ABSTRACT

Title of dissertation: STOCHASTIC OPTIMIZATION: APPROXIMATE BAYESIAN INFERENCE AND COMPLETE EXPECTED IMPROVEMENT

Ye Chen
Doctor of Philosophy, 2018

Dissertation directed by: Professor Ilya Ryzhov
Department of Decision, Operations, and Information Technologies

Stochastic optimization includes modeling, computing and decision making. In practice, due to the limitation of mathematical tools or real budget, many practical solution methods are designed using approximation techniques or taking forms that are efficient to compute and update. These models have shown their practical benefits in different backgrounds, but many of them also lack rigorous theoretical support. Through interfacing with statistical tools, we analyze the asymptotic properties of two important Bayesian models and show their validity by proving consistency or other limiting results, which may be useful to algorithmic scientists seeking to leverage these computational techniques for their practical performance.

The first part of the thesis is the consistency analysis of sequential learning algorithms under approximate Bayesian inference. Approximate Bayesian inference is a powerful methodology for constructing computationally efficient statistical mechanisms for sequential learning from incomplete or censored information.Approximate

Bayesian learning models have proven successful in a variety of operations research and business problems; however, prior work in this area has been primarily computational, and the consistency of approximate Bayesian estimators has been a largely open problem. We develop a new consistency theory by interpreting approximate Bayesian inference as a form of stochastic approximation (SA) with an additional "bias" term. We prove the convergence of a general SA algorithm of this form, and leverage this analysis to derive the first consistency proofs for a suite of approximate Bayesian models from the recent literature.

The second part of the thesis proposes a budget allocation algorithm for the ranking and selection problem. The ranking and selection problem is a well-known mathematical framework for the formal study of optimal information collection. Expected improvement (EI) is a leading algorithmic approach to this problem; the practical benefits of EI have repeatedly been demonstrated in the literature, especially in the widely studied setting of Gaussian sampling distributions. However, it was recently proved that some of the most well-known EI-type methods achieve suboptimal convergence rates. We investigate a recently-proposed variant of EI (known as "complete EI") and prove that, with some minor modifications, it can be made to converge to the rate-optimal static budget allocation without requiring any tuning.

STOCHASTIC OPTIMIZATION: APPROXIMATE BAYESIAN INFERENCE AND COMPLETE EXPECTED IMPROVEMENT

by

Ye Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Paul Smith, Chair/Advisor
Professor Ilya Ryzhov, Co-Chair/Co-Advisor
Professor Michael Fu
Professor Kunpeng Zhang
Professor Paul Schonfeld

Dedication

I dedicate this dissertation to my girlfriend, Qi Wang, and my parents, Xiaoping Song and Xixue Chen.

Acknowledgments

First of all, I would like to express my most sincere gratitude to my advisor Professor Ilya O. Ryzhov, for his vision, advice and support throughout these years. My academic career would not have been possible without his consistent encouragement. What I have learnt from him is not only ideas and skills about how to do research, but also a new attitude and outlook on life. I would also like to thank Professor Paul Smith for his guidance and patience that helps me proceed through my graduate life.

Special thanks to Professor Michael Fu, Professor Kunpeng Zhang and Professor Paul Schonfeld for serving on my committee. I would also like to thank Professor Michael Fu and Professor Yuan Liao for their help during my job searching process.

I would like to thank all my friends and relatives. Forgive me that I cannot put all your names here. I would also like to thank all the music that accompanies me every day and night, and thank all the books, dramas, games and experiences that help me become what I am.

Particularly, I would like to thank my loving girlfriend, Qi Wang, for the happiness and support she has been bringing to me. I also thank my grandfather, Qiang Chen, for letting me realize the power of knowledge since I was a kid. Finally, I would like to thank my parents, Xiaoping Song and Xixue Chen, for their unconditional love and support. They are my very first teacher in this world. I am grateful to everything they have given to me.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 Sequential Learning: Approximate Bayesian Inference

Bayesian statistics allows decision-makers to estimate unknown parameters, but also to include a detailed model of their uncertainty about these estimates. In practice, instead of a fixed dataset, it is more likely to have a stochastic data stream in many applications. For example, in online digital goods auctions [1], each observation is the response of a buyer for a proposed price by the seller in every potential transaction, clearly, the transactions must occur in a sequential manner rather than all occur at the same time. Bayesian models allow the seller to represent the potential for error in the demand model, which in turn allows for more robust adaptive pricing methods. However, these models should be updated sequentially in order to take advantage of new information as soon as it arrives. When the observation comes from a distribution that is conjugate with the prior belief, the posterior distribution then comes from the same distribution family as the prior does, which makes it easy to update the model since it can be completely characterized by the parameter set of this distribution family and updating the model only requires updating the parameter set.

In many situations where the observation is censored or only partially available [2, 3], it is impossible to have a conjugate Bayesian model. Approximate Bayesian inference is a methodology to handle this issue through creating an artificial posterior distribution that comes from the prior's distribution family and letting it mimic the exact posterior distribution according to some criterion. There are different approaches to build approximate Bayesian learning models. For example, one approach is the moment-matching method [4] by solving moment-matching equations in order to make the moments of the artificial posterior distribution equal to corresponding moments of the exact posterior. Other methods include minimizing the Kullback-Leibler divergence between the two distributions [5] and variational Bayesian inference [6, 7] by approximating complicated functions using their Taylor expansions.

Similar to conjugate Bayesian models, approximate Bayesian learning models can be efficiently updated via recursive equations for a small set of parameters, thus avoiding the difficulty for handling the complicated exact posteriors that even may not come from any common distribution family. Simple statistical models make it easy to interface with control policies, thus approximate Bayesian inference is applied in a wide variety of problems, for example, the ranking and selection problem [5]. However, although the numerical advantage of approximate Bayesian models has been repeatedly shown in the past literature, rigorous theoretical analysis of the validity of these models was not studied in any of the prior work. Intuitively, one can see that approximate Bayesian learning models bear a strong resemblance to the classic stochastic approximation (SA) algorithm, whose convergence was fully studied by [8]; however the classic SA framework cannot directly be applied to analyze these methods because of certain important differences. Thus the convergence of the approximate Bayesian algorithms can not be simply explained by the classic SA framework.

1.2 Ranking and Selection: Complete Expected Improvement

In the ranking and selection problem (R&S) problem with finitely many "alternatives" (or "systems"), each alternative has an unknown system value (for simplicity, suppose different alternatives have different system values), and we wish to identify the optimal alternative that has the largest system value among all the alternatives. For any alternative, we are able to observe noisy samples about the unknown system value (population mean); however, we are limited to a fixed budget, i.e., the total number of samples that could be allocated to the alternatives is fixed. Under independent assumptions, the sample of one alternative does not provide any information about other alternatives. After all the sampling budget has been consumed, we select the alternative with the largest sample mean and we say "correct selection" occurs if the selected alternative is the optimal alternative that has the largest population mean. Since the total budget is fixed, we would like to find an allocation strategy that could maximize the probability of correct selection.

With regard to maximizing the probability of correct selection, [9] gives the optimal budget allocation, where the proportions of the total budget assigned to the alternatives satisfy two optimality conditions. However, these optimality conditions depend on the unknown system values, which makes it impossible to solve and apply them directly. Thus the practitioners have preferred to use simple methods that are easy to code and perform well in practice, and one of the most popular methods of this type is the expected improvement (EI) algorithm [10]. EI is a sequential allocation strategy, where every time after an alternative is sampled, a new observation is available and it provides information to help the decision-maker select the next alternative to be sampled. There are many variants of the EI criterion designed for different settings under different sets of modeling assumptions, such as the knowledge gradient criterion [11] and the LL_1 criterion [12]. Although the computational advantage and practical benefit of these methods have been well studied, the theoretical behavior was not fully learnt until [13]. It tuns out that these methods produce different asymptotic allocations, but none of them achieve the optimal budget allocation. Aside from EI and its variants, [14] provides a way to recover the optimal allocation through reverse-engineering the optimality conditions, but this method requires extra computational effort compared to EI and it does not have a natural interpretation as EI does. Recent work such as [15] has shown that it is possible to recover the optimal allocation, but involves an extra tuning parameter, and the optimality conditions are only achieved when the tuning parameter is assigned some specific value, which is, however, unknown without knowledge of the system values.

Recently, [16] proposed complete expected improvement (CEI) criterion. Unlike classic EI, which evaluates the expected improvement over the current-best sample mean from sampling every alternative, CEI evaluates the expected improvement over the current-best alternative from sampling every seemingly-suboptimal alternative. This feature gives CEI the potential to recover the optimal budget allocation while no extra computational effort or tuning work is required.

1.3 Outline of Thesis

In Chapter 2, I present the consistency analysis of sequential learning algorithms under approximate Bayesian inference. Through a motivating example, I first establish a connection between the approximate Bayesian learning algorithm and the traditional stochastic approximation algorithm by showing their similarities as well as the differences. Then I point out the approximate Bayesian learning algorithm does not fit the traditional stochastic approximation framework due to these differences. After that, I define a general stochastic approximation algorithm with some additional "bias" terms involved and show the convergence of this algorithm. Finally, a suite of existing approximate Bayesian models from the recent literature is studied, and by interpreting these algorithms as stochastic approximation algorithms with "bias" terms, I show the convergence of each one under the general framework.

In Chapter 3, I present the modified complete expected improvement algorithm for the ranking and selection problem with finite systems. Complete expected improvement is a recently-proposed criterion that can be viewed as a variant of the expected improvement criterion. Expected improvement (EI) criterion has been widely applied due to its practical benefit, but it was recently shown that some of the well-known EI-type methods are only suboptimal with respect to minimizing the probability of incorrect selection. I propose an algorithm based on the complete expected improvement criterion, which requires no additional tuning work or computational effort than traditional EI-type algorithms, and show this algorithm achieves the optimal budget allocation strategy asymptotically. At last, I conduct a numerical experiment comparing this algorithm with some other allocation algorithms as well as the optimal allocation strategy for illustration.

Chapter 4 provides the conclusion to the thesis.

Chapter 2: Consistency Analysis of Sequential Learning under Approximate Bayesian Inference

2.1 Introduction

Approximate Bayesian inference is a statistical learning methodology with wide-ranging applications in sequential information collection problems, particularly those where a decision-maker must use incomplete or censored information to maintain and update a set of beliefs about one or more unknown population parameters. Approximate Bayesian models are attractive for their computational tractability, and often lead to compact belief representations that can interface with simple and interpretable policies for related decision problems. In the recent literature, approximate Bayesian methods have been successful in the following applications:

- *Market design* [2]. Many financial markets designate official market-makers whose role is to increase liquidity and promote trading by being available to buy and sell securities. By experimenting with bid and ask prices, a market-maker can learn the market value of an asset by observing traders' willingness to buy and sell.

- *Posted-price auctions* [1]. A seller chooses a price for a digital good in order to maximize expected revenue. Buyer valuations of the good cannot be observed directly, and must be inferred from buyers' yes/no responses to posted prices. The seller's problem is characterized by considerable uncertainty about the valuation distribution.

- *E-sports* [4]. Large numbers of players log on to an online gaming service. In order to promote fair and competitive play, the service seeks to match players of similar skill levels. However, "skill" cannot be measured directly; rather, the game master must infer it from a player's win/loss history.

In these problems, sequential learning is needed for improved decision-making: for instance, in the market-making application, each new transaction changes our perception of the optimal bid and ask prices, which should lead to improved earnings over time. Learning is broadly relevant in this way throughout any subdomain of operations research in which decisions are made based on data. Approximate Bayesian models specifically have proved themselves to be useful in the following broad methodological areas:

- *Big data analytics.* Logistic regression is a standard statistical tool for forecasting [17], pricing [18], and order planning [19]. Approximate Bayesian learning models our uncertainty about the regression coefficients and enables us to update them in a computationally efficient manner [6, 20].

- *Approximate dynamic programming.* Many resource allocation problems in transportation [21] and energy [3] are subject to the curse of dimensionality, rendering classic optimization methods intractable [22] and introducing the challenge of exploration in large state spaces [23]. A multivariate Bayesian prior can be used to learn about large parts of the state space in fewer iterations.

- *Ranking and selection.* In the ranking and selection problem [24,25], a limited simulation budget is allocated sequentially in order to discover the best of a finite set of design alternatives. Approximate Bayesian learning with correlated beliefs can discover similarities between designs [5] and learn their values more quickly.

The main contribution of the present paper is a theoretical framework that can be leveraged to produce new consistency proofs in *each* of the above-listed methodological and application areas. Virtually all of the existing work on *sequential* approximate Bayesian learning is computational/algorithmic in nature: approximate Bayesian models have repeatedly demonstrated significant practical benefits (see [26] for an overview), but have remained mostly unamenable to the usual forms of consistency analysis. Our work is among the first to provide broad theoretical support for approximate Bayesian procedures: we prove, for the first time, the statistical consistency of a wide variety of previously-proposed approximate Bayesian estimators, providing insight into their good empirical performance. We also develop theoretical tools that may be used by researchers to develop similar proofs for other problems and applications.

2.1.1 Problem Background

In Bayesian analysis, the prior distribution of an unknown population parameter is an object of belief, chosen by the decision-maker based on past knowledge or other considerations. Given a sample of data, the posterior distribution of the parameter models the change in our beliefs resulting from the acquisition of information. The property of *conjugacy* arises when the posterior belongs to the same family as the prior (e.g., both are normal). If this is the case, the beliefs can be compactly represented by a small number of parameters (such as a mean and a variance), which can often be updated very efficiently.

The problem of approximate Bayesian inference occurs when conjugacy does not hold, i.e., there is a mismatch between the prior distribution and the sampling distribution (this easily happens when the data are censored). In such cases, the traditional approach has been to apply approximate Bayesian computation [27, 28] based on Markov chain Monte Carlo procedures [29]. These techniques are computationally expensive but provably convergent [30,31]. However, this entire literature assumes that the problem is *static*: there is a single dataset and a single stage of inference (i.e., only one posterior distribution to be computed). A rich asymptotic theory has been developed for this class of procedures (see, e.g., recent advances by [32] and [33]), but the underlying assumption is always that there is a single inference problem to be solved.

In sharp contrast with the above, we consider a *dynamic* problem in which information is collected *sequentially*. Our motivating applications all involve multistage optimization where the quality of each new decision (e.g., bid and ask prices) may be improved using feedback from past decisions. In this paper, we do not study the problem of *how* exactly these decisions should be made; however, *any* optimization approach should benefit from adaptive learning. For this reason, we would like to update our beliefs immediately after every new data point, meaning that we are now faced with a sequence of inference problems, each of which has a sample size of 1. Conjugacy now becomes much more valuable: the ability to compactly model a distribution of belief using a small set of parameters enables the decision-maker to create and apply tractable optimization methods that take the belief parameters as inputs and return recommended decisions. Such parametric methods may be required to run very quickly and produce recommendations in real time. Because conjugate learning models are easy to store and update, they greatly simplify the design of algorithms for adaptive decision-making.[1]

In all of the applications considered in this paper, there is no natural choice of prior distribution that is conjugate with the observations. Although a conjugate prior may technically be developed for any distribution belonging to an exponential family [36], in our applications the data are not i.i.d., but rather depend on additional inputs that may be controllable by the decision-maker (for instance, a trader's response to a market-maker depends on both the market value of the asset and the bid/ask prices). This structure may lead us to assume some particular functional

[1] Bayesian learning models in particular enable the design of anticipatory policies that have some form of intelligent experimentation built in; see, e.g., the popular Thompson sampling method [34] or the Gittins index approach of [35].

form for the dependence of the observations on the controls and unknown parameters, which may preclude the use of standard constructions of conjugate priors.

These factors motivate the development of *sequential* approximate Bayesian models, which essentially impose conjugacy by creating an artificial posterior distribution from the same family as the prior (e.g., normal), then choosing the parameters of that artificial distribution in a way that approximates (in some sense) the exact posterior. The approximation may be built using strategies such as moment-matching [37,38], density filtering [5], and variational bounds [6,39]. In many cases, the approximate posterior parameters can be computed in closed form, which is quite convenient for practical implementation and has been the main reason for continued interest in this area. However, despite the large body of empirical evidence that these models work well, they are quite difficult to analyze theoretically. In fact, outside of a few special cases [40], it is unknown whether approximate Bayesian estimators are even consistent. This has also imposed a limitation on algorithmic research in this area, as it is not possible to provide any performance guarantees for any optimization algorithm if the underlying statistical model is invalid.

