=1)|\mathbf{x}\right] \quad \text{(B.202)}$$

$$= \beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1]\beta + \mathbb{E}^S_{\mathbf{Y}}[\tilde{\boldsymbol{\mathcal{E}}}^T Cov(\mathbf{Z}|\mathbf{x},\mathbf{Y},\rho=1)\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{x}] \quad \text{(B.203)}$$

$$+ 2\beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z},\mathbf{Z}|\mathbf{x},\mathbf{y},\rho=1)\mathbb{E}[\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{x}] \quad \text{(B.204)}$$

$$= \beta^T Cov(\tilde{\mathbf{x}}^T\mathbf{Z}|\mathbf{x},\rho=1]\beta + \mathbb{E}^S_{\mathbf{Y}}[Cov(\tilde{\boldsymbol{\mathcal{E}}}^T\mathbf{Z}|\mathbf{x},\rho=1)|\mathbf{x}] \text{ because } \mathbb{E}[\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{x}] = 0. \quad \text{(B.205)}$$

Now we analyze the second term of Eq. B.198:

$$\mathbb{E}^S_{\mathbf{Y}}[Cov(\tilde{\boldsymbol{\mathcal{E}}}^T\mathbf{Z}|\mathbf{x},\rho=1)|\mathbf{x}] \tag{B.206}$$

$$= \mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}[\mathbb{E}_{\mathbf{Z}}[\tilde{\boldsymbol{\mathcal{E}}}^T\mathbf{Z}\mathbf{Z}^T\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{x},\rho=1]|\mathbf{x}] \tag{B.207}$$

$$= \mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}_{\mathbf{Z}}[w_i\mathcal{E}_iZ_iZ_j\mathcal{E}_jw_j|\rho=1]|\mathbf{x}\right] \tag{B.208}$$

$$= \mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}\left[\sum_{i=1}^{n}\mathbb{E}[w_i^2\mathcal{E}_i^2Z_i^2|\rho=1] + \sum_{i\neq j}\mathbb{E}[w_i\mathcal{E}_iZ_iZ_j\mathcal{E}_jw_j|\rho=1]|\mathbf{x}\right] \tag{B.209}$$

$$= \mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}\left[\sum_{i=1}^{n}\mathbb{E}[w_i^2\mathcal{E}_i^21|\rho=1] + \sum_{i\neq j}\mathbb{E}[w_iZ_iZ_j\mathcal{E}_jw_j|\rho=1]\mathbb{E}[\mathcal{E}_i|\rho=1]|\mathbf{x}\right] \text{ because } Z_i^2=1 \tag{B.210}$$

$$= \sum_{i=1}^{n}\mathbb{E}[w_i^2\mathcal{E}_i^2] \text{ because } \mathbb{E}[\mathcal{E}_i|\rho=1]=0 \tag{B.211}$$

$$= \sum_{i=1}^{n}w_i^2\sigma_{\mathcal{E}}^2. \tag{B.212}$$

Putting all together Eq. B.198 becomes:

$$\text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) = \frac{4}{n^2}\left(\beta^TCov(\tilde{\mathbf{x}}^T\mathbf{Z}|\mathbf{x},\rho=1]\beta + \sum_{i=1}^{n}w_i^2\sigma_{\mathcal{E}}^2\right) + +\frac{2}{n^2}\sigma_{\mathcal{E}}^2\sum_{i=1}^{n}w_i^2 \tag{B.213}$$

$$= \frac{4}{n^2}\beta^TCov_{\mathbf{Z}}(\tilde{\mathbf{x}}^T\mathbf{Z}|\mathbf{x},\rho=1]\beta + \frac{6}{n^2}\sigma_{\mathcal{E}}^2\sum_{i=1}^{n}w_i^2. \tag{B.214}$$

□

Proof of Lemma 28. We use the same decomposition of $\beta^TCov_{\mathbf{Z}}(V|\mathbf{x},\Omega)\beta$ as in Harshaw et al. (2019). Let $\mathbf{e}_1,...,\mathbf{e}_n$ and $\lambda_1,..,\lambda_n$ be the normalized eigenvectors and corresponding eigenvalues of matrix $Cov_{\mathbf{Z}}(V|\mathbf{x},\Omega)$. Since $Cov_{\mathbf{Z}}(V|\mathbf{x},\Omega)$ is symmetric, the eigenvectors form an orthonormal basis so we can write β as a linear combination of $\mathbf{e}_1,..,\mathbf{e}_n$ and get:

$$\beta = \|\beta\|\sum_{i=1}^{n}\eta_i\mathbf{e}_i, \tag{B.215}$$

where $\eta_i = \langle \beta, \mathbf{e}_i \rangle / \|\beta\|$ is the coefficient that captures the alignment of the weighted outcome β with respect to the eigenvector $\mathbf{e}_i$. Therefore:

$$\beta^T Cov_{\mathbf{Z}}(V|\mathbf{x}, \Omega)\beta = \|\beta\|^2 \sum_{i=1}^{n} \eta_i^2 \lambda_i. \tag{B.216}$$

Then:

$$\mathbb{E}_\beta \left[\beta^T Cov_{\mathbf{Z}}(V|\mathbf{x}, \Omega)\beta\right] \tag{B.217}$$

$$= \mathbb{E}_\beta \left[\|\beta\|^2 \sum_{i=1}^{n} \eta_i^2 \lambda_i \right] \tag{B.218}$$

$$= l^2 \sum_{i=1}^{n} \lambda_i \mathbb{E}_\beta[\eta_i^2] \tag{B.219}$$

$$= l^2 \sum_{i=1}^{n} \lambda_i \mathbb{E}_\theta cos^2(\theta) \tag{B.220}$$

where θ is the angle between β and $\mathbf{e}_i$. Since β points to any direction with equal probability, θ is uniformly distributed in $[0, 2\pi]$.

$$= \frac{l^2}{2} \pi \sum_{i=1}^{n} \lambda_i \tag{B.221}$$

$$= \frac{l^2}{2} \pi \text{Trace}(Cov_{\mathbf{Z}}(V|\mathbf{x}, \Omega)). \tag{B.222}$$

□

Proof of Lemma 29. Let $p(u)$ be the pmf of U. Define $f(u)$ as follow:

$$f(u) = p(u)\mathbf{1}(u \in \Omega). \tag{B.223}$$

Then:

$$p(u|\Omega) = \frac{p(u)\mathbf{1}(u \in \Omega)}{\mathbb{P}(\Omega)} = \frac{f(u)}{1 - \alpha}. \tag{B.224}$$

Since $\mathbb{P}(\Omega) \geq 1 - \alpha$ we have:

$$\sum_u f(u) \geq 1 - \alpha \tag{B.225}$$

We have:

$$\begin{aligned} Trace(Cov(U|\Omega)) &= Trace(\mathbb{E}[UU^T|\Omega]) && \text{(B.226)} \\ &= Trace(\mathbb{E}[UU^T|\Omega] && \text{(B.227)} \\ &= Trace(\mathbb{E}[U^TU|\Omega] && \text{(B.228)} \\ &= \sum_u u^T u \frac{f(u)}{1-\alpha}\,. && \text{(B.229)} \end{aligned}$$

We want to minimize $Trace(Cov(U|\Omega))$:

$$\sum_u u^T u \frac{f(u)}{1-\alpha}\,. \tag{B.230}$$

subject to:

$$\begin{aligned} f(u) &\in \{0, p(u)\}\ \forall u, && \text{(B.231)} \\ \sum_u f(u) &\geq 1 - \alpha\,. && \text{(B.232)} \end{aligned}$$

This can be done by maximize $f(u)$ so that $f(u) = p(u)$ for the smallest $u^T u$, which is equal to set Ω to be the event $\|U\|^2 < u_\alpha$. □

Proof of Theorem 11. Since $\rho(\mathbf{x}, \mathbf{Z}) = \rho(\mathbf{x}, -\mathbf{Z})$, from Lemma 30 we have $\mathbb{E}_{\mathbf{Z}}[V|\mathbf{x}, \rho = 1] = 0$, which satisfies the criteria in Lemma 29. From Lemma 27:

$$\mathbb{E}_\beta \text{var}^S_{\mathbf{Y},\mathbf{Z}_\rho}(\hat{\tau}^T_Y|\mathbf{x}) \tag{B.233}$$

$$= \frac{4}{n^2}\mathbb{E}_\beta \beta^T Cov(V|\mathbf{x}, \rho = 1)\beta + \frac{6}{n^2}\sigma^2 \sum_{i=1}^n w_i^2 \tag{B.234}$$

$$= \frac{4}{n^2}\frac{l^2}{2}\pi \text{Trace}(Cov(V|\mathbf{x}, \rho = 1)) + \frac{6}{n^2}\sigma^2 \sum_{i=1}^n w_i^2 \text{ by Lemma 28} \tag{B.235}$$

$$\geq \frac{4}{n^2}\frac{l^2}{2}\pi \text{Trace}(Cov(V|\mathbf{x}, \|V\|^2 < v_\alpha)) + \frac{6}{n^2}\sigma^2 \sum_{i=1}^n w_i^2 \text{ by Lemma 29} \tag{B.236}$$

$$= \frac{4}{n^2}\frac{l^2}{2}\pi \text{Trace}(Cov(V|\mathbf{x}, \phi_T^{\alpha\prime} = 1)) + \frac{6}{n^2}\sigma^2 \sum_{i=1}^n w_i^2 \tag{B.237}$$

$$= \frac{4}{n^2}\mathbb{E}_\beta \beta^T Cov(V|\mathbf{x}, \phi_T^{\alpha\prime} = 1)\beta + \frac{6}{n^2}\sigma^2 \sum_{i=1}^n w_i^2 \text{ by Lemma 28} \tag{B.238}$$

$$= \mathbb{E}_\beta \text{var}^S_{\mathbf{Y},\mathbf{Z}_{\phi_T^{\alpha\prime}}}(\hat{\tau}^T_Y|\mathbf{x}). \tag{B.239}$$

□

Proof of Corollary 4. Let ρ being the constant function $\rho(\mathbf{x}, \mathbf{Z}) = 1$ for all $\mathbf{x}, \mathbf{Z}$. Then:

$$\text{var}^S_{\mathbf{Z}_\rho,\mathbf{Y}}(\hat{\tau}^T_Y|\mathbf{x}) = \text{var}^S_{\mathbf{Z},\mathbf{Y}}(\hat{\tau}^T_Y|\mathbf{x}). \tag{B.240}$$

From Theorem 11 we have:

$$\mathbb{E}_\beta \text{var}^S_{\mathbf{Z}_{\phi_T'},\mathbf{Y}}(\hat{\tau}^T_Y|\mathbf{x}) \leq \mathbb{E}_\beta \text{var}^S_{\mathbf{Z}_\rho,\mathbf{Y}}(\hat{\tau}^T_Y|\mathbf{x}) = \mathbb{E}_\beta \text{var}^S_{\mathbf{Z},\mathbf{Y}}(\hat{\tau}^T_Y|\mathbf{x}). \tag{B.241}$$

□

APPENDIX C

SUPPLEMENTARY MATERIAL: TOWARDS PRACTICAL MEAN BOUNDS FOR SMALL SAMPLES

- In Section C.1 we describe the computation of Anderson's bound and present more experiments.
- In Section C.2, as noted in Section 3.1, we show a log-normal distribution where the sample mean distribution is visibly skewed when $n = 80$.
- In Section C.3 we present the proofs of Section 3.2.1.2.
- In Section C.4 we discuss the Monte Carlo convergence result of our approximation in Section 3.3.
- In Section C.5.1 we show that our bound reduces to the Clopper-Pearson bound for binomial distributions as mentioned in Section 3.4.1. In Section C.5.2 we present the proofs of Section 3.4.2. In Section C.5.3 we showed that our bound reduces to Anderson's bound when the lower bound of the support is too small.

C.1 Other Experiments

In this section we perform experiments to find an upper bound of the mean of distributions given a finite upper bound of the support, or to a lower bound of the mean of distributions given a finite lower bound of the support. We find the lower bound of the mean of a random variable X by finding the upper mean bound of $-X$ and negating it to obtain the lower mean bound of X.

First we describe the computation of Anderson's bound with $\mathbf{u}^{\text{And}}$ defined in Eq. 3.44. We compute $\beta(n)$ through Monte Carlo simulations. $\beta(n)$ is the value such that:

$$\mathbb{P}_{\mathbf{U}}(\ \forall i : 1 \leq i \leq n, U_{(i)} \geq i/n - \beta(n)) = 1 - \alpha. \tag{C.1}$$

Therefore

$$\mathbb{P}_{\mathbf{U}}(\beta(n) \geq \max_{i:1\leq i\leq n}(i/n - U_{(i)})) = 1 - \alpha. \tag{C.2}$$

For each sample size n, we generate $L = 1{,}000{,}000$ samples $\mathbf{U}^j \in [0,1]^n, 1 \leq j \leq L$. For each sample $\mathbf{U}^j \in [0,1]^n$ we compute

$$\beta(n)_j = \max_{i:1\leq i\leq n} i/n - U^j_{(i)}.$$

Let $\beta(n)_1 \leq \cdots \leq \beta(n)_L$ be the sorted values from L samples. We output $\hat{\beta}(n) = \beta(n)_{(\lceil(1-\alpha)L\rceil)}$ as an approximation of $\beta(n)$.