2.1.2 Summary of Our Approach and Results

We present a new theoretical framework that enables rigorous study of the consistency problem.[2] First, using a simple illustrative example in Section 2.2, we observe that approximate Bayesian updates can be interpreted as a form of

[2] A brief summary version of our approach, without the full technical details, appeared in the *Proceedings of the 2016 Winter Simulation Conference* [41].

stochastic approximation (or SA; see [8]), a class of provably convergent, frequentist algorithms that optimize nonlinear functions (e.g., likelihood functions) using stochastic observations of their gradients at individual points. The approximate Bayesian update can be viewed as SA with the addition of a "bias" term representing the difference between the frequentist and Bayesian versions of the stochastic gradient. Intuitively, if this bias is "small," the Bayesian procedure should converge. In Section 2.3, we formalize this intuition by proposing a modified Robbins-Monro SA algorithm with a similar bias term. Although there is a rich convergence theory for SA, our algorithm does not fit into the standard convergence conditions [42–44], so we develop a new set of conditions and give a convergence proof.

Our approach should be contrasted with that of [45], which to our knowledge is the only previous effort to address the general consistency problem. [45] also points out the apparent similarity between approximate Bayesian updating and stochastic gradient methods, and sketches out a convergence argument in the context of normal priors and moment-matching. However, this argument assumes that the posterior variance is negligible and that the posterior mean is "sharply peaked" around the true value, i.e., from the start we are already arbitrarily close to the desired limit. In marked contrast, we rigorously handle the asymptotic behaviour of the posterior from any starting conditions, under a standard set of SA assumptions.

We demonstrate the versatility of our SA analogy by using it (in Sections 2.4-2.5) to create consistency proofs for an entire suite of applications taken from existing literature, including previously-proposed approximate Bayesian schemes for market design [2], posted-price auctions [1], and e-sports [4]. In addition, we prove the consistency of three previously-proposed multivariate approximate Bayesian schemes for logistic regression [20], ranking and selection [38] and approximate dynamic programming [46]. Bayesian learning is especially powerful in these examples since the posterior covariance matrix allows us to learn about multiple unknown values after sampling just one. This practical benefit often outweighs any statistical loss incurred by using approximate posteriors [47, 48].

We emphasize that, on one hand, *every one* of these applications comes from an existing paper; these papers proposed the Bayesian models in question and conducted extensive computational experiments and comparisons to other techniques. Yet, on the other hand, *none* of this previous work attempted any rigorous consistency analysis, even within the confines of the specific application of interest. Our paper is the first to show the consistency of *all* of these previously-proposed models, thus contributing to all of the corresponding application areas. We note that our examples include at least one large-scale industry application [4] in which approximate Bayesian inference was successfully deployed in practice, and we also highlight our analysis of Bayesian logistic regression, a model that has existed for nearly 30 years without any progress on consistency. Although there is no way to guarantee that our framework is applicable to every possible approximate Bayesian model, the variety of models and problem domains on display in Sections 2.4-2.5 speaks for itself with regard to applicability.

2.2 Example: Learning from Censored Binary Observations

We first present a simple example that illustrates the main issues of this paper. The goal of this problem is to estimate a single unknown parameter based on censored binary observations. We will use approximate Bayesian inference to construct a computationally tractable estimator that can be easily updated. The analogy to stochastic approximation will then become clear.

Let $(Y_n)_{n=1}^{\infty}$ be a sequence of i.i.d. samples from the common distribution $\mathcal{N}(\theta, \lambda^2)$, where θ is the unknown parameter to be learned, and λ^2 is assumed to be known for simplicity. We impose the Bayesian model $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$, where μ_0 is an estimate of θ and σ_0 represents our uncertainty about that estimate. It is well-known that, if $Y_1, Y_2, ...$ are directly observable, then the posterior distribution of θ given $Y_1, ..., Y_n$ is normal for any n [49]. In that case, the posterior distribution is always completely characterized by the pair (μ_n, σ_n), which can be updated recursively after each observation. The consistency of the estimator μ_n follows trivially, as its update is equivalent to recursive sample averaging.

Now suppose that $Y_1, Y_2, ...$ are not directly observable. Instead, we observe a sequence $(B_n)_{n=1}^{\infty}$ of censored observations defined by

$$B_{n+1} = 1_{\{Y_{n+1} < b_n\}},$$

where the sequence $(b_n)_{n=0}^{\infty}$ represents a control policy. For instance, b_n could be a dosage decision for a drug, with Y_{n+1} representing the maximum allowable dosage level before patient $n+1$ experiences adverse effects and B_{n+1} indicating the presence of those effects. For simplicity, we treat (b_n) as a fixed (deterministic) sequence; however, our convergence analysis will be unaffected if b_n is allowed to be measurable with respect to $B_1, ..., B_n$, as would be the case if the dosage were chosen adaptively based on the outcomes of past trials.[3]

It is easily seen that the posterior density $P(\theta \in dx \mid B_1)$ is not normal, even after just one observation. As more samples are collected, the posterior will become a more complicated mixture, increasingly difficult to store and update. We will address this problem by using approximate Bayesian inference to create an "approximately" normal posterior. After choosing the parameters of this artificial posterior distribution, we will then discard the exact, non-normal posterior and proceed to the next sample using the approximation as our distribution of belief. By doing this, we regain the ability to describe our beliefs using just two parameters, but presumably incur statistical error due to the approximation.

To make the example more concrete, let us apply the method of moment-matching, also known as expectation propagation [37, 45]. This is not the only possible approach (others will be seen in later examples), but in this particular setting it is useful for illustration purposes. Assuming that $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and that B_1 is given, let $\tilde{\theta} \sim \mathcal{N}(\mu_1, \sigma_1^2)$, where μ_1 and σ_1 are chosen to satisfy the equations

$$\int_{\mathbb{R}} xP\left(\tilde{\theta} \in dx\right) \quad = \quad \int_{\mathbb{R}} xP\left(\theta \in dx \mid B_1\right),$$

[3] The exact choice of (b_n) is exogenous to the estimation problem; for instance, one may choose b_n to keep the estimated probability of side effects below some tolerance level. However, the validity of the underlying statistical mechanism, which is our main focus in this paper, is critical to the overall performance of any such approach.

$$\int_{\mathbb{R}} x^2 P\left(\tilde{\theta} \in dx\right) \;\; = \;\; \int_{\mathbb{R}} x^2 P\left(\theta \in dx \,|\, B_1\right).$$

Thus, the first two moments of $\tilde{\theta}$ are equal to those of the non-normal posterior. We then move to the next stage of sampling and repeat the process with the next observation B_2 under the assumption that $\theta \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$. The following result shows that, in the $(n+1)$st stage of sampling, the approximate posterior parameters $(\mu_{n+1}, \sigma_{n+1})$ may be efficiently computed from the parameters (μ_n, σ_n) in the nth stage and the next observation B_{n+1}.

Proposition 2.2.1. *The moment-matching equations in the* $(n+1)$*st stage admit a closed-form solution given by*

$$\begin{aligned}\mu_{n+1} &= \mu_n - \sigma_n^2 \left(B_{n+1} \frac{1}{\sqrt{\lambda^2 + \sigma_n^2}} \frac{\phi\left(p_n\right)}{\Phi\left(p_n\right)} \right. \\ &\qquad \left. - \left(1 - B_{n+1}\right) \frac{1}{\sqrt{\lambda^2 + \sigma_n^2}} \frac{\phi\left(p_n\right)}{1 - \Phi\left(p_n\right)} \right), \qquad (2.1) \\ \sigma_{n+1}^2 &= \sigma_n^2 \left(1 - B_{n+1} \frac{\sigma_n^2}{\lambda^2 + \sigma_n^2} \frac{p_n \phi\left(p_n\right) \Phi\left(p_n\right) + \phi^2\left(p_n\right)}{\Phi^2\left(p_n\right)} \right. \\ &\qquad \left. - \left(1 - B_{n+1}\right) \frac{\sigma_n^2}{\lambda^2 + \sigma_n^2} \frac{\phi^2\left(p_n\right) - p_n \phi\left(p_n\right)\left(1 - \Phi\left(p_n\right)\right)}{\left(1 - \Phi\left(p_n\right)\right)^2} \right), \qquad (2.2)\end{aligned}$$

where ϕ, Φ *are the standard normal pdf and cdf, and*

$$p_n = \frac{b_n - \mu_n}{\sqrt{\lambda^2 + \sigma_n^2}}. \qquad (2.3)$$

Proof. Suppose at the $(n+1)$st stage, the prior distribution of θ is $\theta \sim \mathcal{N}\left(\mu_n, \sigma_n^2\right)$. If $B_{n+1} = 1$, the moment-matching equations yield

$$\mu_{n+1} = \frac{\int \theta \frac{1}{\sigma_n} \phi\left(\frac{\theta - \mu_n}{\sigma_n}\right) \Phi\left(\frac{b_n - \theta}{\lambda}\right) d\theta}{\int \frac{1}{\sigma_n} \phi\left(\frac{\theta - \mu_n}{\sigma_n}\right) \Phi\left(\frac{b_n - \theta}{\lambda}\right) d\theta} = \mu_n + \frac{\int \frac{\theta - \mu_n}{\sigma_n} \phi\left(\frac{\theta - \mu_n}{\sigma_n}\right) \Phi\left(\frac{b_n - \theta}{\lambda}\right) d\theta}{\int \frac{1}{\sigma_n} \phi\left(\frac{\theta - \mu_n}{\sigma_n}\right) \Phi\left(\frac{b_n - \theta}{\lambda}\right) d\theta}$$

and

$$\begin{aligned}
\sigma_{n+1}^2 &= \frac{\int \theta^2 \frac{1}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta}{\int \frac{1}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta} - \mu_{n+1}^2 \\
&= \frac{\int \frac{(\theta-\mu_n)^2}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta}{\int \frac{1}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta} - (\mu_{n+1}-\mu_n)^2 \\
&= \frac{\int \frac{(\theta-\mu_n)^2}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta}{\int \frac{1}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta} - \left(\frac{\int \frac{\theta-\mu_n}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta}{\int \frac{1}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta}\right)^2.
\end{aligned}$$

We then evaluate the integrals

$$\begin{aligned}
\int \frac{1}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta &= \Phi\left(\frac{b_n-\mu_n}{\sqrt{\lambda^2+\sigma_n^2}}\right), \\
\int \frac{\theta-\mu_n}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta &= -\frac{\sigma_n^2}{\sqrt{\lambda^2+\sigma_n^2}}\phi\left(\frac{b_n-\mu_n}{\sqrt{\lambda^2+\sigma_n^2}}\right), \\
\int \frac{(\theta-\mu_n)^2}{\sigma_n}\phi\left(\frac{\theta-\mu_n}{\sigma_n}\right)\Phi\left(\frac{b_n-\theta}{\lambda}\right)d\theta &= \sigma_n^2\Phi\left(\frac{b_n-\mu_n}{\sqrt{\lambda^2+\sigma_n^2}}\right) \\
&\quad -\frac{\sigma_n^4}{\lambda^2+\sigma_n^2}\frac{b_n-\mu_n}{\sqrt{\lambda^2+\sigma_n^2}}\phi\left(\frac{b_n-\mu_n}{\sqrt{\lambda^2+\sigma_n^2}}\right),
\end{aligned}$$

whence

$$\begin{aligned}
\mu_{n+1} &= \mu_n - \frac{\sigma_n^2}{\sqrt{\lambda^2+\sigma_n^2}}\frac{\phi(p_n)}{\Phi(p_n)}, \\
\sigma_{n+1}^2 &= \frac{\sigma_n^2\Phi(p_n) - \frac{\sigma_n^4}{\lambda^2+\sigma_n^2}p_n\phi(p_n)}{\Phi(p_n)} - \left(\frac{\sigma_n^2}{\sqrt{\lambda^2+\sigma_n^2}}\frac{\phi(p_n)}{\Phi(p_n)}\right)^2 \\
&= \sigma_n^2 - \frac{\sigma_n^4}{\lambda^2+\sigma_n^2}\frac{p_n\phi(p_n)\Phi(p_n)+\phi^2(p_n)}{\Phi^2(p_n)},
\end{aligned}$$

as required. A similar argument can be applied when $B_{n+1} = 0$ to obtain the required result. □

It is not obvious whether $\mu_n \to \theta$. In fact, one may intuitively expect that this will *not* happen: first, the censored observations (B_n) carry less information than the complete observations (Y_n), and second, each stage of sampling necessitates a new approximation and thus may compound the statistical error of the model. Thus, it is somewhat surprising that μ_n is, in fact, consistent; that is, it is guaranteed to recover θ w.p. 1.

A rigorous framework for proving this result will be given in Section 2.3. Here, we provide additional intuition for our approach by pointing out that (2.1) can be viewed as a Robbins-Monro stochastic approximation (SA) procedure of the form

$$\mu_{n+1} = \mu_n - \alpha_n G_n\left(B_{n+1}, \mu_n, \sigma_n\right), \tag{2.4}$$

with the posterior variance σ_n^2 serving as the stepsize α_n. More specifically, (2.1) is nearly identical to a version of SA, known as "online gradient descent" or OGD, that was proposed by [50] for frequentist statistical estimation. In the context of our example, OGD is applied as follows. Suppose that θ is fixed; then, the marginal log-likelihood function of B_{n+1} is given by

$$\log L\left(B_{n+1}; \mu\right) = B_{n+1} \log \Phi\left(\frac{b_n - \mu}{\lambda}\right) + \left(1 - B_{n+1}\right) \log\left(1 - \Phi\left(\frac{b_n - \mu}{\lambda}\right)\right). \tag{2.5}$$

The OGD algorithm is given by (2.4) with

$$G_n\left(B_{n+1}, \mu_n\right) = B_{n+1} \frac{1}{\lambda} \frac{\phi\left(q_n\right)}{\Phi\left(q_n\right)} - \left(1 - B_{n+1}\right) \frac{1}{\lambda} \frac{\phi\left(q_n\right)}{1 - \Phi\left(q_n\right)}, \tag{2.6}$$

where $q_n = \frac{b_n - \mu_n}{\lambda}$. In words, (2.6) is the gradient of (2.5) evaluated at the current iterate μ_n. It is easy to see that $\mathbb{E}\left(G_n\left(B_{n+1}, \mu_n\right)\right) = 0$ if and only if $\mu_n = \theta$. Thus, OGD solves a stochastic root-finding problem [51], and $\mu_n \to \theta$ almost surely under the conditions

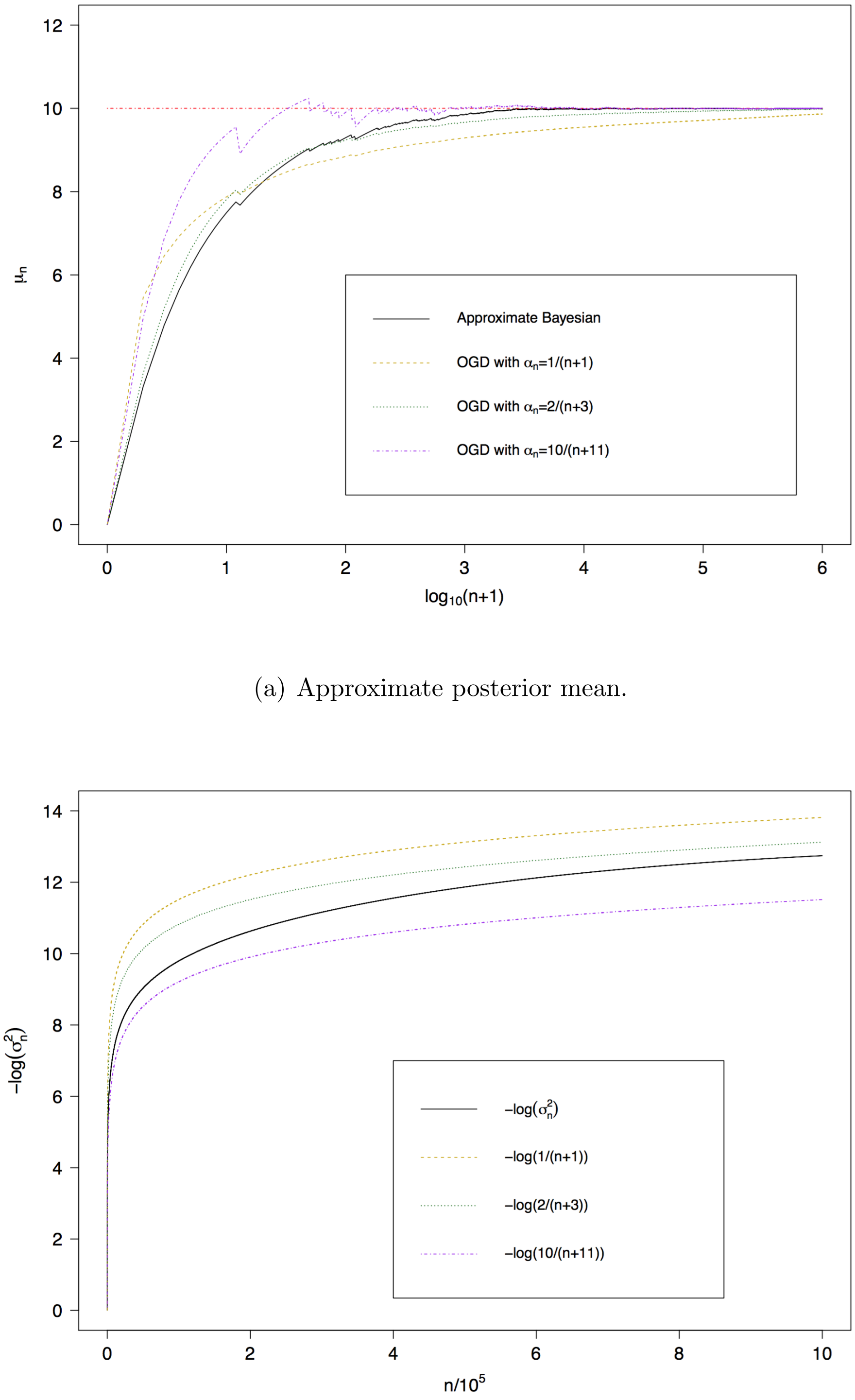
$$\sum_{n=0}^{\infty} \alpha_n = \infty, \qquad \sum_{n=0}^{\infty} \alpha_n^2 < \infty, \tag{2.7}$$

which are usually imposed in SA theory [52]. Thus, the approximate Bayesian update (2.1) can be viewed as a modification of OGD, with the posterior variance σ_n^2 playing two roles: first, it is added to the noise λ^2 in the definition of G_n, and second, it serves instead of the stepsize α_n. Thus, if σ_n^2 satisfies (2.7), and if the difference between the Bayesian and frequentist stochastic gradients is decreasing sufficiently quickly, we may also expect (2.1) to converge.