For each experiment, we used $\alpha = 0.05$ unless specified otherwise. We plot the following:

- The expected value of the bounds versus the sample size. For each sample size, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and compute the average.
- For the upper bound of the mean, we plot the α-quantile of the bound distribution versus the sample size. For each sample size, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and take the α quantile. If the α-quantile is below the true mean, the bound does not have guaranteed coverage.

For the lower bound of the mean, we plot the $1-\alpha$-quantile of the bound distribution versus the sample size. For each sample size, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and take the $1-\alpha$ quantile. If the $1-\alpha$-quantile is above the true mean, the bound does not have guaranteed coverage.

- Coverage of the bounds. For each value of α from 0.02 to 1 with a step size of 0.02, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and plot the percentage of the bounds that are greater than or equal to the true mean (denoted *coverage*). If this percentage is larger than $1-\alpha$, the bound has guaranteed coverage.

We perform the following experiments:

- For the case in which we know a superset D^+ of the distribution's support with a finite lower bound and a finite upper bound (the 2-ended support setting), we compare the following bounds:
  - Anderson's bound.
  - New bound with the function T being Anderson's bound.
  - Student's t.
  - Hoeffding's bound.
  - Maurer and Pontil's bound.

  We find an upper bound of the mean for the following distributions:
  - $\beta(1,5)$, $\texttt{uniform}(0,1)$ and $\beta(5,1)$. The known superset of the support is $[0,1]$. The result is in Figure C.1.
  - $\beta(0.5,0.5)$, $\beta(1,1)$ and $\beta(2,2)$. The known superset of the support is $[0,1]$. The result is in Figure C.2.

- $\texttt{binomial}(10, 0.1)$, $\texttt{binomial}(10, 0.5)$ and $\texttt{binomial}(10, 0.9)$. The known superset of the support is the interval $[0, 10]$. The result is in Figure C.3.

- We also consider the case in which we want an upper bound of the mean without knowing the lower bound of the support (or to find a lower bound without knowing an upper bound of the support). In the main paper we referred to this as the 1-ended support setting. Since Hoeffding's and Maurer and Pontil's bounds require knowing both a finite lower bound and upper bound, they are not applicable in this setting. We compare the following bounds:

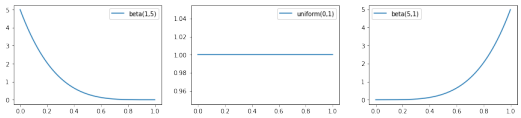
  - Anderson's bound.
  - New bound with T being Anderson's bound.
  - Student's t

  We address the following distributions:

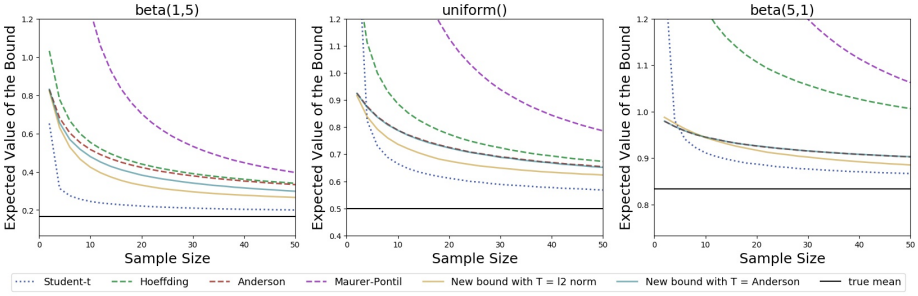
  - $\beta(1, 5)$, $\texttt{uniform}(0, 1)$ and $\beta(5, 1)$. The known superset of the support is $(-\infty, 1]$. We find the upper bound of the mean. The result is in Figure C.4.
  - $\texttt{binomial}(10, 0.1)$, $\texttt{binomial}(10, 0.5)$ and $\texttt{binomial}(10, 0.9)$. The known superset of the support is $(-\infty, 10]$. We find the upper bound of the mean. The result is in Figure C.5.
  - $\texttt{poisson}(2)$, $\texttt{poisson}(10)$ and $\texttt{poisson}(50)$. The known superset of the support is $[0, \infty)$. We find the lower bound of the mean. The result is in Figure C.6.

All the experiments confirm that our bound has guaranteed coverage and is equal to or tighter than Anderson's and Hoeffding's.

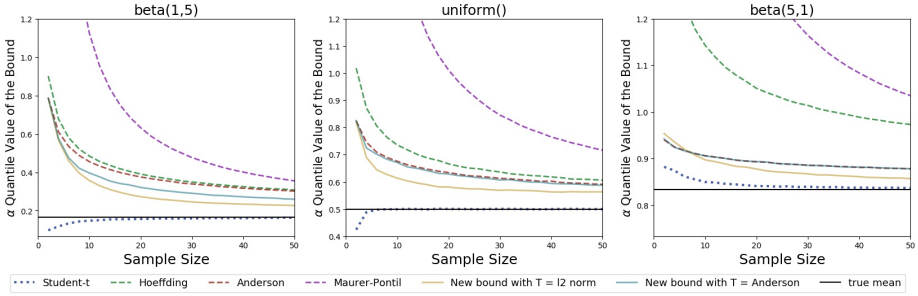
From the experiments, our upper bound performs the best in distributions that are skewed right (respectively, our lower bound will perform the best in distributions that are skewed left), when we know a tight lower bound and upper bound of the support.



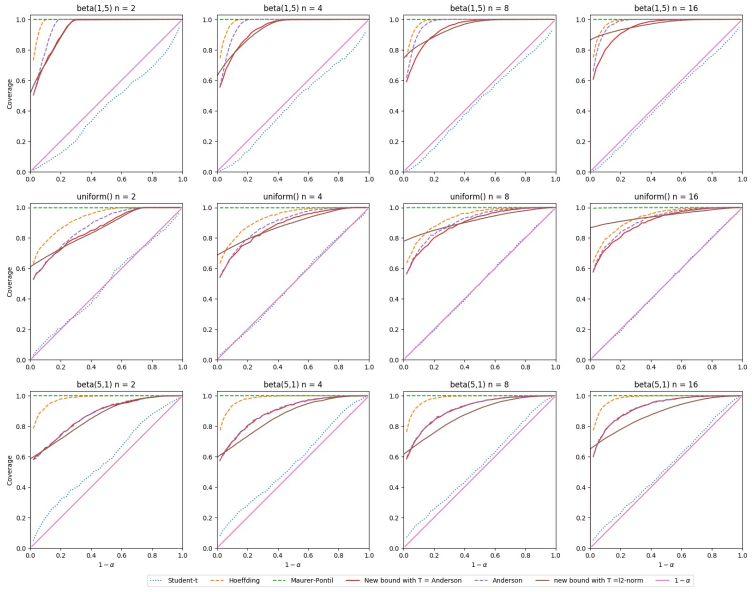
(a) The PDFs of the test distributions.



(b) Expected values of the bounds versus sample size.

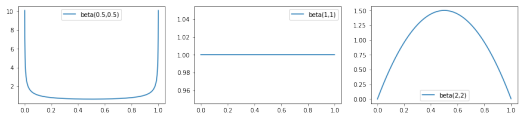


(c) The α-quantiles of bound distributions. If the α-quantile is below the true mean, the bound does not have guaranteed coverage.

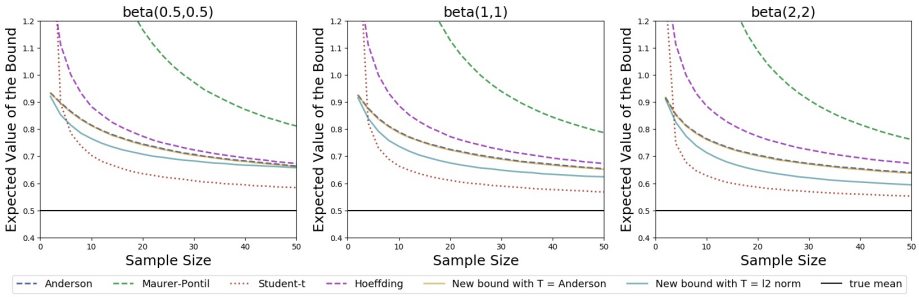


(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

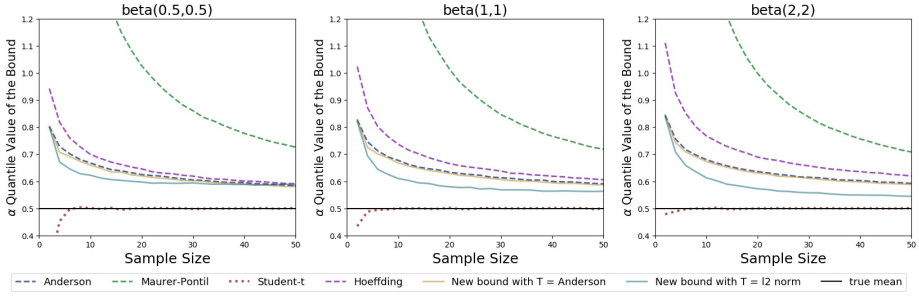
Figure C.1: Finding the upper bound of the mean with $D^+ = [0, 1]$



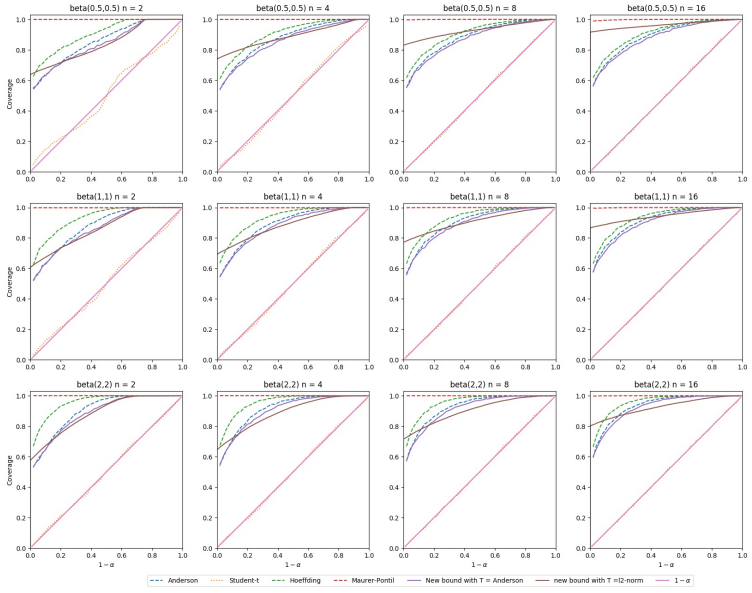
(a) The PDFs of the test distributions.



(b) Expected values of bounds versus sample size.

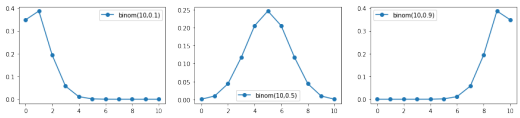


(c) The α-quantiles of bound distributions. If the α-quantile is below the true mean, the bound does not have guaranteed coverage.