Section 2.3 will formalize this approach; here, we provide a numerical illustration. Figure 2.1(a) shows the sequence μ_n produced by (2.1)-(2.2) over 10^6 iterations. We set $\lambda^2 = 1.5$, $\mu_0 = 0$, $\sigma_0^2 = 1$, and the sequence $b_n = 8 + 0.000003n$. The true value of the parameter is set to $\theta = 10$. Convergence is observed after just 1000 iterations. We also plot trajectories for three versions of OGD with stepsizes $\alpha_n = \frac{a}{a+n}$ with $a \in \{1, 2, 10\}$. Figure 2.1(b) compares the trajectories of these three stepsizes with that of the approximate posterior variance. We see that OGD exhibits a classic bias/variance tradeoff: higher values of a lead the procedure to find θ more quickly, but induce less stable behavior in the iterate. By contrast, in the Bayesian procedure, σ_n^2 can be viewed as a kind of adaptive stepsize, whose declining behaviour speeds up in later iterations to produce a more stable iterate.

2.3 A General Convergent Stochastic Approximation Algorithm

Suppose that $(R_n)_{n=0}^{\infty}$ is a sequence of real measurable functions mapping $x \in \mathbb{R}^m$ into $\mathbb{R}^m$. Suppose, furthermore, that the equations $R_n(x) = 0$ all have a unique, common root θ that does not depend on n. SA algorithms produce sequences



(a) Approximate posterior mean.

(b) Approximate posterior variance (log-scale).

Figure 2.1: Empirical convergence of the approximate Bayesian estimator.

$(x_n)_{n=0}^{\infty}$ of recursively updated iterates designed to converge to θ in situations where, for each n, a single stochastic (and not necessarily unbiased) observation of $R_n(x_n)$ is available.

We study a general SA algorithm of the form

$$x_{n+1} = x_n - \alpha_n \left(Q_n\left(W_{n+1}, x_n\right) + \beta_n\left(W_{n+1}, x_n, \alpha_n\right)\right), \ n = 0, 1, ... \tag{2.8}$$

where $x_0 \in \mathbb{R}^m$ is an arbitrary m-vector, $(\alpha_n)_{n=0}^{\infty}$ is a positive (deterministic or random) stepsize sequence satisfying (2.7) almost surely, $(W_n)_{n=1}^{\infty}$ is a sequence of random variables representing exogenous information, $(Q_n)_{n=0}^{\infty}$ is a sequence of real measurable functions mapping (w, x) into $\mathbb{R}^m$ and representing the stochastic observations of (R_n), and $(\beta_n)_{n=0}^{\infty}$ is another sequence of real measurable functions representing the "bias" of the observations.

The main difference between (2.8) and the SA procedures in [8] and other references is the introduction of the bias term β_n. In the example given in Section 2.2, the SA update Q_n would be identical to the OGD gradient G_n in (2.6), while the bias β_n would equal the difference between the OGD update and the approximate Bayesian update in (2.1). The posterior variance σ_n^2 serves as the stepsize, which means that the bias β_n should be allowed to depend on the random variable α_n. This dependence does not fit into the standard SA convergence conditions, such as those in Sec. 5.2 of [8], necessitating a new convergence proof. To prove the convergence of a SA-type algorithm, one has to carefully examine the details of SA convergence proofs to determine whether they can be applied in a particular situation. For example, Assumption A.2.2 in Sec. 5.2.1 of [8] appears to allow a bias term similar to β_n, but the convergence proof requires the stepsize α_n to be deterministic. To give some more recent examples, [53] uses a recursive but deterministic stepsize; [54] uses a random stepsize, but requires the bias to be deterministic; and [55] allows a bias term but imposes a specific linear structure on it that does not apply in our setting. The main differences between (2.8) and other provably convergent SA algorithms are that 1) the bias term β_n is random and may depend on the stepsize, and 2) the stepsize itself may be random.

We define

$$\begin{aligned} \mathcal{F}_n &\triangleq \mathcal{B}\left(W_1, ..., W_n, x_1, ..., x_n, \alpha_1, ..., \alpha_n\right), \\ R_n\left(x\right) &\triangleq \mathbb{E}\left(Q_n\left(W_{n+1}, x\right) \,|\, \mathcal{F}_n\right), \end{aligned}$$

where $\mathcal{B}$ denotes the Borel sigma-algebra, and impose several conditions as follows. First, we ensure that (2.8) is searching for a unique root θ.

Assumption 2.3.1. *For any n, the equation $R_n\left(x\right) = 0$ has a unique root θ, which does not depend on n.*

In the example from Section 2.2, the root is the unknown common mean of (Y_n). In the SA algorithm, however, we treat θ as a fixed value (as in frequentist statistics); thus, we develop a non-Bayesian analysis and later apply it to models that were derived from Bayesian arguments.

The second condition is imposed in many standard SA convergence proofs (e.g., [54]), the idea being that the expected value of the stochastic gradient should point the algorithm toward the root.

Assumption 2.3.2. *For $n = 1, 2, \ldots$ and any $\epsilon > 0$,*

$$\inf_{\|x-\theta\|_2^2 > \epsilon, n\in\mathbb{N}} (x-\theta)^T R_n(x) > 0.$$

The third condition bounds the growth of the second moments of Q_n and β_n.

Assumption 2.3.3. *There exist positive constants C_1 and C_2 such that*

$$\begin{aligned} \sup_{n\in\mathbb{N}} \mathbb{E}\left(\|Q_n(W_{n+1},x)\|_2^2 \,|\, \mathcal{F}_n\right) &\leq C_1\left(1+\|x-\theta\|_2^2\right), & (2.9)\\ \sup_{n\in\mathbb{N}} \mathbb{E}\left(\|\beta_n\left(W_{n+1},x,\alpha_n\right)\|_2^2 \,|\, \mathcal{F}_n\right)/\alpha_n^2 &\leq C_2\left(1+\|x-\theta\|_2^2\right) & (2.10)\end{aligned}$$

for all x.

Equation (2.9) controls the amount of noise in the SA update. Equation (2.10) ensures that the bias of the update (recall that, in Section 2.2, we think of this as the difference between frequentist OGD and approximate Bayesian inference) is "small."

In the remainder of this section, we prove the convergence of (2.8); applications of this result will be given in the following section. Theorems 2.3.1 and 2.3.2 essentially state the same result in two ways; the second version uses an explicit projection operator to ensure the boundedness of the iterates, a widely-used approach in SA convergence theory.

Theorem 2.3.1. *Suppose that Assumptions 2.3.1-2.3.3 hold and (α_n) satisfies (2.7) almost surely. Let x_n be defined by (2.8). Then $x_n \to \theta$ almost surely.*

Proof. In all the proofs of this paper, we assume that a suitable set of measure 0 is discarded, so that we do not need to keep repeating the qualification "almost surely". Without loss of generality, let $\theta = 0$. For any $n \in \mathbb{N}$, by (2.8) and (2.9), we have

$$\begin{aligned}
&\mathbb{E}(\|x_{n+1}\|_2^2 \,|\, \mathcal{F}_n) \\
&\quad = \|x_n\|_2^2 - 2\alpha_n x_n^T \mathbb{E}((Q_n(W_{n+1}, x_n) + \beta_n(W_{n+1}, x_n, \alpha_n)) \,|\, \mathcal{F}_n) \\
&\qquad + \alpha_n^2 \mathbb{E}(\|Q_n(W_{n+1}, x_n) + \beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n) \\
&\quad = \|x_n\|_2^2 - 2\alpha_n x_n^T \mathbb{E}(Q_n(W_{n+1}, x_n) \,|\, \mathcal{F}_n) + \alpha_n^2 \mathbb{E}(\|Q_n(W_{n+1}, x_n)\|_2^2 \,|\, \mathcal{F}_n) \\
&\qquad + 2\alpha_n^2 \mathbb{E}((Q_n(W_{n+1}, x_n))^T \beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) + \alpha_n^2 \mathbb{E}(\|\beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n) \\
&\qquad - 2\alpha_n x_n^T \mathbb{E}(\beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) \\
&\quad \leq \|x_n\|_2^2 - 2\alpha_n x_n^T \mathbb{E}(Q_n(W_{n+1}, x_n) \,|\, \mathcal{F}_n) + \alpha_n^2 C_1(1 + \|x_n\|_2^2) \\
&\qquad + 2\alpha_n^2 \mathbb{E}((Q_n(W_{n+1}, x_n))^T \beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) + \alpha_n^2 \mathbb{E}(\|\beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n) \\
&\qquad - 2\alpha_n x_n^T \mathbb{E}(\beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n). \qquad (2.11)
\end{aligned}$$

By (2.10), there exists a positive constant C_3 such that

$$\begin{aligned}
\alpha_n^2 \mathbb{E}(\|\beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n) &\leq \alpha_n^2 C_2(1 + \|x_n\|_2^2)\alpha_n^2 \\
&\leq \alpha_n^2 C_3(1 + \|x_n\|_2^2), \qquad (2.12)
\end{aligned}$$

and, by Hölder's inequality, there exist positive constants C_4 and C_5 such that

$$\begin{aligned}
&-2\alpha_n x_n^T \mathbb{E}(\beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) \\
&\quad = -2\alpha_n \mathbb{E}(x_n^T \beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) \\
&\quad \leq 2\alpha_n \mathbb{E}(\|x_n^T \beta_n(W_{n+1}, x_n, \alpha_n)\|_1 \,|\, \mathcal{F}_n) \\
&\quad \leq 2\alpha_n (\mathbb{E}(\|x_n\|_2^2 \,|\, \mathcal{F}_n))^{\frac{1}{2}} (\mathbb{E}(\|\beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n))^{\frac{1}{2}} \\
&\quad \leq 2\alpha_n \|x_n\|_2 (C_2(1 + \|x_n\|_2^2)\alpha_n^2)^{\frac{1}{2}} \\
&\quad \leq 2\alpha_n^2 \|x_n\|_2 C_4(1 + \|x_n\|_2^2)^{\frac{1}{2}}
\end{aligned}$$

$$
\begin{aligned}
&\le 2\alpha_n^2 C_4(1 + \|x_n\|_2^2) \\
&\le \alpha_n^2 C_5(1 + \|x_n\|_2^2). \qquad (2.13)
\end{aligned}
$$

Again applying Hölder's inequality with (2.9) and (2.12), there exists some positive constant C_6 such that

$$
\begin{aligned}
&2\alpha_n^2 \mathbb{E}((Q_n(W_{n+1}, x_n))^T \beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) \\
&\le 2\alpha_n^2 (\mathbb{E}(\|Q_n(W_{n+1}, x_n)\|_2^2 \,|\, \mathcal{F}_n))^{\frac{1}{2}} (\mathbb{E}(\|\beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n))^{\frac{1}{2}} \\
&\le 2\alpha_n^2 (C_1(1 + \|x_n\|_2^2))^{\frac{1}{2}} (C_3(1 + \|x_n\|_2^2))^{\frac{1}{2}} \\
&\le \alpha_n^2 C_6(1 + \|x_n\|_2^2). \qquad (2.14)
\end{aligned}
$$

Now, we combine (2.11) with (2.12), (2.13) and (2.14), yielding

$$
\begin{aligned}
\mathbb{E}(\|x_{n+1}\|_2^2 \,|\, \mathcal{F}_n) &\le \|x_n\|_2^2 - 2\alpha_n x_n^T \mathbb{E}(Q_n(W_{n+1}, x_n) \,|\, \mathcal{F}_n) \\
&\quad + \alpha_n^2 (C_1 + C_3 + C_5 + C_6)(1 + \|x_n\|_2^2).
\end{aligned}
$$

Letting $\kappa = C_1 + C_3 + C_5 + C_6$, we have

$$
\begin{aligned}
\mathbb{E}(\|x_{n+1}\|_2^2 \,|\, \mathcal{F}_n) &\le \|x_n\|_2^2(1 + \kappa\alpha_n^2) + \kappa\alpha_n^2 - 2\alpha_n x_n^T \mathbb{E}(Q_n(W_{n+1}, x_n) \,|\, \mathcal{F}_n) \\
&= \|x_n\|_2^2(1 + \kappa\alpha_n^2) + \kappa\alpha_n^2 - 2\alpha_n x_n^T R_n(x_n), \qquad (2.15)
\end{aligned}
$$

where $\alpha_n x_n^T R_n(x_n)$ is nonnegative by Assumption 2.3.2.

Then, by Theorem 1 in [56], (2.15) together with (2.7) implies that $\lim_{n\to\infty} \|x_n\|_2^2$ exists and is finite, and that

$$\sum_{n=1}^{\infty} \alpha_n x_n^T R_n(x_n) < \infty$$

almost surely. Hence, by (2.7), since $\sum_{n=1}^{\infty} \alpha_n = \infty$, we have $\liminf_{n\to\infty} x_n^T R_n(x_n) = 0$ almost surely. Then, by Assumptions 2.3.1 and 2.3.2, there must be a subsequence of (x_n) that converges to 0 almost surely. Finally, since $\lim_{n\to\infty} \|x_n\|_2^2$ exists and is finite, we have $x_n \to 0$ almost surely. $\square$

Theorem 2.3.2. *Suppose that Assumptions 2.3.1-2.3.3 hold and (α_n) satisfies (2.7) almost surely. Define*

$$x_{n+1} = \Pi_H \left(x_n - \alpha_n \left(Q_n \left(W_{n+1}, x_n\right) + \beta_n \left(W_{n+1}, x_n, \alpha_n\right)\right)\right), \ n = 0, 1, ... \qquad (2.16)$$

where $H = [-M, M]^m$ with a large enough constant M such that $x_0, \theta \in H$, and $\Pi_H : \mathbb{R}^m \to H$ is a projection operator defined by

$$\left(\Pi_H \left(x\right)\right)^{(i)} = x^{(i)} \cdot 1_{\{|x^{(i)}| \le M\}} + M \cdot 1_{\{x^{(i)} > M\}} - M \cdot 1_{\{x^{(i)} < -M\}},$$

where $x^{(i)}$ denotes the ith element of a vector x. Then, $x_n \to \theta$ almost surely.

Proof. Without loss of generality, let $\theta = 0$. Under Assumptions 2.3.1-2.3.3, similarly as in Theorem 2.3.1, we have

$$\begin{aligned}
\mathbb{E}(\|x_{n+1}\|_2^2 \,|\, \mathcal{F}_n) &= \mathbb{E}(\|\Pi_H \left(x_n - \alpha_n \left(Q_n \left(W_{n+1}, x_n\right) + \beta_n \left(W_{n+1}, x_n, \alpha_n\right)\right)\right)\|_2^2 \,|\, \mathcal{F}_n) \\
&\le \mathbb{E}(\|x_n - \alpha_n \left(Q_n \left(W_{n+1}, x_n\right) + \beta_n \left(W_{n+1}, x_n, \alpha_n\right)\right)\|_2^2 \,|\, \mathcal{F}_n) \\
&\le \|x_n\|_2^2(1 + \kappa\alpha_n^2) + \kappa\alpha_n^2 - 2\alpha_n x_n^T R_n(x_n),
\end{aligned}$$

where κ is some positive constant. Then by Theorem 1 in [56], this together with (2.7) implies that $\lim_{n\to\infty} \|x_n\|_2^2$ exists and is finite, and that

$$\sum_{n=1}^{\infty} \alpha_n x_n^T R_n(x_n) < \infty$$

almost surely. Applying (2.7) and Assumptions 2.3.1 and 2.3.2, the desired result follows. $\square$

We also prove a version of Theorem 2.3.1 using a relaxed version of Assumption 2.3.3. This result will be useful in cases where Assumption 2.3.3 is too strict or difficult to verify.

Assumption 2.3.4. *There exists a positive constant C_1 such that*

$$\begin{aligned} & \sup_{n\in\mathbb{N}} \mathbb{E}\left(\|Q_n(W_{n+1},x)\|_2^2 + \|\beta_n\left(W_{n+1},x,\alpha_n\right)\|_2^2 \,|\, \mathcal{F}_n\right) \\ \leq \; & C_1\left(1+\|x-\theta\|_2^2\right) \end{aligned} \tag{2.17}$$

for all x. Furthermore,

$$\sum_{n=1}^{\infty} \alpha_n \left|(x_n-\theta)^T\, \mathbb{E}\left(\beta_n\left(W_{n+1},x_n,\alpha_n\right)\,|\,\mathcal{F}_n\right)\right| < \infty \tag{2.18}$$

almost surely.

Theorem 2.3.3. *Suppose that Assumptions 2.3.1, 2.3.2 and 2.3.4 hold and (α_n) satisfies (2.7) almost surely. Let x_n be defined by (2.8) or (2.16). Then, $x_n \to \theta$ almost surely.*

Proof. Without loss of generality, let $\theta = 0$. Similarly as in Theorems 2.3.1 and 2.3.2, we have

$$\begin{aligned} \mathbb{E}(\|x_{n+1}\|_2^2\,|\,\mathcal{F}_n) \;\leq\; & \mathbb{E}(\|x_n - \alpha_n\left(Q_n\left(W_{n+1},x_n\right) + \beta_n\left(W_{n+1},x_n,\alpha_n\right)\right)\|_2^2\,|\,\mathcal{F}_n) \\ = \; & \|x_n\|_2^2 - 2\alpha_n x_n^T \mathbb{E}(Q_n(W_{n+1},x_n)\,|\,\mathcal{F}_n) \\ & +\alpha_n^2\mathbb{E}(\|Q_n(W_{n+1},x_n)\|_2^2\,|\,\mathcal{F}_n) \\ & +2\alpha_n^2\mathbb{E}((Q_n(W_{n+1},x_n))^T\beta_n(W_{n+1},x_n,\alpha_n)\,|\,\mathcal{F}_n) \\ & +\alpha_n^2\mathbb{E}(\|\beta_n(W_{n+1},x_n,\alpha_n)\|_2^2\,|\,\mathcal{F}_n) \\ & -2\alpha_n x_n^T\mathbb{E}(\beta_n(W_{n+1},x_n,\alpha_n)\,|\,\mathcal{F}_n) \end{aligned}$$

$$\begin{aligned} &\leq \|x_n\|_2^2 - 2\alpha_n x_n^T \mathbb{E}(Q_n(W_{n+1}, x_n) \,|\, \mathcal{F}_n) \\ &\quad + \alpha_n^2 \mathbb{E}(\|Q_n(W_{n+1}, x_n)\|_2^2 \,|\, \mathcal{F}_n) \\ &\quad + 2\alpha_n^2 \mathbb{E}((Q_n(W_{n+1}, x_n))^T \beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) \\ &\quad + \alpha_n^2 \mathbb{E}(\|\beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n) \\ &\quad + 2\alpha_n \left| x_n^T \mathbb{E}(\beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n) \right|. \end{aligned}$$

Then by (2.17), there exists some positive constant κ such that

$$\begin{aligned} \kappa\alpha_n^2 \left(1 + \|x_n\|_2^2\right) \;\; \geq \;\; & \alpha_n^2 \mathbb{E}(\|Q_n(W_{n+1}, x_n)\|_2^2 \,|\, \mathcal{F}_n) + \alpha_n^2 \mathbb{E}(\|\beta_n(W_{n+1}, x_n, \alpha_n)\|_2^2 \,|\, \mathcal{F}_n) \\ & + 2\alpha_n^2 \mathbb{E}((Q_n(W_{n+1}, x_n))^T \beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n). \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}(\|x_{n+1}\|_2^2 \,|\, \mathcal{F}_n) \;\; \leq \;\; & \|x_n\|_2^2 (1 + \kappa\alpha_n^2) + \kappa\alpha_n^2 + 2\alpha_n \left| x_n^T \mathbb{E}\left(\beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n\right) \right| \\ & - 2\alpha_n x_n^T \mathbb{E}\left(Q_n(W_{n+1}, x_n) \,|\, \mathcal{F}_n\right) \\ = \;\; & \|x_n\|_2^2 (1 + \kappa\alpha_n^2) + \kappa\alpha_n^2 + 2\alpha_n \left| x_n^T \mathbb{E}\left(\beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n\right) \right| \\ & - 2\alpha_n x_n^T R_n(x_n), \end{aligned}$$

where $\alpha_n x_n^T R_n(x_n)$ is nonnegative by Assumptions 2.3.1 and 2.3.2. Applying (2.18) from Assumption 2.3.4 together with (2.7), we obtain

$$\sum_{n=1}^{\infty} \left(\kappa\alpha_n^2 + 2\alpha_n \left| x_n^T \mathbb{E}\left(\beta_n(W_{n+1}, x_n, \alpha_n) \,|\, \mathcal{F}_n\right) \right| \right) < \infty.$$

By Theorem 1 in [56], this together with (2.7) implies that $\lim_{n\to\infty} \|x_n\|_2^2$ exists and is finite, and that

$$\sum_{n=1}^{\infty} \alpha_n x_n^T R_n(x_n) < \infty$$

almost surely. By (2.7) and Assumptions 2.3.1 and 2.3.2, it follows that $x_n \to 0$, as desired. □

2.4 Univariate Applications

We now present four applications of our convergence analysis to recently-studied problems where the goal is to learn a scalar quantity. First, Section 2.4.1 returns to the example in Section 2.2 and proves the consistency of a projected version of (2.1). Sections 2.4.2-2.4.4 give convergence proofs for three computational learning schemes previously developed for applications in competitive online gaming, market design, and posted-price auctions, respectively. While the computational forms of these schemes are taken from prior work, to our knowledge no consistency results were previously available for any of them.

Applying the theory from Section 2.3 is non-trivial and often requires additional technical material. In addition, we provide in Section 2.5.4 an extension of the example from Section 2.4.1 in which both the mean and variance of the underlying distribution are unknown and have to be learned using a provably consistent approximate Bayesian scheme.

2.4.1 Learning an Unknown Mean from Censored Binary Observations

We consider a slight modification of the example from Section 2.2. Suppose that σ_n^2 is updated using (2.2), and the posterior mean μ_n is updated using

$$\mu_{n+1} = \Pi_H\left(\mu_n - \frac{\sigma_n^2}{\sqrt{\lambda^2+\sigma_n^2}}\left(B_{n+1}\frac{\phi\left(p_n\right)}{\Phi\left(p_n\right)} - \left(1-B_{n+1}\right)\frac{\phi\left(p_n\right)}{1-\Phi\left(p_n\right)}\right)\right), \quad (2.19)$$

where $H = [-M, M]$ for large enough M satisfying $|\mu_0| < M$ and $|\theta| < M$. We can write Π_H explicitly as

$$\Pi_H\left(x\right) = x \cdot 1_{\{|x|\leq M\}} + M \cdot 1_{\{x>M\}} - M \cdot 1_{\{x<-M\}}.$$

Thus, (2.19) is a projected version of (2.1) satisfying the conditions of Theorem 2.3.2. The projection operator ensures that σ_n^2 satisfies (2.7) almost surely. Such projections are widely used for similar purposes in the SA literature; see, e.g., Section 4.3 of [8]. Note that the use of a projection requires us to view θ as a fixed (if unknown) value, as in frequentist statistics. Thus, we used Bayesian arguments to construct the learning model, but our convergence analysis (here and throughout the paper) is non-Bayesian and views the model as searching for a fixed root.

We first state a technical lemma, which was proved in [2]. Theorem 2.3.2 will then be applied to establish consistency. We impose the mild regularity condition that (b_n) is bounded, but otherwise allow any arbitrary control policy.

Lemma 2.4.1. *For all pairs* $(x, y) \in \{-\infty \leq x < y \leq \infty\} \setminus \{x = -\infty, y = \infty\}$,

$$\left(\Phi(y) - \Phi(x)\right)\left(y\phi(y) - x\phi(x)\right) + \left(\phi(y) - \phi(x)\right)^2 > 0.$$

Proposition 2.4.1. *Suppose that μ_n and σ_n^2 are updated using (2.19) and (2.2), and suppose that the sequence $(b_n)_{n=0}^{\infty}$ is bounded. Then, $\mu_n \to \theta$ almost surely.*

Proof. Let p_n be as in (2.3), and define

$$\begin{aligned} q_n &= \frac{b_n - \mu_n}{\lambda}, \\ Q_n(B_{n+1}, b_n, \mu_n) &= B_{n+1}\frac{1}{\lambda}\frac{\phi(q_n)}{\Phi(q_n)} - (1 - B_{n+1})\frac{1}{\lambda}\frac{\phi(q_n)}{1-\Phi(q_n)}, \\ \beta_n(B_{n+1}, b_n, \mu_n, \sigma_n^2) &= B_{n+1}\left(\frac{1}{\sqrt{\lambda^2+\sigma_n^2}}\frac{\phi(p_n)}{\Phi(p_n)} - \frac{1}{\lambda}\frac{\phi(q_n)}{\Phi(q_n)}\right) \\ &\quad -(1-B_{n+1})\left(\frac{1}{\sqrt{\lambda^2+\sigma_n^2}}\frac{\phi(p_n)}{1-\Phi(p_n)} - \frac{1}{\lambda}\frac{\phi(q_n)}{1-\Phi(q_n)}\right). \end{aligned}$$

Let us rewrite (2.2) as

$$\begin{aligned} \frac{1}{\sigma_{n+1}^2} &= \frac{1}{\sigma_n^2\left(1 - \frac{\sigma_n^2}{\lambda^2+\sigma_n^2}\left(B_{n+1}\frac{p_n\phi(p_n)\Phi(p_n)+\phi^2(p_n)}{\Phi^2(p_n)} + (1-B_{n+1})\frac{\phi^2(p_n)-p_n\phi(p_n)(1-\Phi(p_n))}{(1-\Phi(p_n))^2}\right)\right)} \\ &= \frac{1}{\sigma_n^2} + \frac{\frac{1}{\lambda^2+\sigma_n^2}\left(B_{n+1}\frac{p_n\phi(p_n)\Phi(p_n)+\phi^2(p_n)}{\Phi^2(p_n)} + (1-B_{n+1})\frac{\phi^2(p_n)-p_n\phi(p_n)(1-\Phi(p_n))}{(1-\Phi(p_n))^2}\right)}{1 - \frac{\sigma_n^2}{\lambda^2+\sigma_n^2}\left(B_{n+1}\frac{p_n\phi(p_n)\Phi(p_n)+\phi^2(p_n)}{\Phi^2(p_n)} + (1-B_{n+1})\frac{\phi^2(p_n)-p_n\phi(p_n)(1-\Phi(p_n))}{(1-\Phi(p_n))^2}\right)}. \end{aligned}$$

By Lemma 2.4.1, for any $x \in \mathbb{R}$, we have

$$0 < \frac{x\phi(x)\Phi(x)+\phi^2(x)}{\Phi^2(x)} \leq 1, \qquad 0 < \frac{\phi^2(x) - x\phi(x)(1-\Phi(x))}{(1-\Phi(x))^2} \leq 1, \tag{2.20}$$

whence it follows that the sequence (σ_n^2) is positive and monotone decreasing. Since the sequence $(b_n)_{n=0}^{\infty}$ is bounded, and (μ_n) is constrained to a closed and bounded interval by Π_H, it follows that the sequence (p_n) is also constrained to a closed and bounded interval of $\mathbb{R}$. Then, by the continuity of $\frac{x\phi(x)\Phi(x)+\phi^2(x)}{\Phi^2(x)}$ and $\frac{\phi^2(x)-x\phi(x)(1-\Phi(x))}{(1-\Phi(x))^2}$, there exist constants $\gamma_*, \gamma^* > 0$ such that, for all $n \in \mathbb{N}$,

$$\gamma_* \leq \frac{\frac{1}{\lambda^2+\sigma_n^2}\left(B_{n+1}\frac{p_n\phi(p_n)\Phi(p_n)+\phi^2(p_n)}{\Phi^2(p_n)} + (1-B_{n+1})\frac{\phi^2(p_n)-p_n\phi(p_n)(1-\Phi(p_n))}{(1-\Phi(p_n))^2}\right)}{1 - \frac{\sigma_n^2}{\lambda^2+\sigma_n^2}\left(B_{n+1}\frac{p_n\phi(p_n)\Phi(p_n)+\phi^2(p_n)}{\Phi^2(p_n)} + (1-B_{n+1})\frac{\phi^2(p_n)-p_n\phi(p_n)(1-\Phi(p_n))}{(1-\Phi(p_n))^2}\right)} \leq \gamma^*.$$

Therefore,

$$\frac{1}{\sigma_0^2} + n\gamma_* \leq \frac{1}{\sigma_n^2} \leq \frac{1}{\sigma_0^2} + n\gamma^*,$$

whence

$$\sum_{n=1}^{\infty} \sigma_n^2 = \infty, \qquad \sum_{n=1}^{\infty} \sigma_n^4 < \infty,$$

thus verifying (2.7).

Now, define

$$\mathcal{F}_n \triangleq \mathcal{B}(B_1, ..., B_n, \mu_0, ..., \mu_n, \sigma_0^2, ..., \sigma_n^2, b_0, ..., b_n).$$

Recalling that $(Y_n)_{n=1}^{\infty}$ is a sequence of i.i.d. samples from the common distribution $\mathcal{N}(\theta, \lambda^2)$, we calculate

$$\begin{aligned} R_n(x) &\triangleq \mathbb{E}(Q_n(B_{n+1}, b_n, x) \,|\, \mathcal{F}_n) \\ &= \Phi\left(\frac{b_n - \theta}{\lambda}\right) \frac{1}{\lambda} \frac{\phi\left(\frac{b_n - x}{\lambda}\right)}{\Phi\left(\frac{b_n - x}{\lambda}\right)} - \left(1 - \Phi\left(\frac{b_n - \theta}{\lambda}\right)\right) \frac{1}{\lambda} \frac{\phi\left(\frac{b_n - x}{\lambda}\right)}{1 - \Phi\left(\frac{b_n - x}{\lambda}\right)}. \end{aligned}$$

It is easy to see that $R_n(x) = 0$ if and only if $x = \theta$, thus verifying Assumption 2.3.1. Since $(b_n)_{n=0}^{\infty}$ is bounded, it is straightforward to verify Assumption 2.3.2.

Observe that $-\frac{\phi(x)}{\Phi(x)}$ and $\frac{\phi(x)}{1-\Phi(x)}$ are continuously differentiable, and their first derivatives always take values in $(0, 1]$ by (2.20). Since (b_n) is bounded, it follows from the mean value theorem that there exist positive constants C_1, C_2 satisfying

$$\begin{aligned} \sup_{n \in \mathbb{N}} |Q_n(B_{n+1}, b_n, x)| &\leq C_1(1 + |x - \theta|), \\ \sup_{n \in \mathbb{N}} |\beta_n(B_{n+1}, b_n, x, \sigma_n^2)| / \sigma_n^2 &\leq C_2(1 + |x - \theta|). \end{aligned}$$

Consequently, there also exist positive constants C_3, C_4 such that

$$\begin{aligned}\sup_{n\in\mathbb{N}} \mathbb{E}(Q_n^2(B_{n+1}, b_n, x) \,|\, \mathcal{F}_n) &\leq C_3(1+(x-\theta)^2),\\ \sup_{n\in\mathbb{N}} \mathbb{E}(\beta_n^2(B_{n+1}, b_n, x, \sigma_n^2) \,|\, \mathcal{F}_n)/\sigma_n^4 &\leq C_4(1+(x-\theta)^2),\end{aligned}$$

thus verifying Assumption 2.3.3. The desired result then follows by Theorem 2.3.2.

□

An interesting question is whether it is possible to develop a provably consistent approximate Bayesian learning scheme for the case where both θ and λ have to be simultaneously learned from censored binary observations. In brief, the answer is yes; this case is treated in Section 2.5.4.

2.4.2 Learning Player Skills in Competitive Online Gaming

References [57] and [4] describe an approximate Bayesian learning model that was implemented in Microsoft's Xbox Live online gaming service for inferring player skills from the outcomes of competitive events. In this application, large numbers of players log on to the service and ask to play a game; the system then seeks to match players whose skill levels are likely to be similar, in order to promote fair play and create a more rewarding experience.