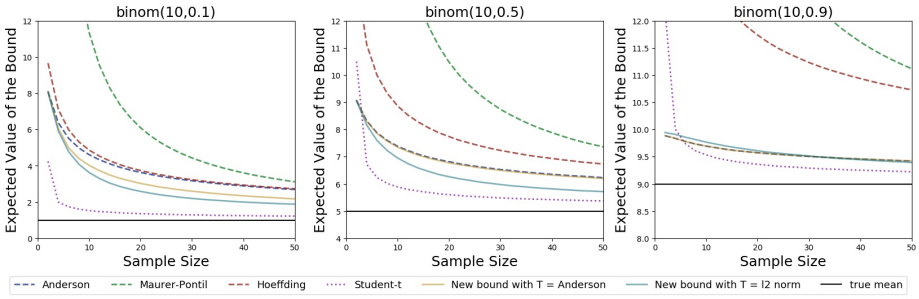


(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

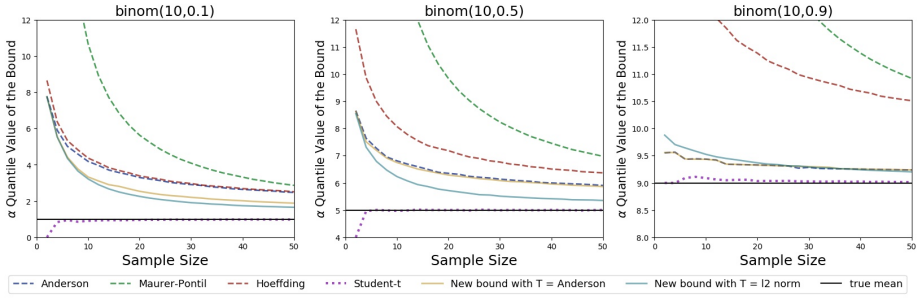
Figure C.2: Finding the upper bound of the mean with $D^+ = [0, 1]$



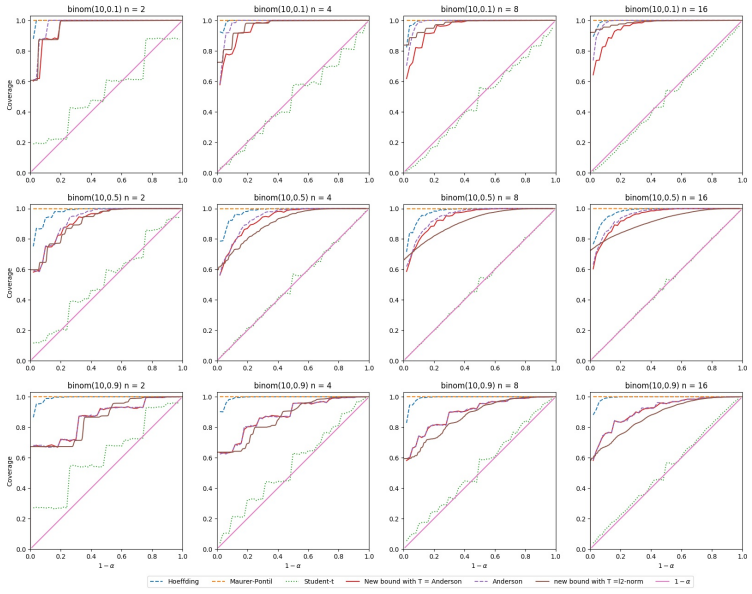
(a) The PMFs of the test distributions.



(b) Expected values of bounds versus sample size.

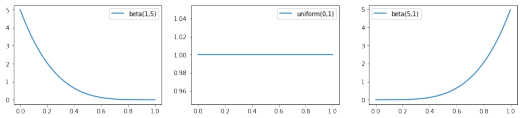


(c) The α-quantiles of bound distributions. If the α-quantile is below the true mean, the bound does not have guaranteed coverage.

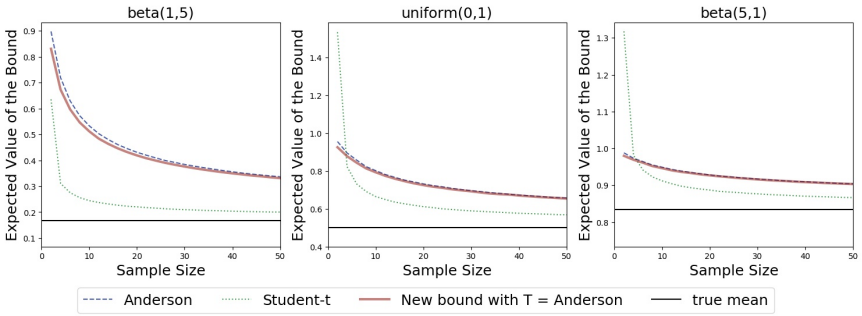


(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

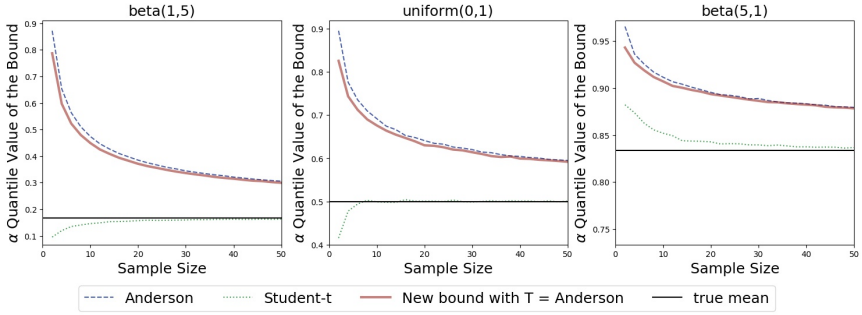
Figure C.3: Finding the upper bound of the mean with $D^+ = [0, 10]$



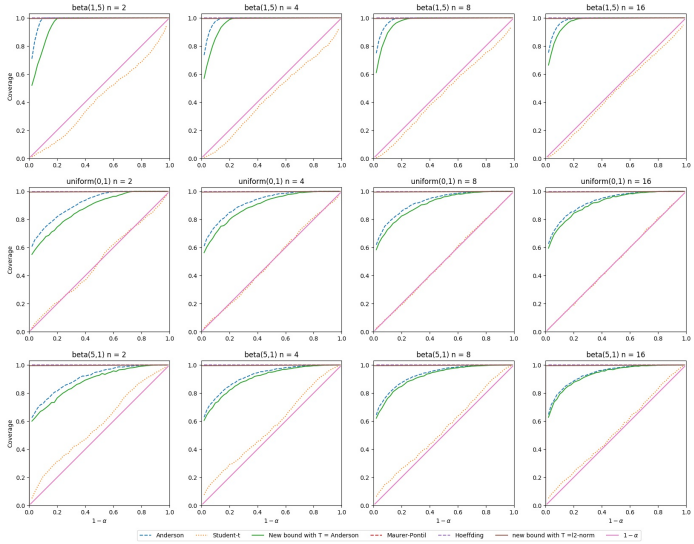
(a) The PDFs of the test distributions.



(b) The expected value of the bounds.

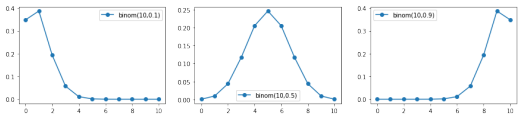


(c) The α-quantile of the bound distribution. If the α-quantile is below the true mean, the bound does not have guaranteed coverage.

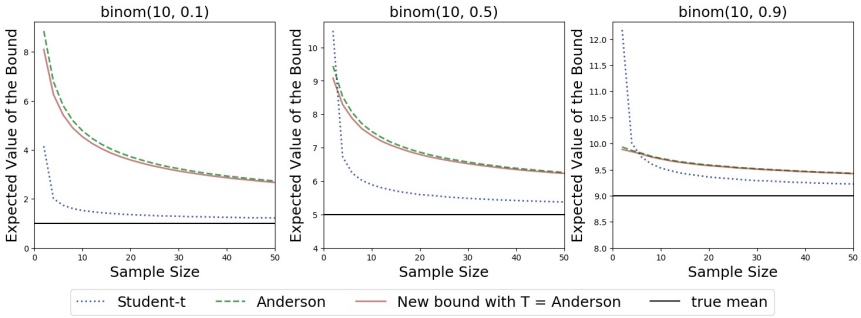


(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

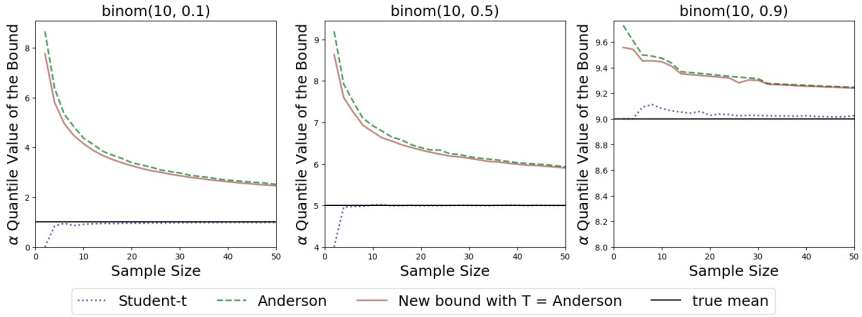
Figure C.4: Finding the upper bound of the mean with $D^+ = (-\infty, 1]$



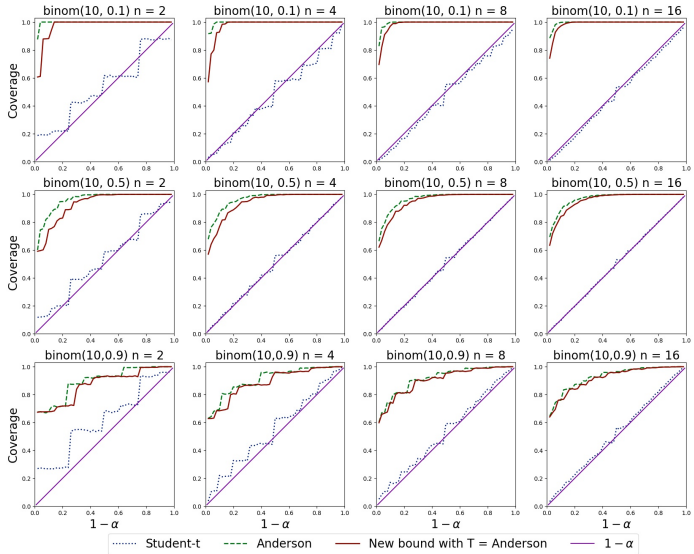
(a) The PMF of the test distributions.



(b) The expected value of the bounds.

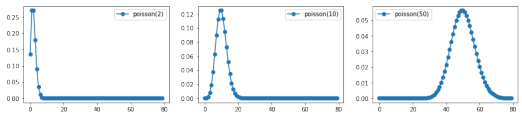


(c) The α-quantile of the bound distribution. If the α-quantile is below the true mean, the bound does not have guaranteed coverage.

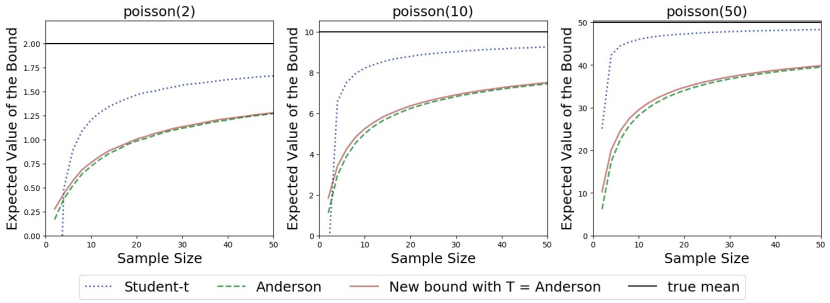


(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

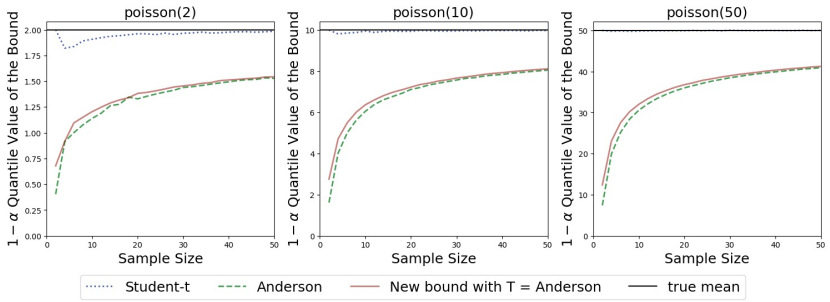
Figure C.5: Finding the upper bound of the mean with $D^+ = (-\infty, 10]$



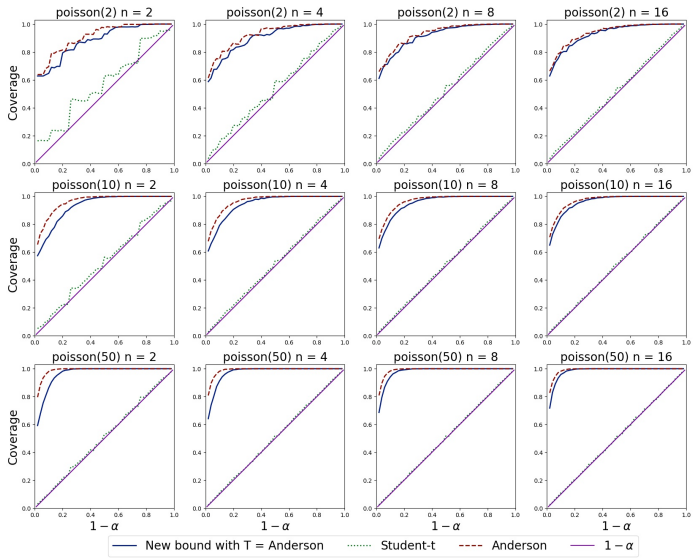
(a) The PMF of the test distributions.



(b) The expected value of the bounds.



(c) The $1-\alpha$-quantile of the bound distribution. If the $1-\alpha$-quantile is above the true mean, the bound does not have guaranteed coverage.