We give a streamlined summary of the model, assuming without loss of generality that there are only two players, and prove a new consistency result. Let $\theta^{(i)}$ represent the "skill" of player $i \in \{1,2\}$. Denote by $Y_n^{(i)}$ the "performance" of player i in the nth game, with the assumption that

$$Y_{n+1}^{(i)} \sim \mathcal{N}\left(\theta^{(i)}, \lambda^2\right),$$

where the variance λ^2 is known. We use the Bayesian prior $\theta^{(i)} \sim \mathcal{N}\left(\mu_0^{(i)}, \left(\sigma_0^{(i)}\right)^2\right)$ for $i \in \{1,2\}$ and assume that all skills and performance values are mutually independent. The game master cannot observe $Y_n^{(i)}$ directly, but rather must infer player skills from the binary outcome of the game, denoted by

$$B_{n+1}^{(i)} = 1_{\left\{Y_{n+1}^{(i)} < Y_{n+1}^{(j)}\right\}},$$

where j denotes the index of the opponent. In words, if player j wins the match against i, we interpret this as $Y_{n+1}^{(j)} > Y_{n+1}^{(i)}$. It is assumed that no game can end in a draw.

Reference [4] used moment-matching to derive the approximate Bayesian updating equations

$$\begin{aligned}
\mu_{n+1}^{(i)} &= \mu_n^{(i)} - \left(\sigma_n^{(i)}\right)^2 \left(\frac{B_{n+1}^{(i)}}{\sqrt{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2}} v \left(\frac{\mu_n^{(j)} - \mu_n^{(i)}}{\sqrt{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2}} \right) \right. \\
&\quad \left. - \frac{\left(1 - B_{n+1}^{(i)}\right)}{\sqrt{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2}} v \left(\frac{\mu_n^{(i)} - \mu_n^{(j)}}{\sqrt{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2}} \right) \right), \qquad (2.21)
\end{aligned}$$

$$\begin{aligned}
\left(\sigma_{n+1}^{(i)}\right)^2 &= \left(\sigma_n^{(i)}\right)^2 \left(1 - \frac{B_{n+1}^{(i)} \left(\sigma_n^{(i)}\right)^2}{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2} w \left(\frac{\mu_n^{(j)} - \mu_n^{(i)}}{\sqrt{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2}} \right) \right. \\
&\quad \left. - \frac{\left(1 - B_{n+1}^{(i)}\right) \left(\sigma_n^{(i)}\right)^2}{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2} w \left(\frac{\mu_n^{(i)} - \mu_n^{(j)}}{\sqrt{\left(\sigma_n^{(i)}\right)^2 + \left(\sigma_n^{(j)}\right)^2 + 2\lambda^2}} \right) \right), \qquad (2.22)
\end{aligned}$$

where

$$\begin{aligned}
v(x) &= \frac{\phi(x)}{\Phi(x)}, \\
w(x) &= v(x)(v(x) + x).
\end{aligned}$$

As in Section 2.4.1, we can replace (2.21) by a projected version where $\mu_n^{(i)}$ is constrained to be within a suitably large interval.

Define

$$\begin{aligned} d_n &\triangleq \mu_n^{(1)} - \mu_n^{(2)}, \\ \delta &\triangleq \theta^{(1)} - \theta^{(2)}. \end{aligned}$$

In this setting, the observable information is insufficient to learn $\theta^{(i)}$ exactly, but the quantity of primary interest to the game master is the difference δ, as this is what is used to evaluate the fairness of a match-up. We prove that d_n is a consistent estimator of δ.

Proposition 2.4.2. *Suppose that $\mu_n^{(i)}$ is updated using a projected version of (2.21), while $\sigma_n^{(i)}$ is updated using (2.22). Then, $d_n \to \delta$ almost surely.*

Proof. Define

$$\begin{aligned} \sigma_n^2 &= \left(\sigma_n^{(1)}\right)^2 + \left(\sigma_n^{(2)}\right)^2, \\ Q_n\left(B_{n+1}^{(1)}, d_n\right) &= B_{n+1}^{(1)} \frac{1}{\sqrt{2\lambda^2}} v\left(\frac{-d_n}{\sqrt{2\lambda^2}}\right) - \left(1 - B_{n+1}^{(1)}\right) \frac{1}{\sqrt{2\lambda^2}} v\left(\frac{d_n}{\sqrt{2\lambda^2}}\right), \\ \beta_n\left(B_{n+1}^{(1)}, d_n, \sigma_n^2\right) &= B_{n+1}^{(1)} \left(\frac{1}{\sqrt{\sigma_n^2 + 2\lambda^2}} v\left(\frac{-d_n}{\sqrt{\sigma_n^2 + 2\lambda^2}}\right) - \frac{1}{\sqrt{2\lambda^2}} v\left(\frac{-d_n}{\sqrt{2\lambda^2}}\right)\right) \\ &\quad - \left(1 - B_{n+1}^{(1)}\right) \left(\frac{1}{\sqrt{\sigma_n^2 + 2\lambda^2}} v\left(\frac{d_n}{\sqrt{\sigma_n^2 + 2\lambda^2}}\right) - \frac{1}{\sqrt{2\lambda^2}} v\left(\frac{d_n}{\sqrt{2\lambda^2}}\right)\right). \end{aligned}$$

Then, with some algebra we can derive

$$d_{n+1} = d_n - \sigma_n^2 \left(Q_n\left(B_{n+1}^{(1)}, d_n\right) + \beta_n\left(B_{n+1}^{(1)}, d_n, \sigma_n^2\right)\right)$$

from (2.21) and (2.22).

If $\mu_n^{(i)}$ is updated using a projected version of (2.21), the sequence (d_n) will also be constrained to some closed and bounded interval. Then, from (2.22) it follows that

$$\frac{1}{\left(\sigma_{n+1}^{(1)}\right)^2} = \frac{1}{\left(\sigma_n^{(1)}\right)^2} + \frac{\frac{1}{\sigma_n^2+2\lambda^2}\left(B_{n+1}^{(1)} w\left(\frac{-d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right) + \left(1-B_{n+1}^{(1)}\right) w\left(\frac{d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right)\right)}{1 - \frac{\left(\sigma_n^{(1)}\right)^2}{\sigma_n^2+2\lambda^2}\left(B_{n+1}^{(1)} w\left(\frac{-d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right) + \left(1-B_{n+1}^{(1)}\right) w\left(\frac{d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right)\right)},$$

and there exist two positive constants γ_*, γ^* such that, for all $n \in \mathbb{N}$,

$$\gamma_* \leq \frac{\frac{1}{\sigma_n^2+2\lambda^2}\left(B_{n+1}^{(1)} w\left(\frac{-d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right) + \left(1-B_{n+1}^{(1)}\right) w\left(\frac{d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right)\right)}{1 - \frac{\left(\sigma_n^{(1)}\right)^2}{\sigma_n^2+2\lambda^2}\left(B_{n+1}^{(1)} w\left(\frac{-d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right) + \left(1-B_{n+1}^{(1)}\right) w\left(\frac{d_n}{\sqrt{\sigma_n^2+2\lambda^2}}\right)\right)} \leq \gamma^*.$$

Consequently,

$$\frac{1}{\left(\sigma_0^{(1)}\right)^2} + n\gamma_* \leq \frac{1}{\left(\sigma_n^{(1)}\right)^2} \leq \frac{1}{\left(\sigma_0^{(1)}\right)^2} + n\gamma^*,$$

whence

$$\sum_{n=1}^{\infty} \left(\sigma_n^{(1)}\right)^2 = \infty, \qquad \sum_{n=1}^{\infty} \left(\sigma_n^{(1)}\right)^4 < \infty.$$

Similar arguments apply to $\left(\left(\sigma_n^{(2)}\right)^2\right)$, whence

$$\sum_{n=1}^{\infty} \sigma_n^2 = \infty, \qquad \sum_{n=1}^{\infty} \sigma_n^4 < \infty,$$

thus verifying (2.7).

Now define

$$\begin{aligned}
\mathcal{F}_n &\triangleq \mathcal{B}\Bigg(B_1^{(1)}, ..., B_n^{(1)}, \mu_0^{(1)}, ..., \mu_n^{(1)}, \left(\sigma_0^{(1)}\right)^2, ..., \left(\sigma_n^{(1)}\right)^2, \\
&\qquad \mu_0^{(2)}, ..., \mu_n^{(2)}, \left(\sigma_0^{(2)}\right)^2, ..., \left(\sigma_n^{(2)}\right)^2\Bigg), \\
R_n(x) &\triangleq \mathbb{E}\left(Q_n\left(B_{n+1}^{(1)}, x\right) \mid \mathcal{F}_n\right) \\
&= \Phi\left(\frac{-\delta}{\sqrt{2\lambda^2}}\right)\frac{1}{\sqrt{2\lambda^2}} v\left(\frac{-x}{\sqrt{2\lambda^2}}\right) - \left(1 - \Phi\left(\frac{-\delta}{\sqrt{2\lambda^2}}\right)\right)\frac{1}{\sqrt{2\lambda^2}} v\left(\frac{x}{\sqrt{2\lambda^2}}\right) \\
&= \Phi\left(\frac{-\delta}{\sqrt{2\lambda^2}}\right)\frac{1}{\sqrt{2\lambda^2}} v\left(\frac{-x}{\sqrt{2\lambda^2}}\right) - \Phi\left(\frac{\delta}{\sqrt{2\lambda^2}}\right)\frac{1}{\sqrt{2\lambda^2}} v\left(\frac{x}{\sqrt{2\lambda^2}}\right).
\end{aligned}$$

It is clear that $R_n(x) = 0$ if and only if $x = \delta$, thus verifying Assumption (2.3.1). Assumption 2.3.2 is straightforward to verify.

Since $v(x) = \frac{\phi(x)}{\Phi(x)}$ is continuously differentiable with $-(v')$ taking values in $(0, 1]$ (as in the proof of Proposition 2.4.1), it follows by the mean value theorem that there exist positive constants C_1, C_2 satisfying

$$\begin{aligned} \sup_{n \in \mathbb{N}} \left| Q_n \left(B_{n+1}^{(1)}, x \right) \right| &\leq C_1(1 + |x - \delta|), \\ \sup_{n \in \mathbb{N}} \left| \beta_n \left(B_{n+1}^{(1)}, x, \sigma_n^2 \right) \right| / \sigma_n^2 &\leq C_2(1 + |x - \delta|). \end{aligned}$$

Consequently, there also exist positive constants C_3, C_4 satisfying

$$\begin{aligned} \sup_{n \in \mathbb{N}} \mathbb{E} \left(Q_n^2 \left(B_{n+1}^{(1)}, x \right) \mid \mathcal{F}_n \right) &\leq C_3(1 + (x - \delta)^2), \\ \sup_{n \in \mathbb{N}} \mathbb{E} \left(\beta_n^2 \left(B_{n+1}^{(1)}, x, \sigma_n^2 \right) \mid \mathcal{F}_n \right) / \sigma_n^4 &\leq C_4(1 + (x - \delta)^2), \end{aligned}$$

whence Assumption 2.3.3 is verified. The desired result then follows by Theorem 2.3.2 □

2.4.3 Learning the Market Value of an Asset

Reference [2] presents a model by which a market-maker may learn the unknown value θ of an asset after a market shock (see also [58] for a case application). The market-maker interacts with a sequence of traders, each of whom may buy or sell one unit of the asset. The sequence $(Y_n)_{n=1}^{\infty}$ denotes the traders' perceptions of the unknown value, which are assumed to be i.i.d. $\mathcal{N}(\theta, \lambda^2)$ random variables with λ^2 known.

The prior $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$ reflects the market-maker's initial belief. Let $(b_n)_{n=0}^{\infty}$ and $(a_n)_{n=0}^{\infty}$ denote sequences of fixed bid and ask prices. If $Y_{n+1} < a_n$, the $(n+1)$st trader buys one unit of the asset from the market-maker; if $a_n \leq Y_{n+1} \leq b_n$, the trader does not make any transaction; and, if $Y_{n+1} > b_n$, the trader sells one unit of the asset to the market-maker. Let

$$B^{(1)}_{n+1} = 1_{\{Y_{n+1} < a_n\}}, \qquad B^{(2)}_{n+1} = 1_{\{a_n \leq Y_{n+1} \leq b_n\}}, \qquad B^{(3)}_{n+1} = 1_{\{Y_{n+1} > b_n\}}$$

represent the $(n+1)$st trader's actions (the three binary variables must sum to 1).

[2] proposed an approximate Bayesian learning model for this problem. We define

$$p_n = \frac{a_n - \mu_n}{\sqrt{\lambda^2 + \sigma_n^2}}, \qquad q_n = \frac{b_n - \mu_n}{\sqrt{\lambda^2 + \sigma_n^2}},$$

and update our beliefs recursively using

$$\begin{aligned}
\mu_{n+1} &= \mu_n - \sigma_n^2 \Bigg(B^{(1)}_{n+1} \frac{1}{\sqrt{\lambda^2 + \sigma_n^2}} \frac{\phi(p_n)}{\Phi(p_n)} + B^{(2)}_{n+1} \frac{1}{\sqrt{\lambda^2 + \sigma_n^2}} \frac{\phi(q_n) - \phi(p_n)}{\Phi(q_n) - \Phi(p_n)} \\
&\quad - B^{(3)}_{n+1} \frac{1}{\sqrt{\lambda^2 + \sigma_n^2}} \frac{\phi(q_n)}{1 - \Phi(q_n)} \Bigg), \qquad (2.23) \\
\sigma^2_{n+1} &= \sigma_n^2 \Bigg(1 - B^{(1)}_{n+1} \frac{\sigma_n^2}{\lambda^2 + \sigma_n^2} \frac{p_n \phi(p_n) \Phi(p_n) + \phi^2(p_n)}{\Phi^2(p_n)} \\
&\quad - B^{(2)}_{n+1} \frac{\sigma_n^2}{\lambda^2 + \sigma_n^2} \frac{\left(q_n \phi(q_n) - p_n \phi(p_n)\right)\left(\Phi(q_n) - \Phi(p_n)\right) + \left(\phi(q_n) - \phi(p_n)\right)^2}{\left(\Phi(q_n) - \Phi(p_n)\right)^2} \\
&\quad - B^{(3)}_{n+1} \frac{\sigma_n^2}{\lambda^2 + \sigma_n^2} \frac{\phi^2(q_n) - q_n \phi(q_n)\left(1 - \Phi(q_n)\right)}{\left(1 - \Phi(q_n)\right)^2} \Bigg). \qquad (2.24)
\end{aligned}$$

As in previous examples, we can use a projected version of (2.23). Consistency of the estimator μ_n then follows.

Proposition 2.4.3. *Suppose that $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=0}^{\infty}$ are bounded, and that μ_n is updated using a projected version of (2.23), while σ_n^2 is updated using (2.24). Then, $\mu_n \to \theta$ almost surely.*

Proof. Define

$$
\begin{aligned}
B_{n+1} &\triangleq \left(B^{(1)}_{n+1}, B^{(2)}_{n+1}, B^{(3)}_{n+1}\right), \\
r_n &= \frac{a_n - \mu_n}{\lambda}, \\
s_n &= \frac{b_n - \mu_n}{\lambda}, \\
Q_n(B_{n+1}, a_n, b_n, \mu_n) &= B^{(1)}_{n+1} \frac{1}{\lambda} \frac{\phi(r_n)}{\Phi(r_n)} + B^{(2)}_{n+1} \frac{1}{\lambda} \frac{\phi(s_n) - \phi(r_n)}{\Phi(s_n) - \Phi(r_n)} - B^{(3)}_{n+1} \frac{1}{\lambda} \frac{\phi(s_n)}{1 - \Phi(s_n)}, \\
\beta_n(B_{n+1}, a_n, b_n, \mu_n, \sigma^2_n) &= B^{(1)}_{n+1} \left(\frac{1}{\sqrt{\lambda^2 + \sigma^2_n}} \frac{\phi(p_n)}{\Phi(p_n)} - \frac{1}{\lambda} \frac{\phi(r_n)}{\Phi(r_n)} \right) \\
&\quad + B^{(2)}_{n+1} \left(\frac{1}{\sqrt{\lambda^2 + \sigma^2_n}} \frac{\phi(q_n) - \phi(p_n)}{\Phi(q_n) - \Phi(p_n)} - \frac{1}{\lambda} \frac{\phi(s_n) - \phi(r_n)}{\Phi(s_n) - \Phi(r_n)} \right) \\
&\quad - B^{(3)}_{n+1} \left(\frac{1}{\sqrt{\lambda^2 + \sigma^2_n}} \frac{\phi(q_n)}{1 - \Phi(q_n)} - \frac{1}{\lambda} \frac{\phi(s_n)}{1 - \Phi(s_n)} \right).
\end{aligned}
$$

Since $(a_n)^\infty_{n=0}$ and $(b_n)^\infty_{n=0}$ are bounded and (μ_n) is constrained in some finite closed interval of $\mathbb{R}$, it can be shown (similarly to the proofs of Propositions 2.4.1 and 2.4.2) that there exist two positive constants γ_*, γ^* such that, for all $n \in \mathbb{N}$,

$$
\frac{1}{\sigma^2_n} + \gamma_* \leq \frac{1}{\sigma^2_{n+1}} \leq \frac{1}{\sigma^2_n} + \gamma^*,
$$

whence

$$
\frac{1}{\sigma^2_0} + n\gamma_* \leq \frac{1}{\sigma^2_n} \leq \frac{1}{\sigma^2_0} + n\gamma^*,
$$

and

$$
\sum_{n=1}^{\infty} \sigma^2_n = \infty, \qquad \sum_{n=1}^{\infty} \sigma^4_n < \infty,
$$

thus verifying (2.7).