(d) The coverage of the bound. If the coverage is below the line $1-\alpha$, the bound does not have guaranteed coverage.

Figure C.6: Finding the lower bound of the mean with $D^+ = [0, \infty)$

C.2 Discussion on Section 3.1: Skewed Sample Mean Distribution with $n = 80$

In this section, as noted in Section 3.1, we show a log-normal distribution where the sample mean distribution is visibly skewed when $n = 80$ (Figure C.7). Student's t is not a good candidate in this case because the sample mean distribution is not approximately normal. This example is a variation on the one provided by Frost (2021).

While the log-normal distribution is an extreme example of skew, this example illustrates the danger of assuming the validity of arbitrary thresholds on the sample size, such as the traditional threshold of $n = 30$, for using the Student's t method. Clearly there are cases where such a threshold, and even much larger thresholds, are not adequate.

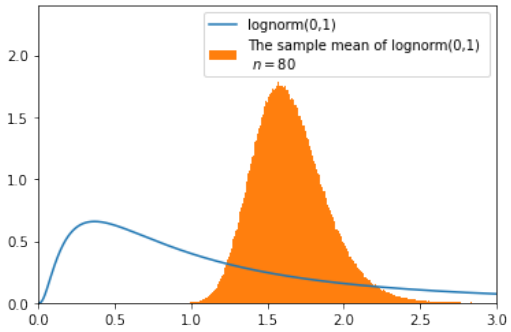


Figure C.7: The PDFs of $lognorm(0, 1)$ and the sample mean distribution of $lognorm(0, 1)$. The sample mean distribution of $lognorm(0, 1)$ is visibly skewed when the sample size $n = 80$.

C.3 Proof of Section 3.2.1.2

Proof of Lemma 9. Let $F^{-1}(y) = \inf\{x : F(x) \geq y\}$ for $0 < y < 1$ and U be an uniform random variable on $(0, 1)$. Since F is non-decreasing and right-continuous, $F(F^{-1}(y)) \geq y$. By Angus (1994), $F^{-1}(U)$ has CDF F. For $0 < y < 1$, then:

$$\mathbb{P}(Y \leq y) = \mathbb{P}(F(X) \leq y) \tag{C.3}$$

$$= \mathbb{P}(F(F^{-1}(U)) \leq y) \tag{C.4}$$

$$\leq \mathbb{P}(U \leq y) \tag{C.5}$$

$$= y. \tag{C.6}$$

If F is continuous, Angus (1994) shows that Y is uniformly distributed on $(0, 1)$. □

Proof of Lemma 10. Let $x_{(0)} \stackrel{\text{def}}{=} -\infty$ and $x_{(n+1)} \stackrel{\text{def}}{=} s_D$. Then:

$$\mu_F = \int x \, dF(x) \tag{C.7}$$

$$= \sum_{i=1}^{n+1} \int_{x_{(i-1)}}^{x_{(i)}} x \, dF(x) \tag{C.8}$$

$$\leq \sum_{i=1}^{n+1} \int_{x_{(i-1)}}^{x_{(i)}} x_{(i)} \, dF(x) \tag{C.9}$$

$$= \sum_{i=1}^{n+1} x_{(i)}(F(x_{(i)}) - F(x_{(i-1)})) \tag{C.10}$$

$$= s_D - \sum_{i=1}^{n} F(x_{(i)})(x_{(i+1)} - x_{(i)}) \tag{C.11}$$

$$\leq s_D - \sum_{i=1}^{n} G(x_{(i)})(x_{(i+1)} - x_{(i)}) \tag{C.12}$$

$$= \mu_G. \tag{C.13}$$

□

Proof of Lemma 11. Let $\cup$ denote the union of events and $\{\}$ denote an event. Let $\mathbf{Z}$ be a sample from F. Then for any sample $\mathbf{x}$:

$$\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) = \mathbb{P}_{\mathbf{Z}}(\mathbf{Z} \in \mathbb{S}(\mathbf{x})) \tag{C.14}$$

$$= \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{\mathbf{Z} = \mathbf{y}\}) \tag{C.15}$$

$$\leq \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\mathbf{Z} \preceq \mathbf{y}) \text{ because } \mathbf{Z} = \mathbf{y} \text{ implies } \mathbf{Z} \preceq \mathbf{y} \tag{C.16}$$

$$\leq \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{F(\mathbf{Z}) \preceq F(\mathbf{y})\}) \tag{C.17}$$

because F is non-decreasing, so $Z_{(i)} \leq y_{(i)}$ implies $F(Z_{(i)}) \leq F(y_{(i)})$. Let $U_1, \cdots, U_n$ be n samples from the uniform distribution on $(0, 1)$. From Lemma 9, for any $u \in (0, 1)$, $\mathbb{P}(F(Z_i) \leq u) \leq \mathbb{P}(U_i \leq u)$. Therefore

$$\mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{F(\mathbf{Z}) \preceq F(\mathbf{y})\}) \leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{\mathbf{U} \preceq F(\mathbf{y})\}). \tag{C.18}$$

Recall that $m_D(\mathbf{y}, \mathbf{U}) = s_D - \sum_{i=1}^{n} U_{(i)}(y_{(i+1)} - y_{(i)})$ where $\forall i, y_{(i+1)} - y_{(i)} \geq 0$. Therefore if $\forall i, U_{(i)} \leq F(y_{(i)})$ then $m_D(\mathbf{y}, \mathbf{U}) \geq m_D(\mathbf{y}, F(\mathbf{y}))$:

$$\mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{\mathbf{U} \preceq F(\mathbf{y})\}) \tag{C.19}$$

$$\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{m_D(\mathbf{y}, \mathbf{U}) \geq m_D(\mathbf{y}, F(\mathbf{y}))\}), \text{by Lemma 10} \tag{C.20}$$

$$\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{m_D(\mathbf{y}, \mathbf{U}) \geq \mu\}), \text{by Lemma 10} \tag{C.21}$$

$$\leq \mathbb{P}_{\mathbf{U}}(\sup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})} m_D(\mathbf{y}, \mathbf{U}) \geq \mu) \tag{C.22}$$

$$= \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu). \tag{C.23}$$

The inequality in Eq. C.22 is because if there exists $\mathbf{y} \in \mathbb{S}(\mathbf{x})$ such that $m_D(\mathbf{y}, \mathbf{U}) \geq \mu$, then $\sup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})} m_D(\mathbf{y}, \mathbf{U}) \geq \mu$. Therefore the event $\cup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})}\{m_D(\mathbf{y}, \mathbf{U}) \geq \mu\}$ is a subset of the event $\sup_{\mathbf{y}\in\mathbb{S}(\mathbf{x})} m_D(\mathbf{y}, \mathbf{U}) \geq \mu$, and Eq. C.22 follows.

From Eqs. C.17, C.18 and Eq. C.23:

$$\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) \quad \leq \quad \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu). \tag{C.24}$$

□

C.4 Discussion on Section 3.3: Monte Carlo Convergence

In Section 3.3, we discussed the use of Monte Carlo sampling of the induced mean function $b(\mathbf{x}, \mathbf{U})$ via sampling of the uniform random variable U, to approximate the $1-\alpha$ quantile of $b(\mathbf{x}, \mathbf{U})$. Let $\hat{q}_\ell$ denote the output of the Monte Carlo algorithm (Algorithm 1) using ℓ Monte Carlo samples. In this section we show that our estimator converges to the true quantile as the number of Monte Carlo samples grows, and, given a desired threshold ϵ, we can compute an upper bound at most $Q(1-\alpha, b(\mathbf{x}, \mathbf{U})) + \epsilon$ with guaranteed coverage.

Theorem 12. *Let $\epsilon > 0$.*

If $\sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)} = s_D$, then we can output s_D as an estimate and with probability at least $1-\alpha$:

$$\mu \leq s_D \leq Q(1-\alpha, b(\mathbf{x}, \mathbf{U})) + \epsilon. \tag{C.25}$$

If $\sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)} < s_D$, let $\gamma = \left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x}))} z_{(1)})}\right)^n$. Suppose $\gamma < \alpha$. Use $\ell = \left\lceil \frac{-\ln(\gamma/2)}{2} \left(\frac{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})}{\epsilon}\right)^n \right\rceil$ Monte Carlo samples to compute $Q(1-\alpha+\gamma, b(\mathbf{x}, \mathbf{U}))$ using Algorithm 1. Let $\hat{q}_\ell$ be the output of the algorithm. We output $\hat{q}_\ell + \epsilon/3$ as the final estimator. Then with probability at least $1-\alpha$:

$$\mu \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha, b(\mathbf{x}, \mathbf{U})) + \epsilon. \tag{C.26}$$

To prove Theorem 12, we first show some lemmas.

The Monte Carlo approximation error is quantified in the following lemma due to Serfling (1980). Let $F(m-) \stackrel{\text{def}}{=} \lim_{x\to m^-} F(x)$.

Lemma 38 (Theorem 2.3.2 in Serfling (1980))**.** *Let $0 < p < 1$. If $Q(p, M)$ is the unique solution m of $F(m-) \leq p \leq F(m)$, then for every $\epsilon > 0$,*

$$\mathbb{P}(|\mathbf{M}_{\lceil pl \rceil} - Q(p, M)| > \epsilon) \leq 2e^{-2l\delta}, \tag{C.27}$$

where $\mathbf{M}_k$ denotes the k-th order statistic of the sample $\mathbf{M}$ and

$$\delta = \min\left(p - F(Q(p, M) - \epsilon), F(Q(p, M) + \epsilon) - p\right).$$

Note that when the condition that $Q(p, M)$ is the unique solution m of $F(m-) \leq p \leq F(m)$ is satisfied, $\delta > 0$. Let $M \stackrel{\text{def}}{=} b_{D,T}(\mathbf{x}, \mathbf{U}) \in [0, 1]$. In Lemma 39 we will show that the CDF of M satisfies the condition in Lemma 38. Therefore the error incurred by computing the bound via Monte Carlo sampling can be decreased to an arbitrarily small value by choosing a large enough number of Monte Carlo samples l. The Monte Carlo estimation of $b^{\alpha}_{D^+,T}(\mathbf{x})$ where $D^+ = [0, 1]$ is presented in Algorithm 1.

We will show that for any $\mathbf{x}$, for any T, for any $p \in (0, 1)$, $F_M(m-) \leq p \leq F_M(m)$ has a unique solution by showing that for any $\mathbf{x}$ and T, F_M is strictly increasing on its support. To do so, for any c_1, c_2 in the support such that $c_1 < c_2$ we will show that

$$F_M(c_2) - F_M(c_1) > 0. \tag{C.28}$$

Lemma 39. *Let $M \stackrel{\text{def}}{=} b(\mathbf{x}, \mathbf{U})$. Let F_M be the CDF of M.*

For any $\mathbf{x}$, for any scalar function T,

- *If $\sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)} = s_D$, then for any $\epsilon > 0$, $M \in [s_D - \epsilon, s_D]$, or*
- *If $\sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)} < s_D$, then for any c_1, c_2 such that $0 \leq c_1 < c_2 \leq 1$,*

$$F_M(c_2) - F_M(c_1) \geq \left(\frac{c_2 - c_1}{s_D - \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)}}\right)^n > 0. \tag{C.29}$$

Proof. Recall the definition of the induced mean as

$$b(\mathbf{x}, \boldsymbol{\ell}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(\ell_{(i)} - \ell_{(i-1)}), \tag{C.30}$$

$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} \ell_{(i)}(z_{(i+1)} - z_{(i)}), \tag{C.31}$$

where $\ell_{(0)} \stackrel{\text{def}}{=} 0$, $\ell_{(n+1)} \stackrel{\text{def}}{=} 1$ and $z_{(n+1)} \stackrel{\text{def}}{=} s_D$.

We now find the support of M. Let $\phi \stackrel{\text{def}}{=} \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)}$. We will show that for any $\mathbf{u}$ where $0 \leq u_i \leq 1$, we have $\phi \leq b(\mathbf{x}, \mathbf{u}) \leq s_D$, and therefore the support of M is a subset of $[\phi, s_D]$. We have

$$b(\mathbf{x}, \mathbf{u}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} u_{(i)}(z_{(i+1)} - z_{(i)}) \tag{C.32}$$

$$\leq \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} 0(z_{(i+1)} - z_{(i)}) \tag{C.33}$$

$$= s_D. \tag{C.34}$$

Similarly we have

$$b(\mathbf{x}, \mathbf{u}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} u_{(i)}(z_{(i+1)} - z_{(i)}) \tag{C.35}$$

$$\geq \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} 1(z_{(i+1)} - z_{(i)}) \tag{C.36}$$

$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - (z_{(n+1)} - z_{(1)}) \tag{C.37}$$

$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)} \tag{C.38}$$

$$= \phi. \tag{C.39}$$

Therefore $M = b(\mathbf{x}, \mathbf{U}) \in [\phi, s_D]$. We consider two cases: where $\phi = s_D$ and where $\phi < s_D$.