Now, define

$$
\mathcal{F}_n \triangleq \mathcal{B}\left(B_1, ..., B_n, \mu_0, ..., \mu_n, \sigma^2_0, ..., \sigma^2_n, a_0, ..., a_n, b_0, ..., b_n\right),
$$

$$
\begin{aligned}
R_n(x) &\triangleq \mathbb{E}(Q_n(B_{n+1}, a_n, b_n, x) \,|\, \mathcal{F}_n) \\
&= \Phi\left(\frac{a_n - \theta}{\lambda}\right) \frac{1}{\lambda} \frac{\phi\left(\frac{a_n - x}{\lambda}\right)}{\Phi\left(\frac{a_n - x}{\lambda}\right)} \\
&\quad + \left(\Phi\left(\frac{b_n - \theta}{\lambda}\right) - \Phi\left(\frac{a_n - \theta}{\lambda}\right)\right) \frac{1}{\lambda} \frac{\phi\left(\frac{b_n - x}{\lambda}\right) - \phi\left(\frac{a_n - x}{\lambda}\right)}{\Phi\left(\frac{b_n - x}{\lambda}\right) - \Phi\left(\frac{a_n - x}{\lambda}\right)} \\
&\quad - \left(1 - \Phi\left(\frac{b_n - \theta}{\lambda}\right)\right) \frac{1}{\lambda} \frac{\phi\left(\frac{b_n - x}{\lambda}\right)}{1 - \Phi\left(\frac{b_n - x}{\lambda}\right)}.
\end{aligned}
$$

It is easy to see that $R_n(x) = 0$ if and only if $x = \theta$, verifying Assumption 2.3.1. The boundedness of $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=0}^{\infty}$ straightforwardly implies Assumption 2.3.2.

Observe that, for any n, $\frac{\phi\left(\frac{a_n - x}{\lambda}\right)}{\Phi\left(\frac{a_n - x}{\lambda}\right)}$, $\frac{\phi\left(\frac{b_n - x}{\lambda}\right) - \phi\left(\frac{a_n - x}{\lambda}\right)}{\Phi\left(\frac{b_n - x}{\lambda}\right) - \Phi\left(\frac{a_n - x}{\lambda}\right)}$ and $\frac{\phi\left(\frac{b_n - x}{\lambda}\right)}{1 - \Phi\left(\frac{b_n - x}{\lambda}\right)}$ are continuously differentiable with first derivatives taking values in $\left(0, \frac{1}{\lambda}\right]$. The boundedness of (a_n) and (b_n), together with the mean value theorem, implies the existence of positive constants C_1, C_2 satisfying

$$
\begin{aligned}
\sup_{n \in \mathbb{N}} |Q_n(B_{n+1}, a_n, b_n, x)| &\leq C_1(1 + |x - \theta|), \\
\sup_{n \in \mathbb{N}} |\beta_n(B_{n+1}, a_n, b_n, x, \sigma_n^2)| / \sigma_n^2 &\leq C_2(1 + |x - \theta|).
\end{aligned}
$$

Consequently, there also exist positive constants C_3, C_4 satisfying

$$
\begin{aligned}
\sup_{n \in \mathbb{N}} \mathbb{E}(Q_n^2(B_{n+1}, a_n, b_n, x) \,|\, \mathcal{F}_n) &\leq C_3(1 + (x - \theta)^2), \\
\sup_{n \in \mathbb{N}} \mathbb{E}(\beta_n^2(B_{n+1}, a_n, b_n, x, \sigma_n^2) \,|\, \mathcal{F}_n) / \sigma_n^4 &\leq C_4(1 + (x - \theta)^2),
\end{aligned}
$$

whence Assumption 2.3.3 is verified. The desired result then follows by Theorem 2.3.2. □

2.4.4 Learning Buyer Valuations in Online Posted-Price Auctions

Reference [1] describes the following model for dynamic pricing in online digital goods auctions. The sequence $(Y_n)_{n=1}^{\infty}$ represents independent buyer valuations of a product. The seller sets a sequence $(q_n)_{n=0}^{\infty}$ of prices, and the nth price is accepted if $Y_{n+1} > q_n$, i.e., the value of the item to the buyer exceeds the price. Otherwise, the price is rejected and no revenue is earned. The term "demand curve" refers to the acceptance probability $\rho(q) = P(Y_{n+1} > q)$ viewed as a function of the price q; two valuations are i.i.d. given the same price. In revenue management, a commonly-used model is a linear demand curve [59]

$$\rho(q) = 1 - \gamma q.$$

The slope γ is unknown and must be learned. We suppose that the prices are normalized, i.e., $q_n \in [0,1]$ for all n, and can then assume that $\gamma \in (0,1)$. A natural choice of prior in this setting is the beta distribution $\gamma \sim Beta(a_0, b_0)$. Let I_{n+1} be a binary variable that equals 1 if the $(n+1)$st buyer accepts the price q_n, and zero otherwise.

The following learning mechanism, based on moment-matching, was proposed by [1]. Define

$$\begin{aligned}
\mu_n &= \frac{a_n}{a_n + b_n},\\
\tau_n &= a_n + b_n,\\
A_n &= \mu_n(1-\mu_n),\\
B_n &= 2(1-q_n) + (3 - 2q_n - 2\mu_n q_n + \mu_n)\tau_n + (1-\mu_n q_n)^2 \tau_n^2,\\
C_n &= q_n \tau_n \mu_n (1-q_n)(1+\mu_n \tau_n),\\
D_n &= q_n \tau_n (1-\mu_n)(1+(1-\mu_n)\tau_n),
\end{aligned}$$

and apply the updating equations

$$\begin{aligned} a_{n+1} &= a_n - I_{n+1}\frac{C_n}{B_n} + (1 - I_{n+1}), & (2.25) \\ b_{n+1} &= b_n + I_{n+1}\frac{D_n}{B_n}, & (2.26) \\ \tau_{n+1} &= a_{n+1} + b_{n+1}, & (2.27) \\ \mu_{n+1} &= \mu_n - \frac{1}{\tau_n + 1}\left(I_{n+1}\frac{A_n q_n}{1 - q_n\mu_n} - (1 - I_{n+1})\frac{A_n}{\mu_n}\right). & (2.28) \end{aligned}$$

Again, we make a slight modification to (2.28) by using a projection operator to ensure that $\inf_n \mu_n > 0$ and $\sup_n \mu_n < 1$. Consistency can then be obtained.

Proposition 2.4.4. *Suppose that $\inf_n q_n > 0$ and $\sup_n q_n < 1$, and that μ_n is updated using a suitable projected version of (2.28), while (2.25)-(2.27) are used to update a_n, b_n and τ_n. Then, $\mu_n \to \gamma$ a.s.*

Proof. First, notice that Theorem 2.3.2 still holds if we replace $\mathbb{R}^m$ by the interval $(0, 1)$ with H chosen to be a large enough closed interval in $(0, 1)$ such that $\mu_0, \gamma \in H$. Since we use a projected version of (2.28), the sequence (μ_n) is constrained in some interval $[\mu_*, \mu^*]$, where $0 < \mu_* < \mu^* < 1$. Let

$$\begin{aligned} \eta_n &= a_n(1 - \mu^*)^2 + b_n, \\ E_n &= \frac{\eta_n + 1}{\tau_n + 1}, \\ Q_n(I_{n+1}, q_n, A_n, E_n, \mu_n) &= I_{n+1}\frac{A_n E_n q_n}{1 - q_n\mu_n} - (1 - I_{n+1})\frac{A_n E_n}{\mu_n}. \end{aligned}$$

Then, the updating equation (2.28) can be rewritten as

$$\mu_{n+1} = \mu_n - \frac{1}{\eta_n + 1} Q_n(I_{n+1}, q_n, A_n, E_n, \mu_n),$$

which is an SA algorithm with no bias term and $\frac{1}{\eta_n+1}$ as the step size. We observe that

$$\begin{aligned}\eta_{n+1} - \eta_n &= (a_{n+1} - a_n)(1-\mu^*)^2 + (b_{n+1} - b_n)\\ &= \left(-I_{n+1}\frac{C_n}{B_n} + (1 - I_{n+1})\right)(1-\mu^*)^2 + I_{n+1}\frac{D_n}{B_n}\\ &= I_{n+1}\frac{D_n - C_n(1-\mu^*)^2}{B_n} + (1 - I_{n+1})(1-\mu^*)^2.\end{aligned}$$

It is obvious that $(1-\mu^*)^2 > 0$ and $B_n, C_n, D_n > 0$. Then, since $\mu_n \in [\mu_*, \mu^*]$ and $0 < \inf_n q_n \le \sup_n q_n < 1$, we have $D_n - C_n(1-\mu^*)^2 > 0$ and $\frac{D_n}{B_n} \le 1$. By the continuity of B_n, C_n and D_n, there exist positive constants η_* and η^* such that, for any $n \in \mathbb{N}$,

$$\eta_* \le \eta_{n+1} - \eta_n \le \eta^*,$$

whence it follows that the sequence (η_n) is monotone increasing and

$$\sum_{n=1}^{\infty} \frac{1}{\eta_n + 1} = \infty, \qquad \sum_{n=1}^{\infty} \frac{1}{(\eta_n+1)^2} < \infty.$$

Now, define

$$\begin{aligned}\mathcal{F}_n &\triangleq \mathcal{B}(I_1, ..., I_n, q_0, ..., q_n, \mu_0, ..., \mu_n, \eta_0, ..., \eta_n),\\ R_n(x) &\triangleq \mathbb{E}(Q_n(I_{n+1}, q_n, A_n, E_n, x) \,|\, \mathcal{F}_n)\\ &= (1 - q_n\gamma)\frac{A_n E_n q_n}{1 - q_n x} - q_n\gamma\frac{A_n E_n q_n}{q_n x}.\end{aligned}$$

Since $0 < A_n < 1$ and $(1-\mu^*)^2 \le E_n \le 1$, we can see $R_n(x) = 0$ if and only if $x = \gamma$, thus verifying Assumption 2.3.1. Assumption 2.3.2 is verified straightforwardly from the facts $\mu_n \in [\mu_*, \mu^*]$ and $0 < \inf_n q_n \le \sup_n q_n < 1$. From the same facts, it follows that there exists a positive constant C_1 such that

$$\sup_{n\in\mathbb{N}} |Q_n(I_{n+1}, q_n, A_n, E_n, x)| \leq C_1.$$

Consequently, there exists another positive constant C_2 such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}(Q_n^2(I_{n+1}, q_n, A_n, E_n, x) \,|\, \mathcal{F}_n) \leq C_2,$$

thus verifying Assumption 2.3.3. The desired result follows by Theorem 2.3.2. □

2.5 Multivariate Applications

We present three more applications of our convergence analysis to problems with multivariate priors in which covariance matrices are used to quantify similarities or differences between unknown values. Section 2.5.1 gives the first consistency proof for a Bayesian logistic regression method, thus solving a problem that has been open since at least [20]. Section 2.5.2 proves, for the first time, the convergence of an approximate value iteration algorithm in a Markov decision problem with correlated Bayesian beliefs about the values of different states. Section 2.5.3 proves a new result for ranking and selection with unknown correlation structures.[5]

2.5.1 Bayesian Logistic Regression

Let $(X_n, Y_n)_{n=0}^{\infty}$ be a sequence of pairs consisting of a binary observation $Y_n \in \{0, 1\}$ and a vector $X_n \in \mathbb{R}^K$ of covariates. We assume that the covariates

[5] It bears repeating that *none* of these applications fits into the framework of [45]. Sections 2.5.1 and 2.5.2 use multivariate normal priors, but not moment-matching. Section 2.5.3 uses a Wishart prior to model unknown correlations.

$(X_n)_{n=0}^{\infty}$ are drawn independently from some common, but unknown distribution. The observations (Y_n) are independent and satisfy $P(Y_n = 1 \mid X_n) = \ell(X_n; \theta)$ where

$$\ell(x;\theta) = \frac{1}{1 + \exp(-x^\top \theta)}, \tag{2.29}$$

with $\theta \in \mathbb{R}^K$ being a vector of regression coefficients. Equation (2.29) denotes a standard logistic regression model; in classical statistics, θ has to be learned through maximum likelihood estimation given a fixed sample of data.

Suppose, however, that we wish to update our estimate of θ after each new observation. This may happen if these estimates are being used to solve an optimization problem (as in the setting of [60]; for instance, the covariates may represent product attributes, which help us learn about demand distributions, which in turn are important for making stocking decisions). A multivariate normal prior $\theta \sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_0)$ allows us to model beliefs about similarities and differences between the regression coefficients. For instance, suppose that two covariates X_i, X_j are dummy variables representing two distinct products, and that product i is observed much more frequently than product j. If the (i,j)th entry of $\mathbf{\Sigma}_0$ is positive, this suggests a degree of similarity between i and j, so that we can make use of what we have learned about i when we do finally observe j. See [7] for an example of such an application.

Unfortunately, the multivariate normal prior is not conjugate with the binary observations encountered in logistic regression. For this reason, researchers going back to at least [20] have used approximate Bayesian methods to create tractable updates. The predominant approach in this literature is to use an update of the form

$$\begin{aligned}\mu_{n+1} &= \mu_n - \left(\ell\left(X_n;\mu_n\right) - Y_{n+1}\right)\boldsymbol{\Sigma}_{n+1}X_n, \qquad (2.30)\\ \boldsymbol{\Sigma}_{n+1}^{-1} &= \boldsymbol{\Sigma}_n^{-1} + vX_nX_n^T. \qquad (2.31)\end{aligned}$$

It is easy to see that (2.30)-(2.31) are virtually identical to the well-known recursive least squares update. In other words, the approximation strategy in this case is to simply treat logistic regression as if it were *linear* regression; the quantity $\ell\left(X_n;\mu_n\right) - Y_{n+1}$ in (2.30) acts as a "residual," whereas $v > 0$ is an artificial parameter standing in for the residual variance (there being no exact analog of this concept in logistic regression). Later work by [6] showed that (2.30)-(2.31) can be obtained by applying a first-order Taylor approximation (variational bound) to the logistic likelihood function, in line with the idea of "linearizing" the logistic regression model. This and subsequent work focused on computational issues, such as how to choose v optimally (see also [7]), and never formally studied the consistency of the procedure.

Using our framework, we obtain (for the first time) the surprising result that (2.30) is consistent, that is, $\mu_n \to \theta$ almost surely under (2.30)-(2.31). We first give the assumptions used in our analysis, then state the result and give the proof.

Assumption 2.5.1. *The covariate vectors* $(X_n)_{n=0}^{\infty}$ *are drawn i.i.d. from a common distribution satisfying* $\mathbb{E}\left(X_nX_n^T\right) = \mathbf{A}$*, where* $\mathbf{A}$ *is a positive definite symmetric matrix.*

Assumption 2.5.2. *The sequence* $(X_n)_{n=0}^{\infty}$ *satisfies* $0 < \inf_n \|X_n\|_1 \leq \sup_n \|X_n\|_1 < \infty$ *almost surely.*

Together, these assumptions lead to a law of large numbers for the data generating process, i.e., $\lim_{n\to\infty}\frac{1}{n}\sum_{n=0}^{\infty}X_nX_n^T=\mathbf{A}$ almost surely. For simplicity, we also suppose that the noise parameter $v>0$ in (2.30)-(2.31) is some fixed but arbitrary constant.