Case 1: $\phi = s_D$.

We have $\sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)} = s_D$. Then for any $\epsilon > 0$, there exists $\mathbf{z}^*$ such that $z^*_{(1)} \geq s_D - \epsilon$, and therefore $z^*(i) \geq s_D - \epsilon$ for all $i, 1 \leq i \leq n$. Therefore:

$$b(\mathbf{x}, \boldsymbol{\ell}) = \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(\ell_{(i)} - \ell_{(i-1)}) \tag{C.40}$$

$$\geq \sum_{i=1}^{n+1} (s_D - \epsilon)(\ell_{(i)} - \ell_{(i-1)}) \tag{C.41}$$

$$= s_D - \epsilon. \tag{C.42}$$

Therefore $s_D - \epsilon \leq M = b(\mathbf{x}, \mathbf{U}) \leq s_D$, and the $1-\alpha$ quantile of M is in $[s_D - \epsilon, s_D]$.

Case 2: $\phi < s_D$.

Let $c_1, c_2 \in \mathcal{R}$ be such that $\phi \leq c_1 < c_2 \leq s_D$. We will now show that

$$F_M(c_2) - F_M(c_1) > 0. \tag{C.43}$$

Let $v \stackrel{\text{def}}{=} \frac{s_D - c_2}{s_D - \phi}$ and $w \stackrel{\text{def}}{=} \frac{s_D - c_1}{s_D - \phi}$. If $\phi \leq c_1 < c_2 \leq s_D$ then $v < w$ and $v, w \in [0, 1]$. Let $\mathbf{v} \stackrel{\text{def}}{=} (v_1, \cdots, v_n)$ and $\mathbf{w} \stackrel{\text{def}}{=} (w_1, \cdots, w_n)$ where $\forall i, v_i = v$ and $w_i = w$. Then

$$b(\mathbf{x}, \mathbf{v}) = \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(v_{(i)} - v_{(i-1)}) \tag{C.44}$$

$$= \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(n+1)}(v_{(n+1)} - v_{(n)}) + z_{(1)}(v_{(1)} - v_{(0)}) \tag{C.45}$$

$$= \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} s_D(1 - v) + z_{(1)}(v - 0) \tag{C.46}$$

$$= \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} s_D - (s_D - z_{(1)})\frac{s_D - c_2}{s_D - \phi} \tag{C.47}$$

$$= s_D - (s_D - \phi)\frac{s_D - c_2}{s_D - \phi} \text{ because } \frac{s_D - c_2}{s_D - \phi} \geq 0 \tag{C.48}$$

$$= c_2. \tag{C.49}$$

Similarly,

$$b(\mathbf{x}, \mathbf{w}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(w_{(i)} - w_{(i-1)}) \tag{C.50}$$

$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(n+1)}(w_{(n+1)} - w_{(n)}) + z_{(1)}(w_{(1)} - w_{(0)}) \tag{C.51}$$

$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D(1 - w) + z_{(1)}(w - 0) \tag{C.52}$$

$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - (s_D - z_{(1)})\frac{s_D - c_1}{s_D - \phi} \tag{C.53}$$

$$= s_D - (s_D - \phi)\frac{s_D - c_1}{s_D - \phi} \text{ because } \frac{s_D - c_1}{s_D - \phi} \geq 0 \tag{C.54}$$

$$= c_1. \tag{C.55}$$

Since $b(\mathbf{x}, \mathbf{u})$ is constructed from a linear function of $\mathbf{u}$ with non-positive coefficients, for any $\mathbf{u}$ such that $v \leq u_{(1)} \leq \cdots \leq u_{(n)} < w$ we have:

$$b(\mathbf{x}, \mathbf{w}) < b(\mathbf{x}, \mathbf{u}) \leq b(\mathbf{x}, \mathbf{v}), \tag{C.56}$$

which is equivalent to:

$$c_1 < b(\mathbf{x}, \mathbf{u}) \leq c_2. \tag{C.57}$$

So we have $v \leq u_{(1)} \leq \cdots \leq u_{(n)} < w$ implies $c_1 < b(\mathbf{x}, \mathbf{u}) \leq c_2$. Therefore for any c_1, c_2 such that $\phi \leq c_1 < c_2 \leq s_D$:

$$F_M(c_2) - F_M(c_1) = \mathbb{P}(c_1 < M \leq c_2) \tag{C.58}$$

$$= \mathbb{P}_{\mathbf{U}}(c_1 < b(\mathbf{x}, \mathbf{U}) \leq c_2) \tag{C.59}$$

$$\geq \mathbb{P}_{\mathbf{U}}(v \leq U_{(1)} \leq \cdots \leq U_{(n)} < w) \tag{C.60}$$

$$= \mathbb{P}_{\mathbf{U}}(\ \forall i, 1 \leq i \leq n : v \leq U_i < w) \tag{C.61}$$

$$= (w - v)^n \tag{C.62}$$

$$= \left(\frac{c_2 - c_1}{s_D - \phi}\right)^n \tag{C.63}$$

$$> 0 \text{ because } c_1 < c_2. \tag{C.64}$$

Since the support of M is in $[\phi, s_D]$ we have that F_M is strictly increasing on the support. □

In summary, the Monte Carlo estimate of our bound will converge to the correct value as the number of samples grows.

Now we prove Theorem 12.

Proof of Theorem 12. To simplify the notation, we use $Q(\alpha)$ to denote $Q(\alpha, M)$. From Lemma 39, consider 2 cases:

- If $\sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)} = s_D$, then for any $\epsilon > 0$, $s_D - \epsilon \leq M \leq s_D$. So $s_D - \epsilon \leq Q(1-\alpha, M) \leq s_D$ which is equivalent to $Q(1-\alpha, M) \leq s_D \leq Q(1-\alpha, M) + \epsilon$. Since with probability at least $1-\alpha$, $Q(1-\alpha, M) \geq \mu$, we have with probability at least $1-\alpha$, $s_D \geq \mu$. Therefore with probability at least $1-\alpha$,

$$\mu \leq s_D \leq Q(1-\alpha, M) + \epsilon. \tag{C.65}$$

- If $\sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)} < s_D$: since F_M is strictly increasing on the support, for γ such that $0 < \gamma < \alpha$, $Q(1-\alpha) < Q(1-\alpha+\gamma)$ and:

$$\gamma = F(Q(1-\alpha), Q(1-\alpha+\gamma)) \geq \left(\frac{Q(1-\alpha+\gamma) - Q(1-\alpha)}{s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)}} \right)^n. \quad \text{(C.66)}$$

Therefore, letting $\gamma = \left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x}))} z_{(1)})} \right)^n$ we have that

$$Q(1-\alpha+\gamma) \leq Q(1-\alpha) + \gamma^{1/n}(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)}) \quad \text{(C.67)}$$

$$\leq Q(1-\alpha) + \epsilon/3. \quad \text{(C.68)}$$

Let $p \stackrel{\text{def}}{=} 1-\alpha+\gamma$. From Lemma 38 and Lemma 39,

$$\mathbb{P}(|\hat{q}_\ell - Q(p)| > \epsilon/3) \leq 2e^{-2l\delta}, \quad \text{(C.69)}$$

where

$$\delta = \min\left(p - F(Q(p) - \epsilon/3), F(Q(p) + \epsilon/3) - p\right) \quad \text{(C.70)}$$

$$= \min\left(F(Q(p) - \epsilon/3, Q(p)), F(Q(p), Q(p) + \epsilon/3)\right) \quad \text{(C.71)}$$

$$\geq \left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})} \right)^n. \quad \text{(C.72)}$$

Therefore letting $\ell = \left\lceil \frac{-\ln(\gamma/2)}{2} \left(\frac{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})}{\epsilon} \right)^n \right\rceil$,

$$\mathbb{P}(|\hat{q}_\ell - Q(1-\alpha+\gamma)| > \epsilon/3) \leq 2e^{-2l\left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})} \right)^n} \quad \text{(C.73)}$$

$$\leq \gamma. \quad \text{(C.74)}$$

Since $\mathbb{P}(Q(1-\alpha+\gamma) < \mu) \leq \alpha - \gamma$, using the union bound we have

$$\mathbb{P}(|\hat{q}_\ell - Q(1-\alpha+\gamma)| > \epsilon/3 \text{ OR } Q(1-\alpha+\gamma) < \mu) \leq \gamma + \alpha - \gamma \quad \text{(C.75)}$$

$$= \alpha. \quad \text{(C.76)}$$

And therefore,

$$1-\alpha \leq \mathbb{P}(|\hat{q}_\ell - Q(1-\alpha+\gamma)| \leq \epsilon/3 \text{ AND } Q(1-\alpha+\gamma) \geq \mu) \tag{C.77}$$

$$\leq \mathbb{P}(Q(1-\alpha+\gamma) \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha+\gamma) + 2\epsilon/3 \text{ AND } Q(1-\alpha+\gamma) \geq \mu) \tag{C.78}$$

$$\leq \mathbb{P}(\mu \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha+\gamma) + 2\epsilon/3) \tag{C.79}$$

$$\leq \mathbb{P}(\mu \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha) + \epsilon) \text{ from Eq. C.68.} \tag{C.80}$$

□

C.5 Discussion on Section 3.4

We discuss the case when the distribution is Bernoulli in Section C.5.1, and present the proofs of Section 3.4.2 in Section C.5.2. In Section C.5.3 we show that our bound is equal to Anderson's when T is Anderson's bound and the lower bound of the support is $-\infty$, and could be better than Anderson's when T is Anderson's bound and the lower bound of the support is finite and tight.

C.5.1 Special Case: Bernoulli Distribution

When we know that $D = \{0, 1\}$, the distribution is Bernoulli. If we choose T to be the sample mean, we will show that our bound becomes the same as the Clopper-Pearson confidence bound for binomial distributions (Clopper and Pearson, 1934).

If $\mathbf{x}, \mathbf{z} \in \{0, 1\}^n$ and $T(\mathbf{z}) \leq T(\mathbf{x})$ then $m(\mathbf{z}, \mathbf{u}) \leq m(\mathbf{x}, \mathbf{u})$. Therefore for any $\mathbf{u} \in [0, 1]^n$,

$$b_{D,T}(\mathbf{x}, \mathbf{u}) = \sup_{\mathbf{z} \in \{0,1\}^n : T(\mathbf{z}) \leq T(\mathbf{x})} m_D(\mathbf{z}, \mathbf{u}) = m_D(\mathbf{x}, \mathbf{u}). \tag{C.81}$$

Let $p_{\mathbf{x}}$ be the number of 0's in $\mathbf{x}$. Therefore the bound becomes the $1-\alpha$ quantile of $m_D(\mathbf{x}, \mathbf{U})$ where

$$m_D(\mathbf{x}, \mathbf{U}) = 1 - \sum_{i=1}^{n} U_{(i)}(x_{(i+1)} - x_{(i)}) = 1 - U_{(p_{\mathbf{x}})}. \tag{C.82}$$

Therefore the bound is the $1-\alpha$ quantile of $1 - U_{(p_{\mathbf{x}})}$. Then

$$\mathbb{P}(U_{(p_{\mathbf{x}})} \leq 1 - b^{\alpha}(\mathbf{x})) = \mathbb{P}(1 - U_{(p_{\mathbf{x}})} \geq b^{\alpha}(\mathbf{x})) = \alpha. \tag{C.83}$$

Let $\beta(i, j)$ denote a beta distribution with parameters i and j. We use the fact that the order statistics of a uniform distribution are beta-distributed. Since $U_{(p_{\mathbf{x}})} \sim \beta(p_{\mathbf{x}}, n + 1 - p_{\mathbf{x}})$, we have $1 - U_{(p_{\mathbf{x}})} \sim \beta(n - p_{\mathbf{x}} + 1, p_{\mathbf{x}})$

$$b^{\alpha}(\mathbf{x}) = Q(1 - \alpha, \beta(n - p_{\mathbf{x}} + 1, p_{\mathbf{x}})). \tag{C.84}$$

This is the same as the Clopper-Pearson upper confidence bound for binomial distributions.

C.5.2 Proof of Section 3.4.2

Proof of Lemma 14. [1] If $\forall y \in \mathcal{R}, F(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)$ then

$$\forall i : 1 \leq i \leq n, F(X_{(i)}) \geq \ell_{(i)}. \tag{C.85}$$

Recall that $m_D(\mathbf{X}, \boldsymbol{\ell}) = s_D - \sum_{i=1}^{n} \ell_{(i)}(z_{(i+1)} - z_{(i)})$. Therefore if $\forall i : 1 \leq i \leq n,\ F(X_{(i)}) \geq \ell_{(i)}$ then $m(\mathbf{X}, \boldsymbol{\ell}) \geq m(\mathbf{X}, F(\mathbf{X}))$.