Theorem 2.5.1. *Suppose Assumptions 2.5.1-2.5.2 hold and μ_n is updated using (2.30), while $\mathbf{\Sigma}_n$ is updated using (2.31). Then $\mu_n\to\theta$ almost surely.*

Proof. Without loss of generality, we assume that $\theta=0$. Recalling that $v>0$ is a fixed constant, we let $\mathbf{B}=v\mathbf{A}$, where $\mathbf{A}$ is the matrix from Assumption 2.5.1, and rewrite (2.30) as

$$\begin{aligned}\mathbf{B}^{\frac{1}{2}}\mu_{n+1} &= \mathbf{B}^{\frac{1}{2}}\left(\mu_n-\left(\ell\left(X_n;\mu_n\right)-Y_{n+1}\right)\mathbf{\Sigma}_{n+1}X_n\right)\\ &= \mathbf{B}^{\frac{1}{2}}\left(\mu_n-\frac{1}{n+1}\left(\ell\left(X_n;\mu_n\right)-Y_{n+1}\right)\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}+\mathbf{B}^{-1}\right)X_n\right)\\ &= \mathbf{B}^{\frac{1}{2}}\mu_n-\frac{1}{n+1}\left(\ell\left(X_n;\mu_n\right)-Y_{n+1}\right)\mathbf{B}^{-\frac{1}{2}}X_n \qquad (2.32)\\ &\quad -\frac{1}{n+1}\left(\ell\left(X_n;\mu_n\right)-Y_{n+1}\right)\mathbf{B}^{\frac{1}{2}}\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}\right)X_n. \qquad (2.33)\end{aligned}$$

We will demonstrate the convergence of the transformed sequence $\left(\mathbf{B}^{\frac{1}{2}}\mu_n\right)$, which implies the consistency of the original sequence. Equations (2.32)-(2.33) represent a stochastic approximation algorithm; we define

$$\begin{aligned}\mathcal{F}_n &\triangleq \mathcal{B}\left(Y_1,...,Y_n,X_0,...,X_n,\mu_0,...,\mu_n,\mathbf{\Sigma}_0,...,\mathbf{\Sigma}_n\right),\\ R_n\left(\mu\right) &\triangleq \mathbb{E}\left(\left(\ell\left(X_n;\mu\right)-Y_{n+1}\right)\mathbf{B}^{-\frac{1}{2}}X_n|\mathcal{F}_n\right)\\ &= \left(\left(\frac{1}{1+e^{-X_n^T\mu}}-1\right)\frac{1}{2}+\frac{1}{1+e^{-X_n^T\mu}}\frac{1}{2}\right)\mathbf{B}^{-\frac{1}{2}}X_n\\ &= \frac{1}{2}\frac{1-e^{-X_n^T\mu}}{1+e^{-X_n^T\mu}}\mathbf{B}^{-\frac{1}{2}}X_n.\end{aligned}$$

The structure of R_n poses the main technical challenge for the proof, as Assumption 2.3.1 is not applicable; instead of $\theta = 0$ being the unique root of R_n for all n, we have $R_n(\mu) = 0$ if and only if $X_n^T \mu = 0$ individually for *each* n. This also introduces complications for the other assumptions in Section 2.3, which are expressed in terms of the unique root. Nonetheless, the overall structure of the proof is the same as that of Theorem 2.3.3; we discuss how the remaining assumptions should be modified and then complete the argument.

Convexity condition. We calculate the inner product of the iterate $\mathbf{B}^{\frac{1}{2}}\mu_n$ and the gradient in (2.32), yielding

$$\left(\mathbf{B}^{\frac{1}{2}}\mu_n\right)^T \left(\ell\left(X_n;\mu_n\right) - Y_{n+1}\right)\mathbf{B}^{-\frac{1}{2}}X_n = \left(\ell\left(X_n;\mu_n\right) - Y_{n+1}\right)\mu_n^T X_n.$$

Taking the conditional expectation, we find

$$\mathbb{E}\left(\left(\ell\left(X_n;\mu_n\right) - Y_{n+1}\right)\mu_n^T X_n \,|\, \mathcal{F}_n\right) = \frac{1}{2}\mu_n^T X_n \frac{1 - e^{-X_n^T\mu_n}}{1 + e^{-X_n^T\mu_n}} \geq 0, \tag{2.34}$$

and, for $\varepsilon > 0$ and $n = 1, 2, ...$, we also have

$$\inf_{(X_n^T\mu)^2 > \varepsilon, n\in\mathbb{N}} \frac{1}{2}X_n^T\mu \frac{1 - e^{-X_n^T\mu}}{1 + e^{-X_n^T\mu}} > 0, \tag{2.35}$$

the relevant analog of the convexity condition in Assumption 2.3.2.

Bias condition. Recall that (2.33) serves as the bias term. From the LLN obtained from Assumptions 2.5.1 and 2.5.2, we have

$$(n+1)\,\mathbf{\Sigma}_{n+1} - \mathbf{B}^{-1} \to 0, \tag{2.36}$$

suggesting that the bias eventually vanishes. However, analogously to Assumption 2.3.4, it is necessary to ensure that this happens fast enough in some sense. This is established in the following auxiliary technical lemma, which is proved right after the current proof,

Lemma 2.5.1. *For $p = 1, 2$,*

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_p &< \infty, \\ \sum_{n=1}^{\infty} \frac{1}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_p^2 &< \infty.\end{aligned}$$

Remainder of proof. Similarly to Theorem 2.3.3, we calculate

$$\begin{aligned}&\left\| \mathbf{B}^{\frac{1}{2}} \mu_{n+1} \right\|_2^2 \\ &= \mu_{n+1}^T \mathbf{B} \mu_{n+1} \\ &= \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2 + \frac{1}{(n+1)^2} \left(\ell\left(X_n; \mu_n\right) - Y_{n+1} \right)^2 \left\| \mathbf{B}^{-\frac{1}{2}} X_n \right\|_2^2 &(2.37) \\ &\quad + \frac{1}{(n+1)^2} \left(\ell\left(X_n; \mu_n\right) - Y_{n+1} \right)^2 \left\| \mathbf{B}^{\frac{1}{2}} \left((n+1) \mathbf{\Sigma}_{n+1} - \mathbf{B}^{-1} \right) X_n \right\|_2^2 &(2.38) \\ &\quad - \frac{2}{n+1} \left(\ell\left(X_n; \mu_n\right) - Y_{n+1} \right) X_n^T \mu_n &(2.39) \\ &\quad - \frac{2}{n+1} \left(\ell\left(X_n; \mu_n\right) - Y_{n+1} \right) \mu_n^T \mathbf{B} \left((n+1) \mathbf{\Sigma}_{n+1} - \mathbf{B}^{-1} \right) X_n &(2.40) \\ &\quad + \frac{2}{(n+1)^2} \left(\ell\left(X_n; \mu_n\right) - Y_{n+1} \right)^2 X_n^T \left((n+1) \mathbf{\Sigma}_{n+1} - \mathbf{B}^{-1} \right) X_n. &(2.41)\end{aligned}$$

From (2.36) and the boundedness of (X_n) (Assumption 2.5.2), we can bound terms in (2.37), (2.38), and (2.41): there must exist a positive constant C_1 such that, for all n,

$$\begin{aligned}\left(\ell\left(X_n, \mu_n\right) - Y_{n+1} \right)^2 \left\| \mathbf{B}^{-\frac{1}{2}} X_n \right\|_2^2 &\leq C_1, \\ \left(\ell\left(X_n, \mu_n\right) - Y_{n+1} \right)^2 \left\| \mathbf{B}^{\frac{1}{2}} \left((n+1) \mathbf{\Sigma}_{n+1} - \mathbf{B}^{-1} \right) X_n \right\|_2^2 &\leq C_1, \\ 2 \left(\ell\left(X_n, \mu_n\right) - Y_{n+1} \right)^2 X_n^T \left((n+1) \mathbf{\Sigma}_{n+1} - \mathbf{B}^{-1} \right) X_n &\leq C_1.\end{aligned}$$

We now handle (2.40); applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&-\frac{2}{n+1}\left(\ell\left(X_n,\mu_n\right)-Y_{n+1}\right)\mu_n^T\mathbf{B}\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}\right)X_n \\
&\leq \frac{2}{n+1}\left|\ell\left(X_n,\mu_n\right)-Y_{n+1}\right|\left\|\mathbf{B}^{\frac{1}{2}}\mu_n\right\|_2\left\|\mathbf{B}^{\frac{1}{2}}\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}\right)X_n\right\|_2 \\
&= \left(\frac{2}{(n+1)^{\frac{5}{8}}}\left|\ell\left(X_n,\mu_n\right)-Y_{n+1}\right|\left\|\mathbf{B}^{\frac{1}{2}}\mu_n\right\|_2\right) \\
&\quad\times\left(\frac{1}{(n+1)^{\frac{3}{8}}}\left\|\mathbf{B}^{\frac{1}{2}}\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}\right)X_n\right\|_2\right) \\
&\leq \frac{4}{(n+1)^{\frac{5}{4}}}\left(\ell\left(X_n,\mu_n\right)-Y_{n+1}\right)^2\left\|\mathbf{B}^{\frac{1}{2}}\mu_n\right\|_2^2 \\
&\quad+\frac{1}{(n+1)^{\frac{3}{4}}}\left\|\mathbf{B}^{\frac{1}{2}}\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}\right)X_n\right\|_2^2. \qquad (2.42)
\end{aligned}$$

We now bound each of the terms in (2.42). First, there exists a positive constant C_2 such that

$$\frac{4}{(n+1)^{\frac{5}{4}}}\left(\ell\left(X_n,\mu_n\right)-Y_{n+1}\right)^2\left\|\mathbf{B}^{\frac{1}{2}}\mu_n\right\|_2^2 \leq \frac{C_2}{(n+1)^{\frac{5}{4}}}\left\|\mathbf{B}^{\frac{1}{2}}\mu_n\right\|_2^2.$$

Second, by Assumption 2.5.2 together with (2.36), there exists a positive constant C_3 such that

$$\begin{aligned}
&\frac{1}{(n+1)^{\frac{3}{4}}}\left\|\mathbf{B}^{\frac{1}{2}}\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}\right)X_n\right\|_2^2 \\
&= \frac{1}{(n+1)^{\frac{3}{4}}}\left\|\mathbf{B}^{\frac{1}{2}}(n+1)\mathbf{\Sigma}_{n+1}\left(\frac{1}{n+1}\mathbf{\Sigma}_{n+1}^{-1}-\mathbf{B}\right)\mathbf{B}^{-1}X_n\right\|_2^2 \\
&\leq \frac{1}{(n+1)^{\frac{3}{4}}}\left\|\mathbf{B}^{\frac{1}{2}}\right\|_2^2\left\|(n+1)\mathbf{\Sigma}_{n+1}\right\|_2^2\left\|\frac{1}{n+1}\mathbf{\Sigma}_{n+1}^{-1}-\mathbf{B}\right\|_2^2\left\|\mathbf{B}^{-1}\right\|_2^2\left\|X_n\right\|_2^2 \\
&\leq \frac{C_3}{(n+1)^{\frac{3}{4}}}\left\|\frac{1}{n+1}\mathbf{\Sigma}_{n+1}^{-1}-\mathbf{B}\right\|_2^2,
\end{aligned}$$

where the first inequality holds because of the submultiplicativity of the norm $\|\cdot\|_2$. Thus, the desired bound on (2.40) is given by

$$-\frac{2}{n+1}\left(\ell\left(X_n;\mu_n\right)-Y_{n+1}\right)\mu_n^T\mathbf{B}\left((n+1)\mathbf{\Sigma}_{n+1}-\mathbf{B}^{-1}\right)X_n$$

$$\leq \frac{C_2}{(n+1)^{\frac{5}{4}}} \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2 + \frac{C_3}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_2^2.$$

Putting together the bounds on (2.37)-(2.41), we obtain

$$\begin{aligned}
\left\| \mathbf{B}^{\frac{1}{2}} \mu_{n+1} \right\|_2^2 &\leq \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2 + \frac{3C_1}{(n+1)^2} - \frac{2}{n+1} \left(\ell \left(X_n; \mu_n \right) - Y_{n+1} \right) X_n^T \mu_n \\
&\quad + \frac{C_2}{(n+1)^{\frac{5}{4}}} \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2 + \frac{C_3}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_2^2 \\
&\leq \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2 \left(1 + \frac{C_2}{(n+1)^{\frac{5}{4}}} \right) + \frac{3C_1}{(n+1)^2} + \frac{C_3}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_2^2 \\
&\quad - \frac{2}{n+1} \left(\ell \left(X_n; \mu_n \right) - Y_{n+1} \right) X_n^T \mu_n, \qquad (2.43)
\end{aligned}$$

where the final term in (2.43) is carried over from (2.39). Taking the conditional expectation, we obtain

$$\begin{aligned}
\mathbb{E} \left(\left\| \mathbf{B}^{\frac{1}{2}} \mu_{n+1} \right\|_2^2 | \mathcal{F}_n \right) &\leq \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2 \left(1 + \frac{C_2}{(n+1)^{\frac{5}{4}}} \right) + \frac{3C_1}{(n+1)^2} \\
&\quad + \frac{C_3}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_2^2 \\
&\quad - \frac{1}{n+1} \frac{1 - e^{-X_n^T \mu_n}}{1 + e^{-X_n^T \mu_n}} X_n^T \mu_n.
\end{aligned}$$

It is obvious that

$$\sum_n \frac{C_2}{(n+1)^{\frac{5}{4}}} < \infty,$$

and by Lemma 2.5.1, we also have

$$\sum_n \frac{3C_1}{(n+1)^2} + \frac{C_3}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_2^2 < \infty.$$

These facts together with (2.34) enable us to apply Theorem 1 of [56]. It follows that $\lim_{n \to \infty} \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2$ exists and

$$\sum_{n=0}^{\infty} \frac{1}{n+1} \frac{1 - e^{-X_n^T \mu_n}}{1 + e^{-X_n^T \mu_n}} X_n^T \mu_n < \infty$$

almost surely. Therefore, for every sample path, there must exist a subsequence $\left(X_{n_k}^T \mu_{n_k}\right)$ of $\left(X_n^T \mu_n\right)$ such that, as $k \to \infty$, $X_{n_k}^T \mu_{n_k} \to 0$. On the other hand, since $\lim_{n\to\infty} \left\|\mathbf{B}^{\frac{1}{2}} \mu_n\right\|_2^2$ exists, the sequence (μ_n) is bounded (although the precise value of this bound depends on the sample path). Therefore, there must exist a subsequence $\left(\mu_{n_{k_j}}\right)$ of (μ_{n_k}) such that, as $j \to \infty$, we have $\mu_{n_{k_j}} \to \nu$ where ν is some fixed vector. Applying Assumption 2.5.2, we have

$$\begin{aligned}
\lim_{j\to\infty} \left| X_{n_{k_j}}^T \nu \right| &= \lim_{j\to\infty} \left| X_{n_{k_j}}^T \left(\nu - \mu_{n_{k_j}} + \mu_{n_{k_j}} \right) \right| \\
&\leq \lim_{j\to\infty} \left| X_{n_{k_j}}^T \left(\nu - \mu_{n_{k_j}} \right) \right| + \lim_{j\to\infty} \left| X_{n_{k_j}}^T \mu_{n_{k_j}} \right| \\
&= 0.
\end{aligned}$$

Thus, for any arbitrary $\varepsilon > 0$, there exists an integer J such that, for all $j \geq J$,

$$\left| X_{n_{k_j}}^T \nu \right| < \varepsilon. \tag{2.44}$$

However, since $\left(X_{n_{k_j}}\right)_{j=J}^{\infty}$ is also an infinite sequence of i.i.d. samples from the distribution of X, there must exist K linearly independent vectors $X_{n_{k_{j_1}}}, ..., X_{n_{k_{j_K}}}$ from $\left(X_{n_{k_j}}\right)_{j=J}^{\infty}$ that can be a basis of $\mathbb{R}^K$. To show this, suppose that all $\left(X_{n_{k_j}}\right)_{j=J}^{\infty}$ come from a subspace V of $\mathbb{R}^K$ and $V \neq \mathbb{R}^K$; then, there must be a nonzero vector $\gamma \in V^\perp$ such that

$$\begin{aligned}
\gamma^T \mathbf{A} \gamma &= \gamma^T \left(\lim_{J'\to\infty} \frac{1}{J'} \sum_{j=J}^{J'} X_{n_{k_j}} X_{n_{k_j}}^T \right) \gamma \\
&= \lim_{J'\to\infty} \frac{1}{J'} \sum_{j=J}^{J'} \left(X_{n_{k_j}}^T \gamma \right)^2 \\
&= 0,
\end{aligned}$$

where the first equality holds by Assumption 2.5.2, but the last line contradicts Assumption 2.5.1, which holds that $\mathbf{A}$ is positive definite.