[1] The proof is implied in Anderson (1969b) but we provide it here for completeness

From Lemma 10, $m(\mathbf{X}, F(\mathbf{X})) \geq \mu$. Therefore $m(\mathbf{X}, \boldsymbol{\ell}) \geq \mu$. And hence, finally,

$$\mathbb{P}(m(\mathbf{X}, \boldsymbol{\ell}) \geq \mu) \geq \mathbb{P}_{\mathbf{X}} \left(\forall y \in \mathcal{R}, F(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)\right) \tag{C.86}$$

$$= 1 - \alpha. \tag{C.87}$$

□

We now show that if $G_{\mathbf{X},\boldsymbol{\ell}}$ (Figure 3.1a) is a lower confidence bound, then the order statistics of $\boldsymbol{\ell}$ are element-wise smaller than the order statistics of a sample of size n from the uniform distribution with high probability:

Lemma 40. *Let $\mathbf{U} = U_1, ..., U_n$ be a sample of size n from the continuous uniform distribution on $[0, 1]$. Let $\boldsymbol{\ell} \in [0, 1]^n$ and $\alpha \in (0, 1)$. If $\mathcal{D}^+$ is continous and $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $(1 - \alpha)$ lower confidence bound for the CDF then:*

$$\mathbb{P}_{\mathbf{U}}(\ \forall i : 1 \leq i \leq n, U_{(i)} \geq \ell_{(i)}) \geq 1 - \alpha. \tag{C.88}$$

Proof. Let K be the CDF of a distribution such that K is continuous and strictly increasing on $\mathcal{D}^+$ (since $\mathcal{D}^+$ is continuous, K exists). Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a sample of size n from the distribution with CDF K. By Lemma 9, $K(X)$ is uniformly distributed on $[0, 1]$.

By the definition of $G_{\mathbf{X},\boldsymbol{\ell}}$, if $\ \forall x \in C, K(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)$ then:

$$K(y) \geq 0, \qquad \text{if } y < X_{(1)} \tag{C.89}$$

$$K(y) \geq \ell_{(i)}, \qquad \text{if } X_{(i)} \leq y < X_{(i+1)} \tag{C.90}$$

$$K(y) \geq 1, \qquad \text{if } y \geq s_{D^+}. \tag{C.91}$$

which is equivalent to:

$$\forall i : 1 \leq i \leq n, K(y) \geq \ell_{(i)}, \text{ if } X_{(i)} \leq y < X_{(i+1)}. \tag{C.92}$$

Since $K(y)$ is non-decreasing, this is equivalent to:

$$\forall i : 1 \leq i \leq n, K(X_{(i)}) \geq \ell_{(i)}. \tag{C.93}$$

Since $G_{\mathbf{X},\boldsymbol{\ell}}$ is a lower confidence bound,

$$1 - \alpha \leq \mathbb{P}_{\mathbf{X}}(\ \forall y \in \mathcal{R}, K(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)) \tag{C.94}$$

$$= \mathbb{P}_{\mathbf{X}}(\forall i : 1 \leq i \leq n, K(X_{(i)}) \geq \ell_{(i)}) \tag{C.95}$$

$$= \mathbb{P}_{\mathbf{U}}(\forall i : 1 \leq i \leq n, U_{(i)} \geq \ell_{(i)}) \text{ by Lemma 9.} \tag{C.96}$$

□

To prove Theorem 9, we prove the more general version where $G_{\mathbf{X},\boldsymbol{\ell}}$ is a (possibly not exact) lower confidence bound for the CDF.

Theorem 13. *Let $\boldsymbol{\ell} \in [0,1]^n$. Let $D^+ = [-\infty, b]$. If $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $1-\alpha$ lower confidence bound for the CDF, then for any sample size n, for all sample values $\mathbf{x} \in D^n$ and all $\alpha \in (0,1)$, using $T(\mathbf{x}) = m_{D^+}(\mathbf{x}, \boldsymbol{\ell})$ to compute $b^{\alpha}_{D^+,T}(\mathbf{x})$ yields:*

$$b^{\alpha}_{D^+,T}(\mathbf{x}) \leq m_{D^+}(\mathbf{x}, \boldsymbol{\ell}). \tag{C.97}$$

Proof. Since $G_{\mathbf{X},\boldsymbol{\ell}}$ is a lower confidence bound for the CDF F, from Lemma 40,

$$\mathbb{P}(\forall i, U_{(i)} \geq \ell_{(i)}) \geq 1 - \alpha. \tag{C.98}$$

First we note that

$$b_{D^+,T}(\mathbf{x}, \boldsymbol{\ell}) = \sup_{\mathbf{y}:\mathbf{y}\in\mathbb{S}_{D^+,T}(\mathbf{x})} m_{D^+}(\mathbf{y}, \boldsymbol{\ell}) \tag{C.99}$$

$$= \sup_{m_{D^+}(\mathbf{y},\boldsymbol{\ell}) \leq m_{D^+}(\mathbf{x},\boldsymbol{\ell})} m_{D^+}(\mathbf{y}, \boldsymbol{\ell}) \tag{C.100}$$

$$= m_{D^+}(\mathbf{x}, \boldsymbol{\ell}). \tag{C.101}$$

Recall that $b^{\alpha}_{D^+,T}(\mathbf{x})$ is the $1-\alpha$ quantile of $b_{D^+,T}(\mathbf{x},\mathbf{U})$. In order to show that $b^{\alpha}_{D^+,T}(\mathbf{x}) \leq b_{D^+,T}(\mathbf{x},\boldsymbol{\ell})$, we will show that

$$\mathbb{P}(b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq b_{D^+,T}(\mathbf{x},\boldsymbol{\ell})) \geq 1-\alpha. \tag{C.102}$$

Recall that $b_{D^+,T}(\mathbf{x},\mathbf{U}) = \sup_{\mathbf{y}\in\mathbb{S}_T(\mathbf{x})} s_{D^+} - \sum_{i=1}^{n} U_{(i)}(y_{(i+1)} - y_{(i)})$. Then if $\forall i, U_{(i)} \geq \ell_{(i)}$ then $b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq b_{D^+,T}(\mathbf{x},\boldsymbol{\ell})$. Therefore,

$$\begin{aligned}\mathbb{P}(b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq b_{D^+,T}(\mathbf{x},\boldsymbol{\ell})) &\geq \mathbb{P}(\forall i, U_{(i)} \geq \ell_{(i)}) &\text{(C.103)}\\ &\geq 1-\alpha, \text{ by Lemma 40.} &\text{(C.104)}\end{aligned}$$

□

We can now show the comparison with Anderson's bound and Hoeffding's bound.

Proof of Theorem 9. We have $b^{\alpha,\text{Anderson}}_{\boldsymbol{\ell}}(\mathbf{x}) = m_{D^+}(\mathbf{x},\boldsymbol{\ell})$ where $\boldsymbol{\ell}$ satisfies $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $1-\alpha$ lower confidence bound for the CDF. Therefore applying Theorem 13 yields the result. □

Proof of Theorem 10. The proof follows directly from Lemma 15 and Theorem 9.

Recall that $G_{\mathbf{X},\mathbf{u}^{\text{And}}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF and therefore:

$$\mathbb{P}_{\mathbf{U}}(\forall i : 1 \leq i \leq n, U_{(i)} \geq u^{\text{And}}_{(i)}) = 1-\alpha. \tag{C.105}$$

From Theorem 9, using $T(\mathbf{x}) = b^{\alpha,\text{Anderson}}_{\mathbf{u}^{\text{And}}}(\mathbf{x})$ yields

$$b^{\alpha}_{D^+,T}(\mathbf{x}) \leq b^{\alpha,\text{Anderson}}_{\mathbf{u}^{\text{And}}}(\mathbf{x}). \tag{C.106}$$

Let $\boldsymbol{\ell} \in [0,1]^n$ be defined such that

$$\ell_i \stackrel{\text{def}}{=} \max\left\{0, i/n - \sqrt{\ln(1/\alpha)/(2n)}\right\}. \tag{C.107}$$

Since $G_{\mathbf{X},\boldsymbol{\ell}}(x)$ is an $1-\alpha$ lower confidence bound, from Lemma 40:

$$\mathbb{P}_{\mathbf{U}}(\forall i : 1 \le i \le n, U_{(i)} \ge \ell_{(i)}) \ge 1-\alpha \tag{C.108}$$

$$= \mathbb{P}_{\mathbf{U}}(\forall i : 1 \le i \le n, U_{(i)} \ge u^{\text{And}}_{(i)}). \tag{C.109}$$

Since $u^{\text{And}}_i = \max\{0, i/n - \beta(n)\}$ and $\ell_i = \max\left\{0, i/n - \sqrt{\ln(1/\alpha)/(2n)}\right\}$, we have $\beta(n) \le \sqrt{\ln(1/\alpha)/(2n)}$, and therefore $u^{\text{And}}_i \ge \ell_i$ for all i. Therefore $m(\mathbf{x}, \boldsymbol{\ell}) \ge m(\mathbf{x}, \mathbf{u}^{\text{And}})$, i.e.

$$b^{\alpha,\text{Anderson}}_{\mathbf{u}^{\text{And}}}(\mathbf{x}) \le b^{\alpha,\text{Anderson}}_{\boldsymbol{\ell}}(\mathbf{x}). \tag{C.110}$$

From Eq. C.106, Eq. C.110 and Lemma 15 we have the result. □

C.5.3 Special Case: Reduction to Anderson's Bound

In this section we present a more detailed comparison to Anderson's. We show that our bound is equal to Anderson's when T is Anderson's bound and the lower bound of the support is $-\infty$, and can be better than Anderson's when T is Anderson's bound and the lower bound of the support is tight.

Lemma 41. *Let $\boldsymbol{\ell} \in [0,1]^n$ be such that $\ell_i \ge 0\ \forall i, 1 \le i \le n$ and $0 < \ell_{(n)} < 1$. Let $D^+ = [a,b]$. Let $i_0 \stackrel{\text{def}}{=} \min\{i : 1 \le i \le n, \ell_{(i)} > 0\}$. If $G_{\mathbf{X},\boldsymbol{\ell}}$ is an exact $1-\alpha$*

lower confidence bound for the CDF, then for any sample size n, for all sample values $\mathbf{x} \in D^n$ and all $\alpha \in (0,1)$, using $T(\mathbf{x}) = m_{D^+}(\mathbf{x}, \boldsymbol{\ell})$ to compute $b^{\alpha}_{D^+,T}(\mathbf{x})$ yields:

$$b^{\alpha}_{D^+,T}(\mathbf{x}) = m_{D^+}(\mathbf{x}, \boldsymbol{\ell}) \text{ if } a \leq b - \frac{b - m(\mathbf{x}, \boldsymbol{\ell})}{\ell_{(i_0)}}, \tag{C.111}$$

$$b^{\alpha}_{D^+,T}(\mathbf{x}) < m_{D^+}(\mathbf{x}, \boldsymbol{\ell}) \text{ if } a > b - \frac{b - m(\mathbf{x}, \boldsymbol{\ell})}{\ell_{(i_0)}} \text{ and } n > i_0. \tag{C.112}$$

In particular, if $x_i < b - \frac{(b-a)\ell_{(i_0)}}{\ell_{(n)}}$ $\forall i, 1 \leq i \leq n$ and $n > i_0$ then $b^{\alpha}_{D^+,T}(\mathbf{x}) < m_{D^+}(\mathbf{x}, \boldsymbol{\ell})$.

Proof. From the proof of Theorem 13, if $U_{(i)} \geq \ell_{(i)}$ for all i, then $b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})$ and therefore:

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) \geq \mathbb{P}\left(\cap_{i:1\leq i\leq n}\{U_{(i)} \geq \ell_{(i)}\}\right) \tag{C.113}$$

$$= 1 - \alpha. \tag{C.114}$$

We will show that:

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) \leq 1 - \alpha \text{ if } a \leq \frac{b\ell_{(i_0)} - b + m(\mathbf{x}, \boldsymbol{\ell})}{\ell_{(i_0)}} \tag{C.115}$$

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) > 1 - \alpha \text{ otherwise,} \tag{C.116}$$

which implies:

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) = 1 - \alpha \text{ if } a \leq \frac{b\ell_{(i_0)} - b + m(\mathbf{x}, \boldsymbol{\ell})}{\ell_{(i_0)}} \tag{C.117}$$

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) > 1 - \alpha \text{ otherwise,} \tag{C.118}$$

- First we will show that $\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) \leq 1-\alpha$ if $a \leq \frac{b\ell_{(i_0)} - b + m(\mathbf{x}, \boldsymbol{\ell})}{\ell_{(i_0)}}$.
  Recall that $b_{D^+,T}(\mathbf{x}, \mathbf{U}) = \sup_{\mathbf{y} \in \mathbb{S}_{D^+,T}(\mathbf{x})} m_{D^+,T}(\mathbf{x}, \mathbf{U})$. We have:

$$\mathbb{P}(b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})) = \mathbb{P}\left(\sup_{\mathbf{y}\in\mathbb{S}_{D^+,T}(\mathbf{x})} m_{D^+,T}(\mathbf{y},\mathbf{U}) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})\right). \tag{C.119}$$

Consider the set of points $\mathbf{v}^i$ of the form

$$\mathbf{v}^i = (\underbrace{\gamma_i,\cdots,\gamma_i}_{i \text{ times}}, b, \cdots, b), \tag{C.120}$$

where γ_i satisfy $a \leq \gamma_i \leq b$ and $\sum_{i=1}^{n+1} v_i(\ell_{(i)} - \ell_{(i-1)}) = m(\mathbf{x},\boldsymbol{\ell})$, which is equivalent to:

$$a \leq \gamma_i \leq b, \tag{C.121}$$

$$\ell_{(i)} > 0, \tag{C.122}$$

$$\gamma_i = b - \frac{b - m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i)}}. \tag{C.123}$$

Therefore if $a \leq b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i)}}$ for all i such that $\ell_{(i)} > 0$ then $\mathbf{v}^i \in \mathbb{S}_{D^+,T}(\mathbf{x})$ for all i and:

$$\mathbb{P}(b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})) \tag{C.124}$$

$$\leq \mathbb{P}\left(\cap_{i:\ell_{(i)}>0}\{m_{D^+,T}(\mathbf{v}^i,\mathbf{U}) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})\}\right) \tag{C.125}$$

$$= \mathbb{P}\left(\cap_{i:\ell_{(i)}>0}\{b - U_{(i)}(b-\gamma_i) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})\}\right) \tag{C.126}$$

$$= \mathbb{P}\left(\cap_{i:\ell_{(i)}>0}\{b - U_{(i)}(b-(b-\frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i)}})) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})\}\right) \tag{C.127}$$

$$= \mathbb{P}\left(\cap_{i:\ell_{(i)}>0}\{U_{(i)} \geq \ell_{(i)}\}\right) \tag{C.128}$$

$$= 1 - \alpha. \tag{C.129}$$

Since $\ell_{(1)} \leq \cdots \leq \ell_{(n)}$, if $a \leq b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i_0)}}$ then $a \leq b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i)}}$ for all i. Therefore if $a \leq b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i_0)}}$, then $\mathbb{P}(b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})) = 1-\alpha$ and $b^{\alpha}_{D^+,T}(\mathbf{x}) = m_{D^+}(\mathbf{x},\boldsymbol{\ell})$.

- Now we will show that if $a > b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i_0)}}$ and $n > i_0$ then $\mathbb{P}(b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq m_{D^+,T}(\mathbf{x},\boldsymbol{\ell})) > 1-\alpha$.

  Let $\epsilon = \min\left(\inf_{\mathbf{y}\in\mathbb{S}_{D^+,T}(\mathbf{x})} \frac{(b-y_{(i_0+1)})(1-\ell_{(n)})}{b-a}, \ell_{(i_0)}\right)$.

  We will show that if $a = b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i_0)}} + \delta$ where $\delta > 0$ then $\ell_{(n)} > \ell_{(i_0)}$ and $b \geq y_{(i_0+1)} + \frac{\delta\ell_{(i_0)}}{\ell_{(n)}-\ell_{(i_0)}}$ for all $\mathbf{y} \in \mathbb{S}_{D^+,T}(\mathbf{x})$ and therefore $\epsilon \geq \min\left(\frac{\delta\ell_{(i_0)}}{\ell_{(n)}-\ell_{(i_0)}}\frac{1-\ell_{(n)}}{b-a}, \ell_{(i_0)}\right) > 0$. Given $\mathbf{y}$, construct $\mathbf{y}'$ of the form:

$$\mathbf{y}' = (\underbrace{a,\cdots,a}_{i_0 \text{ times}}, \underbrace{y_{(i_0+1)},\cdots,y_{(i_0+1)}}_{n-i_0 \text{ times}}). \tag{C.130}$$

Since $\mathbf{y}'$ is component-wise smaller than $\mathbf{y}$ and $m(\mathbf{y},\boldsymbol{\ell}) \leq m(\mathbf{x},\boldsymbol{\ell})$ we have:

$$m(\mathbf{y}',\boldsymbol{\ell}) \leq m(\mathbf{y},\boldsymbol{\ell}) \leq m(\mathbf{x},\boldsymbol{\ell}). \tag{C.131}$$

And therefore:

$$m(\mathbf{y}',\boldsymbol{\ell}) \leq m(\mathbf{x},\boldsymbol{\ell}) \tag{C.132}$$

$$\iff b(1-\ell_{(n)}) + y_{(i_0+1)}(\ell_n - \ell_{(i_0)}) + (b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i_0)}} + \delta)\ell_{(i_0)} \leq m(\mathbf{x},\boldsymbol{\ell}) \tag{C.133}$$

$$\iff b(1-\ell_{(n)}) + y_{(i_0+1)}(\ell_n - \ell_{(i_0)}) + (b\ell_{(i_0)} + \delta\ell_{(i_0)} - b + m(\mathbf{x},\boldsymbol{\ell})) \leq m(\mathbf{x},\boldsymbol{\ell}) \tag{C.134}$$

$$\iff \delta\ell_{(i_0)} \leq (b - y_{(i_0+1)})(\ell_{(n)} - \ell_{(i_0)}). \tag{C.135}$$

Therefore $\ell_{(n)} - \ell_{(i_0)} > 0$ and $b \geq y_{(i_0+1)} + \frac{\delta\ell_{(i_0)}}{\ell_{(n)}-\ell_{(i_0)}}$, and $\epsilon > 0$.

Let $\mathcal{U}_\epsilon = \{\mathbf{U} : 0 \leq U_j \leq \epsilon \ \forall j : 1 \leq j < i_0, \ell_{(i_0)} - \epsilon \leq U_{i_0} < \ell_{(i_0)}, 1 - \epsilon \leq U_j \leq 1 \ \forall j : i_0 < j \leq n\}$. Since $\epsilon > 0$, $\mathbb{P}(\mathcal{U}_\epsilon) > 0$. We will show that $\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) \geq \mathbb{P}\left(\cap_{i:1\leq i\leq n}\{U_{(i)} \geq \ell_{(i)}\}\right) + \mathbb{P}(\mathcal{U}_\epsilon)$, which implies:

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) \geq \mathbb{P}\left(\cap_{i:1\leq i\leq n}\{U_{(i)} \geq \ell_{(i)}\}\right) + \mathbb{P}(\mathcal{U}_\epsilon) \tag{C.136}$$

$$= 1 - \alpha + \mathbb{P}(\mathcal{U}_\epsilon) \tag{C.137}$$

$$> 1 - \alpha. \tag{C.138}$$

We will show that if $\ell_{(i_0)} - \epsilon \leq U_{(i_0)} \leq \ell_{(i_0)}, 1 - \epsilon \leq U_j \leq 1 \ \forall j : i_0 < j \leq n$ then $b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})$. Then the set $\mathbf{U}$ satisfying $b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})$ contains 2 disjoint sets $\mathcal{U}_\epsilon$ and the set $\mathbf{U}$ satisfying $U_i \geq \ell_i$ for all i, which implies Eq. C.136.

Let $\mathbf{U}'$ be the component-wise smallest element in $\mathcal{U}_\epsilon$: $U'_j = 0$ when $1 \leq j < i_0$, $U'_{(i_0)} = \ell_{(i_0)} - \epsilon \geq 0$ (because $\epsilon \leq \ell_{(i_0)}$) and $U'_{(j)} = 1 - \epsilon$ when $i_0 < j \leq n$. We will show that $m_{D^+,T}(\mathbf{y}, \mathbf{U}') \leq m_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})$ for all $\mathbf{y} \in \mathbb{S}_{D^+,T}(\mathbf{x})$.

We have:

$$m_{D^+,T}(\mathbf{y},\mathbf{U}') = b(1-U'_{(n)}) + \sum_{i=1}^{n} y_{(i)}(U'_{(i)} - U'_{(i-1)}) \tag{C.139}$$

$$= b\epsilon + y_{(i_0+1)}(1-\epsilon-(\ell_{(i_0)}-\epsilon)) + y_{(i_0)}(\ell_{(i_0)}-\epsilon) \tag{C.140}$$

$$= b\epsilon + y_{(i_0+1)}(1-\ell_{(i_0)}) + y_{(i_0)}(\ell_{(i_0)}-\epsilon) \tag{C.141}$$

$$= (b-y_{(i_0)})\epsilon + y_{(i_0+1)}(1-\ell_{(i_0)}) + y_{(i_0)}\ell_{(i_0)} \tag{C.142}$$

$$\leq (b-y_{(i_0)})\frac{(b-y_{(i_0+1)})(1-\ell_{(n)})}{b-a} + y_{(i_0+1)}(1-\ell_{(i_0)}) + y_{(i_0)}\ell_{(i_0)} \tag{C.143}$$

$$\leq (b-y_{(i_0)})\frac{(b-y_{(i_0+1)})(1-\ell_{(n)})}{b-y_{(i_0)}} + y_{(i_0+1)}(1-\ell_{(i_0)}) + y_{(i_0)}\ell_{(i_0)} \tag{C.144}$$

$$= (b-y_{(i_0+1)})(1-\ell_{(n)}) + y_{(i_0+1)}(1-\ell_{(i_0)}) + y_{(i_0)}\ell_{(i_0)} \tag{C.145}$$

$$= b(1-\ell_{(n)}) + y_{(i_0+1)}(\ell_{(n)}-\ell_{(i_0)}) + y_{(i_0)}\ell_{(i_0)} \tag{C.146}$$

$$= b(1-\ell_{(n)}) + y_{(i_0+1)}\sum_{i=i_0+1}^{n}(\ell_{(i)}-\ell_{(i-1)}) + y_{(i_0)}\ell_{(i_0)} \tag{C.147}$$

$$\leq b(1-\ell_{(n)}) + \sum_{i=i_0+1}^{n} y_{(i)}(\ell_{(i)}-\ell_{(i-1)}) + y_{(i_0)}\ell_{(i_0)} \tag{C.148}$$

$$\leq m(\mathbf{y},\boldsymbol{\ell}) \tag{C.149}$$

$$\leq m(\mathbf{x},\boldsymbol{\ell}). \tag{C.150}$$

Since $\mathbf{U}'$ is the component-wise smallest element in $\mathcal{U}_\epsilon$ and $m(\mathbf{x},\mathbf{U})$ is a linear function of $\mathbf{U}$ with negative coefficient, we have $m_{D^+,T}(\mathbf{y},\mathbf{U}) \leq m_{D^+,T}(\mathbf{y},\mathbf{U}') \leq m(\mathbf{x},\boldsymbol{\ell})$, and therefore $b_{D^+,T}(\mathbf{x},\mathbf{U}) \leq m(\mathbf{x},\boldsymbol{\ell})$ for all $\mathbf{U} \in \mathcal{U}_\epsilon$.

Note that if $x_i < b - \frac{(b-a)\ell_{(i_0)}}{\ell_{(n)}}\ \forall i, 1 \leq i \leq n$ then $a > b - \frac{b-m(\mathbf{x},\boldsymbol{\ell})}{\ell_{(i_0)}}$ and therefore if $n > i_0$ then $b^{\alpha}_{D^+,T}(\mathbf{x}) < m_{D^+}(\mathbf{x},\boldsymbol{\ell})$.

□

For the specific case where $\boldsymbol{\ell} = \mathbf{u}^{And}$ we have the following result.

Theorem 14. *Let* $D^+ = [a, b]$. *Let* $i_0 \stackrel{\text{def}}{=} \min\{i : 1 \leq i \leq n, u^{And}_{(i)} > 0\}$. *For any sample value* $\mathbf{x} \in D^n$, *for any sample size* n *and for all* $\alpha \in (0, 1)$, *using* $T(\mathbf{x}) = b^{\alpha,Anderson}_{\mathbf{u}^{And}}(\mathbf{x})$ *yields:*

$$b^{\alpha}_{D^+,T}(\mathbf{x}) = b^{\alpha,Anderson}_{\mathbf{u}^{And}}(\mathbf{x}) \text{ if } a \leq \frac{bu^{And}_{(i_0)} - b + m(\mathbf{x}, \boldsymbol{\ell})}{u^{And}_{(i_0)}} \tag{C.151}$$

For any sample value $\mathbf{x} \in D^n$, *for any sample size* n *and for all* $\alpha \in (0, 1)$ *satisfying* $\frac{(n-1)^2}{n} > \frac{ln(1/\alpha)}{2}$ [2], *using* $T(\mathbf{x}) = b^{\alpha,Anderson}_{\mathbf{u}^{And}}(\mathbf{x})$ *yields:*

$$b^{\alpha}_{D^+,T}(\mathbf{x}) < b^{\alpha,Anderson}_{\mathbf{u}^{And}}(\mathbf{x}) \text{ if } a > \frac{bu^{And}_{(i_0)} - b + m(\mathbf{x}, \boldsymbol{\ell})}{u^{And}_{(i_0)}} \tag{C.152}$$

Proof. The proof follows from Lemma 41.

First we note that $\mathbf{u}^{And}$ satisfies:

$$\mathbb{P}_{\mathbf{X}}(\ \forall x \in R, F(x) \geq G_{\mathbf{X},\mathbf{u}^{And}}(x)) = 1 - \alpha, \tag{C.153}$$

and that $\mathbf{u}^{And}_{(n)} < 1$ by definition. We will now show that if $\frac{(n-1)^2}{n} > \frac{ln(1/\alpha)}{2}$, then $u^{And}_{(n-1)} > 0$ and therefore $i_0 \leq n - 1$, which implies $n > i_0$.

Using the Dvoretsky-Kiefer-Wolfowitz inequality (Dvoretzky et al., 1956) to define the $1 - \alpha$ CDF lower bound via $\beta(n) = \sqrt{\ln(1/\alpha)/(2n)}$, we can compute a lower bound for u^{And}_i as follows:

$$u^{And}_i \geq \max\left(0, i/n - \sqrt{\frac{\ln(1/\alpha)}{2n}}\right). \tag{C.154}$$

Therefore if $\frac{n-1}{n} - \sqrt{\frac{\ln(1/\alpha)}{2n}}) > 0$ then $u^{And}_{(n-1)} > 0$. The condition $\frac{n-1}{n} - \sqrt{\frac{\ln(1/\alpha)}{2n}} > 0$ is equivalent to $\frac{(n-1)^2}{n} > \frac{\ln(1/\alpha)}{2}$. □

[2] To satisfy this condition, when $\alpha = 0.01$, $n \geq 5$. When $\alpha = 0.05$, $n \geq 4$. When $\alpha = 0.1$, $n \geq 3$.

BIBLIOGRAPHY

Agrawal, S. and Goyal, N. (2013). Further optimal regret bounds for Thompson sampling. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2013)*, volume 31 of *Proceedings of Machine Learning Research*, pages 99–107. PMLR.

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165.

Anderson, T. W. (1969a). Confidence limits for the value of an arbitrary bounded random variable with a continuous distribution function. *Bulletin of The International and Statistical Institute*, 43:249–251.

Anderson, T. W. (1969b). Confidence limits for the value of an arbitrary bounded random variable with a continuous distribution function. *Technical Report Number 1, Department of Statistics, Stanford University.*

Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M. I. (2003). An introduction to MCMC for machine learning. *Machine Learning*, 50(1):5–43.

Angus, J. E. (1994). The probability integral transform and related results. *SIAM Rev.*, 36(4):652–654.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Mollan, K. R., Sax, P. E., Daar, E. S., Adimora, A. A., Eron, J. J., and Mugavero, M. J. (2018). Generalizing evidence from randomized trials using inverse probability of sampling weights. *Journal of the Royal Statistical Society. Series A,(Statistics in Society)*, 181(4):1193.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Cichocki, A. and Amari, S. (2010). Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities. *Entropy*, 12:1532–1568.

Clopper, C. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413.

Coppock, A., Leeper, T. J., and Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, 115(49):12441–12446.

Dahabreh, I. J., Haneuse, S. J.-P. A., Robins, J. M., Robertson, S. E., Buchanan, A. L., Stuart, E. A., and Hernán, M. A. (2019). Study designs for extending causal inferences from a randomized trial to a target population.

Dahabreh, I. J., Robertson, S. E., Steingrimsson, J. A., Stuart, E. A., and Hernan, M. A. (2018). Extending inferences from a randomized trial to a new target population.

Dehejia, R., Pop-Eleches, C., and Samii, C. (2019). From local to global: External validity in a fertility natural experiment. *Journal of Business & Economic Statistics*, pages 1–27.

Diouf, M. A. and Dufour, J. M. (2005). Improved nonparametric inference for the mean of a bounded random variable with application to poverty measures.

Doucet, A. and Johansen, A. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 12:656–704.

DuGoff, E. H., Schuler, M., and Stuart, E. A. (2014). Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Services Research*, 49(1):284–303.

Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., McIntosh, C., Nellis, G., Adida, C. L., Arias, E., Bicalho, C., Boas, T. C., Buntaine, M. T., Chauchard, S., Chowdhury, A., Gottlieb, J., Hidalgo, F. D., Holmlund, M., Jablonski, R., Kramon, E., Larreguy, H., Lierl, M., Marshall, J., McClendon, G., Melo, M. A., Nielson, D. L., Pickering, P. M., Platas, M. R., Querubín, P., Raffler, P., and Sircar, N. (2019). Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials. *Science Advances*, 5(7).

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of a sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, 27:642–669.

Eaton, M. L. and Perlman, M. D. (1973). The Non-Singularity of Generalized Sample Covariance Matrices. *The Annals of Statistics*, 1(4):710 – 717.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.

Fienberg, S. E., Neter, J., and Leitch, R. A. (1977). Estimating the total overstatement error in accounting populations. *Journal of the American Statistical Association*, 72(358):295–302.

Fisher, J. A. and Kalbaugh, C. A. (2011). Challenging assumptions about minority participation in us clinical research. *American journal of public health*, 101(12):2217–2222.

Frost, J. (2021). Statistics by Jim: Central limit theorem explained.

Gaffke, N. (2005). Three test statistics for a nonparametric one-sided hypothesis on the mean of a nonnegative variable. *Mathematical Methods of Statistics*, 14(4):451–467.

Gopalan, A., Mannor, S., and Mansour, Y. (2014). Thompson sampling for complex online problems. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 100–108, Bejing, China. PMLR.

Greevy, R., Lu, B., Silber, J. H., and Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics*, 5(2):263–275.

Harshaw, C., Sävje, F., Spielman, D., and Zhang, P. (2019). Balancing covariates in randomized experiments using the gram-schmidt walk.

Hartman, E., Grieve, R., Ramsahai, R., and Sekhon, J. S. (2015). From sate to patt: combining experimental with observational studies to estimate population treatment effects. *JR Stat. Soc. Ser. A Stat. Soc.(forthcoming). doi*, 10:1111.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Most people are not weird. *Nature*, 466(7302):29–29.

Higgins, M. J., Sävje, F., and Sekhon, J. S. (2016). Improving massive experiments with threshold blocking. *Proceedings of the National Academy of Sciences*, 113(27):7369–7376.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Institute of Medicine (2001). *Small clinical trials: Issues and challenges*. The National Academies Press.

Kallus, N. (2018). Optimal a priori balance in the design of controlled experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):85–112.

Kawale, J., Bui, H. H., Kveton, B., Tran-Thanh, L., and Chawla, S. (2015). Efficient Thompson sampling for online matrix-factorization recommendation. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 1297–1305. Curran Associates, Inc.

Kveton, B., Szepesvari, C., Vaswani, S., Wen, Z., Lattimore, T., and Ghavamzadeh, M. (2019). Garbage in, reward out: Bootstrapping exploration in multi-armed bandits. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3601–3610, Long Beach, California, USA. PMLR.

Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press.

Learned-Miller, E. and DeStefano, J. (2008). A probabilistic upper bound on differential entropy. *IEEE Transactions on Information Theory*, 54(11):5223–5230.

Learned-Miller, E. and Thomas, P. S. (2019). A new confidence interval for the mean of a bounded random variable. *arXiv preprint arXiv:1905.06208*.

Li, X., Ding, P., and Rubin, D. B. (2018). Asymptotic theory of rerandomization in treatment–control experiments. *Proceedings of the National Academy of Sciences*, 115(37):9157–9162.

Liu, C.-Y. and Li, L. (2016). On the prior sensitivity of thompson sampling. In *Algorithmic Learning Theory*, pages 321–336, Cham. Springer International Publishing.

Lu, X. and Van Roy, B. (2017). Ensemble sampling. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 3260–3268. Curran Associates, Inc.

Maurer, A. and Pontil, M. (2009). Empirical Bernstein bounds and sample variance penalization. In *Proceedings of the Twenty-Second Annual Conference on Learning Theory*, pages 115–124.

Meager, R. (2019). Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, 11(1):57–91.

Minka, T. (2005). Divergence measures and message passing. Technical Report MSR-TR-2005-173, Microsoft Research.

Minka, T., Winn, J., Guiver, J., Zaykov, Y., Fabian, D., and Bronskill, J. (2018). /Infer.NET 0.3. Microsoft Research Cambridge. http://dotnet.github.io/infer.

Miratrix, L. W., Sekhon, J. S., Theodoridis, A. G., and Campos, L. F. (2018). Worth weighting? how to think about and use weights in survey experiments. *Political Analysis*, 26(3):275–291.

Morgan, K. L., Rubin, D. B., et al. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2):1263–1282.

Munger, K. (2018). Temporal validity in online social science.

Ordentlich, E. and Weinberger, M. J. (2004). A distribution dependent refinement of Pinsker's inequality. *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, pages 29–.

Osband, I., Van Roy, B., and Wen, Z. (2016). Generalization and exploration via randomized value functions. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pages 2377–2386. JMLR.org.

Pap, G. and van Zuijlen, M. C. A. (1995). The Stringer bound in case of uniform taintings. *Computers and Mathematics with Applications*, 29(10):51–59.

Pearl, J. and Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 25.

Phan, M., Abbasi Yadkori, Y., and Domke, J. (2019). Thompson sampling and approximate inference. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Phan, M., Arbour, D., Dimmery, D., and Rao, A. (2021a). Designing transportable experiments under s-admissability. In Banerjee, A. and Fukumizu, K., editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 2539–2547. PMLR.

Phan, M., Thomas, P. S., and Learned-Miller, E. (2021b). Towards practical mean bounds for small samples. In *Proceedings of the 38th International Conference on Machine Learning (ICML-21)*.

Qin, C., Klabjan, D., and Russo, D. (2017). Improving the expected improvement algorithm. In *Advances in Neural Information Processing Systems 30*, pages 5381–5391. Curran Associates, Inc.

Riquelme, C., Tucker, G., and Snoek, J. (2018). Deep Bayesian bandits showdown: An empirical comparison of bayesian deep networks for thompson sampling. In *International Conference on Learning Representations (ICLR 2018)*.

Romano, J. P. and Wolf, M. (2000). Finite sample nonparametric inference and large sample efficiency. *Annals of Mathematical Statistics*, 28(3):756–778.

Royden, H. L. (2018 - 2010). *Real analysis / H.L. Royden, Stanford University, P.M. Fitzpatrick, University of Maryland, College Park.* Pearson modern classic. Pearson, New York, NY, fourth edition [2018 reissue]. edition.

Rubin, D. B. (2011). Causal inference using potential outcomes. *Journal of the American Statistical Association.*

Rudolph, K. E. and van der Laan, M. J. (2017). Robust estimation of encouragement-design intervention effects transported across sites. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 79(5):1509.

Russo, D. (2016). Simple Bayesian algorithms for best arm identification. In *29th Annual Conference on Learning Theory (COLT 2016)*, volume 49 of *Proceedings of Machine Learning Research*, pages 1417–1418. PMLR.

Russo, D. and Roy, B. V. (2016). An information-theoretic analysis of Thompson sampling. *Journal of Machine Learning Research*, 17(68):1–30.

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

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55.

Serfling, R. (1980). *Approximation theorems of mathematical statistics*. Wiley series in probability and mathematical statistics : Probability and mathematical statistics. Wiley, New York, NY [u.a.], [nachdr.] edition.

Stringer, K. W. (1963). Practical aspects of statistical sampling. *Proceedings of Business and Economic Statistics Section, American Statistical Association*.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2):369–386.

Student (1908). The probable error of a mean. *Biometrika*, pages 1–25.

Thomas, P. S., Theocharous, G., and Ghavamzadeh, M. (2015). High-confidence off-policy evaluation. In *AAAI*.

Tran, D., Kucukelbir, A., Dieng, A. B., Rudolph, M., Liang, D., and Blei, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint arXiv:1610.09787*.

Urteaga, I. and Wiggins, C. (2018). Variational inference for the multi-armed contextual bandit. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS 2018)*, volume 84 of *Proceedings of Machine Learning Research*, pages 698–706. PMLR.

Vivalt, E. (2015). Heterogeneous treatment effects in impact evaluation. *American Economic Review*, 105(5):467–70.

Wang, Y., Gupta, S., Lu, J., Mahmoudzadeh, A., and Liu, S. (2019). On heavy-user bias in a/b testing. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 2425–2428.

Waudby-Smith, I. and Ramdas, A. (2021). Estimating means of bounded random variables by betting. *arXiv preprint arXiv:2010.09686*.