Then, to satisfy (2.44), since ε can be arbitrarily small, by Assumption 2.5.2, ν has to be the zero vector. Thus, $\mu_{n_{k_j}} \to 0$, so

$$\lim_{j\to\infty} \left\| \mathbf{B}^{\frac{1}{2}} \mu_{n_{k_j}} \right\|_2^2 = 0,$$

but $\left(\mu_{n_{k_j}}\right)$ is a subsequence of (μ_n) and $\lim_{n\to\infty} \left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2$ exists; therefore, $\left\| \mathbf{B}^{\frac{1}{2}} \mu_n \right\|_2^2 \to 0$, whence $\mu_n \to 0$ a.s., as desired. □

Proof of Lemma 2.5.1. For notational convenience, let $\mathbf{M}_n = vX_nX_n^T$. From (2.31), we obtain

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} &= \lim_{n\to\infty} \frac{1}{n+1} \left(\mathbf{\Sigma}_0^{-1} + v \sum_{k=0}^{n} X_k X_k^T \right) \\
&= v \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} X_k X_k^T \\
&= v\mathbf{A} \\
&= \mathbf{B}, \qquad (2.45)
\end{aligned}$$

where the third equality holds because of Assumption 2.5.2. For any two integers $i, j \in \{1, ..., K\}$, denote the (i, j)th element of $\mathbf{\Sigma}_{n+1}^{-1}$ by $\left(\mathbf{\Sigma}_{n+1}^{-1}\right)^{(i,j)}$. We will first show that

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)^{\frac{3}{4}}} \left| \frac{1}{n+1} \left(\mathbf{\Sigma}_{n+1}^{-1}\right)^{(i,j)} - \mathbf{B}^{(i,j)} \right| < \infty. \qquad (2.46)$$

By Kolmogorov's three-series theorem [61], the convergence of (2.46) follows from the convergence of the three series

$$\sum P(|\xi_n| \geq c), \quad \sum \mathbb{E}(\xi_n 1_{\{|\xi_n| \leq c\}}), \quad \sum Var(\xi_n 1_{\{|\xi_n| \leq c\}}),$$

where c is some positive constant and

$$\xi_n = \frac{1}{(n+1)^{\frac{3}{4}}} \left| \frac{1}{n+1} \left(\mathbf{\Sigma}_{n+1}^{-1}\right)^{(i,j)} - \mathbf{B}^{(i,j)} \right|$$

$$
\begin{aligned}
&= \frac{1}{(n+1)^{\frac{3}{4}}} \left| \frac{1}{n+1} \left(\mathbf{\Sigma}_0^{-1} + v \sum_{k=0}^{n} X_k X_k^T \right)^{(i,j)} - \mathbf{B}^{(i,j)} \right| \\
&= \frac{1}{(n+1)^{\frac{3}{4}}} \left| \frac{1}{n+1} \sum_{k=0}^{n} \left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)} \right) + \frac{1}{n+1} \left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right|. \qquad (2.47)
\end{aligned}
$$

To show the convergence of the first series, notice that by (2.45), $\xi_n = \frac{1}{(n+1)^{\frac{3}{4}}} O(1) = o(1)$. Thus, there must exist a large enough positive constant c such that $\xi_n < c$ for all n. It then follows that $P\left(|\xi_n| \geq c\right) = 0$ for all n, whence the first series converges.

Next, we show convergence of the last (third) series. From (2.47), we have

$$
\begin{aligned}
&\sum Var(\xi_n 1_{\{|\xi_n| \leq c\}}) \\
&= \sum Var(\xi_n) \\
&= \sum_{n=0}^{\infty} Var \left(\frac{1}{(n+1)^{\frac{3}{4}}} \left| \frac{1}{n+1} \sum_{k=0}^{n} \left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)} \right) + \frac{1}{n+1} \left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right| \right) \\
&= \sum_{n=0}^{\infty} \frac{1}{(n+1)^{\frac{3}{2}}} Var \left(\left| \frac{1}{n+1} \sum_{k=0}^{n} \left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)} \right) + \frac{1}{n+1} \left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right| \right) \\
&= \sum_{n=0}^{\infty} \frac{1}{(n+1)^{\frac{7}{2}}} Var \left(\left| \sum_{k=0}^{n} \left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)} \right) + \left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right| \right) \\
&\leq \sum_{n=0}^{\infty} \frac{1}{(n+1)^{\frac{7}{2}}} \mathbb{E} \left(\sum_{k=0}^{n} \left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)} \right) + \left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right)^2 \\
&\leq \sum_{n=0}^{\infty} \frac{2}{(n+1)^{\frac{7}{2}}} \left(\mathbb{E} \left(\sum_{k=0}^{n} \left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)} \right) \right)^2 + \left(\left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right)^2 \right) \\
&= \sum_{n=0}^{\infty} \frac{2}{(n+1)^{\frac{7}{2}}} \mathbb{E} \left(\sum_{k=0}^{n} \left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)} \right) \right)^2 + \sum_{n=0}^{\infty} \frac{2 \left(\left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right)^2}{(n+1)^{\frac{7}{2}}}.
\end{aligned}
$$

To handle the second term on the last line, it is obvious that

$$
\sum_{n=0}^{\infty} \frac{2 \left(\left(\mathbf{\Sigma}_0^{-1} \right)^{(i,j)} \right)^2}{(n+1)^{\frac{7}{2}}} < \infty.
$$

To handle the first term, by Assumptions 2.5.1 and 2.5.2, there must exist a large enough positive constant C_1 such that

$$\begin{aligned}
&\mathbb{E}\left(\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)\right)^2 \\
&= \mathbb{E}\left(\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)^2+2\sum_{0\le k<k'\le n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)\left(\mathbf{M}_{k'}^{(i,j)}-\mathbf{B}^{(i,j)}\right)\right) \\
&= \sum_{k=0}^{n}\mathbb{E}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)^2+2\sum_{0\le k<k'\le n}\mathbb{E}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)\mathbb{E}\left(\mathbf{M}_{k'}^{(i,j)}-\mathbf{B}^{(i,j)}\right) \\
&= \sum_{k=0}^{n}\mathbb{E}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)^2 \\
&\le (n+1)C_1, \qquad (2.48)
\end{aligned}$$

where the second and third equality hold from Assumption 2.5.1 and the first inequality holds from Assumption 2.5.2. Thus, together we have

$$\begin{aligned}
\sum Var(\xi_n 1_{\{|\xi_n|\le c\}}) &\le \sum_{n=0}^{\infty}\frac{2}{(n+1)^{\frac{7}{2}}}\mathbb{E}\left(\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)\right)^2+\sum_{n=0}^{\infty}\frac{2\left(\left(\mathbf{\Sigma}_0^{-1}\right)^{(i,j)}\right)^2}{(n+1)^{\frac{7}{2}}} \\
&\le \sum_{n=0}^{\infty}\frac{2C_1}{(n+1)^{\frac{5}{2}}}+\sum_{n=0}^{\infty}\frac{2\left(\left(\mathbf{\Sigma}_0^{-1}\right)^{(i,j)}\right)^2}{(n+1)^{\frac{7}{2}}}<\infty,
\end{aligned}$$

showing the convergence of the third series.

To show the convergence of the second series, from (2.47) we have

$$\begin{aligned}
\sum \mathbb{E}(\xi_n 1_{\{|\xi_n|\le c\}}) &= \sum \mathbb{E}(\xi_n) \\
&= \sum_{n=0}^{\infty}\mathbb{E}\left(\frac{1}{(n+1)^{\frac{3}{4}}}\left|\frac{1}{n+1}\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)+\frac{1}{n+1}\left(\mathbf{\Sigma}_0^{-1}\right)^{(i,j)}\right|\right) \\
&= \sum_{n=0}^{\infty}\frac{1}{(n+1)^{\frac{3}{4}}}\mathbb{E}\left(\left|\frac{1}{n+1}\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)+\frac{1}{n+1}\left(\mathbf{\Sigma}_0^{-1}\right)^{(i,j)}\right|\right) \\
&= \sum_{n=0}^{\infty}\frac{1}{(n+1)^{\frac{7}{4}}}\mathbb{E}\left(\left|\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)+\left(\mathbf{\Sigma}_0^{-1}\right)^{(i,j)}\right|\right) \\
&\le \sum_{n=0}^{\infty}\frac{1}{(n+1)^{\frac{7}{4}}}\left(\mathbb{E}\left|\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)}-\mathbf{B}^{(i,j)}\right)\right|+\left|\left(\mathbf{\Sigma}_0^{-1}\right)^{(i,j)}\right|\right)
\end{aligned}$$

$$\begin{aligned} &\le \sum_{n=0}^{\infty} \frac{1}{(n+1)^{\frac{7}{4}}} \left(\sqrt{\mathbb{E}\left(\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)}\right)\right)^2} + \left|\left(\boldsymbol{\Sigma}_0^{-1}\right)^{(i,j)}\right| \right) \\ &= \sum_{n=0}^{\infty} \frac{1}{(n+1)^{\frac{7}{4}}} \sqrt{\mathbb{E}\left(\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)}\right)\right)^2} + \sum_{n=0}^{\infty} \frac{\left|\left(\boldsymbol{\Sigma}_0^{-1}\right)^{(i,j)}\right|}{(n+1)^{\frac{7}{4}}}. \end{aligned}$$

To handle the second term on the last line, we note that

$$\sum_{n=0}^{\infty} \frac{\left|\left(\boldsymbol{\Sigma}_0^{-1}\right)^{(i,j)}\right|}{(n+1)^{\frac{7}{4}}} < \infty.$$

To handle the first term, by (2.48) we have

$$\sqrt{\mathbb{E}\left(\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)}\right)\right)^2} \le \sqrt{(n+1)C_1}.$$

Thus, together we have

$$\begin{aligned} \sum \mathbb{E}(\xi_n 1_{\{|\xi_n| \le c\}}) &\le \sum_{n=0}^{\infty} \frac{1}{(n+1)^{\frac{7}{4}}} \sqrt{\mathbb{E}\left(\sum_{k=0}^{n}\left(\mathbf{M}_k^{(i,j)} - \mathbf{B}^{(i,j)}\right)\right)^2} + \sum_{n=0}^{\infty} \frac{\left|\left(\boldsymbol{\Sigma}_0^{-1}\right)^{(i,j)}\right|}{(n+1)^{\frac{7}{4}}} \\ &\le \sum_{n=0}^{\infty} \frac{\sqrt{C_1}}{(n+1)^{\frac{5}{4}}} + \sum_{n=0}^{\infty} \frac{\left|\left(\boldsymbol{\Sigma}_0^{-1}\right)^{(i,j)}\right|}{(n+1)^{\frac{7}{4}}} < \infty, \end{aligned}$$

proving the convergence of the second series. Therefore, (2.46) holds for any two integers $i, j \in \{1, ..., K\}$, so we have

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{1}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \boldsymbol{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_1 &\le \sum_{n=1}^{\infty} \frac{1}{(n+1)^{\frac{3}{4}}} \sum_{i,j} \left| \frac{1}{n+1} \left(\boldsymbol{\Sigma}_{n+1}^{-1}\right)^{(i,j)} - \mathbf{B}^{(i,j)} \right| \\ &= \sum_{i,j} \sum_{n=1}^{\infty} \frac{1}{(n+1)^{\frac{3}{4}}} \left| \frac{1}{n+1} \left(\boldsymbol{\Sigma}_{n+1}^{-1}\right)^{(i,j)} - \mathbf{B}^{(i,j)} \right| \\ &< \infty. \end{aligned}$$

Now from (2.45), there must exist a large enough integer N such that, for all $n \ge N$,

$$\left\| \frac{1}{n+1} \boldsymbol{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_1 \le 1.$$

Then,

$$\sum_{n\geq N} \frac{1}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_1^2 \;\leq\; \sum_{n\geq N} \frac{1}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_1,$$

whence

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)^{\frac{3}{4}}} \left\| \frac{1}{n+1} \mathbf{\Sigma}_{n+1}^{-1} - \mathbf{B} \right\|_1^2 \;<\; \infty.$$

For $p = 2$, since $\|\cdot\|_2 \leq \sqrt{K}\,\|\cdot\|_1$, we have the desired results. □

2.5.2 Reinforcement Learning with Correlated Beliefs

Consider a Markov decision process [62] with finite state space $\mathcal{S}$, finite decision space $\mathcal{X}$, and single-period reward function $C : \mathcal{S} \times \mathcal{X} \to \mathbb{R}$ with discount factor $\gamma \in (0, 1)$. The maximum cumulative infinite-horizon discounted reward obtainable from state $s \in \mathcal{S}$ is given by the well-known Bellman equation [22]

$$V(s) = \max_x C(s, x) + \gamma \sum_{s'} P(s'|s, x)\, V(s').$$

In reinforcement learning [63], it is useful to redefine V as a function of a state-action pair, i.e.,

$$V(s, x) = C(s, x) + \gamma \sum_{s'} P(s'|s, x) \left(\max_{x'} V(s', x') \right). \tag{2.49}$$

The optimal action to take in state s is given by $\arg\max_x V(s, x)$. In practice, however, (2.49) is difficult to solve as the state and action spaces may be large and the transition probabilities may be completely unknown.

Approximate value iteration algorithms address this issue by solving (2.49) approximately. Suppose that we are in state s_n in the nth stage of the algorithm, and choose the action x_n. The next state s_{n+1} is then drawn from the transition distribution $P\left(\cdot|s_n,x_n\right)$ and observed. We then compute the quantity

$$v_{n+1} = C\left(s_n,x_n\right) + \gamma \max_x \bar{V}_n\left(s_{n+1},x\right), \tag{2.50}$$

and interpret this as an approximate observation of the unknown value $V\left(s_n,x_n\right)$, bootstrapped from an existing approximation function $\bar{V}_n$. Some form of stochastic approximation can then be used to smooth v_{n+1} together with $\bar{V}_n\left(s_n,x_n\right)$. If every state-action pair is visited infinitely often, SA is provably convergent [64–66] despite the fact that (2.50) is a biased estimate of $V\left(s_n,x_n\right)$.

However, if the state and action spaces are large, convergence may be too slow for any practical time horizon [67], driving interest in "spreading" methods that are able to learn about multiple state-action pairs from one observation [68]. For this purpose, [46] proposed the following approximate Bayesian scheme. We begin with the multivariate normal prior $V \sim \mathcal{N}\left(\bar{V}_0,\mathbf{\Sigma}_0\right)$, where $\bar{V}_0$ is our initial approximation of V and $\mathbf{\Sigma}_0$ includes correlated beliefs about different state-action pairs. After calculating (2.50), we update

$$\begin{aligned}
\bar{V}_{n+1}\left(s,x\right) &= \bar{V}_n\left(s,x\right) && (2.51)\\
&\quad -\frac{\mathbf{\Sigma}_n\left(\left(s,x\right),\left(s_n,x_n\right)\right)\left(\bar{V}_n\left(s_n,x_n\right)-v_{n+1}\right)}{\lambda_n^2+\mathbf{\Sigma}_n\left(\left(s_n,x_n\right),\left(s_n,x_n\right)\right)}, && (2.52)\\
\mathbf{\Sigma}_{n+1}\left(\left(s,x\right),\left(s',x'\right)\right) &= \mathbf{\Sigma}_n\left(\left(s,x\right),\left(s',x'\right)\right) && (2.53)\\
&\quad -\frac{\mathbf{\Sigma}_n\left(\left(s,x\right),\left(s_n,x_n\right)\right)\mathbf{\Sigma}_n\left(\left(s_n,x_n\right),\left(s',x'\right)\right)}{\lambda_n^2+\mathbf{\Sigma}_n\left(\left(s_n,x_n\right),\left(s_n,x_n\right)\right)}, && (2.54)
\end{aligned}$$

for all state-action pairs $(s,x) \in \mathcal{S} \times \mathcal{X}$. If v_{n+1} were an unbiased observation of $V\left(s_n,x_n\right)$ with variance λ_n^2, (2.52)-(2.54) would describe a conjugate model. However, no such unbiased observation is available, so we simply apply this update with the biased observation from (2.50), treating λ_n^2 as a tunable parameter (analogous to a stepsize sequence). We note that, in practice, $\mathbf{\Sigma}_n$ would be expensive to store if the state and action spaces are large; however, the concept of the correlated Bayesian model can potentially be extended to more compact belief representations [69]. Here, we focus on applying our theory from Section 2.3 to show convergence in the base model where the value function is represented by a lookup table.

If $\mathbf{\Sigma}_0$ is diagonal, (2.52) is equivalent to recursive sample averaging and thus is provably convergent by standard SA theory [70]. We will prove convergence for a modified version of (2.52)-(2.54) that includes correlations (non-diagonal priors). For our analysis, we work with the sequence

$$\lambda_n^2 = (n+1)\mathbf{\Sigma}_n\left(\left(s_n, x_n\right), \left(s_n, x_n\right)\right).$$

We also impose some additional assumptions on the prior covariance matrix $\mathbf{\Sigma}_0$. The prior covariances are crucial to the asymptotic performance of the procedure since they govern the magnitude of the effect that an observation of (s, x) can have on other state-action pairs (this issue also arises in the analysis of conjugate models; see [71]).

Assumption 2.5.3. *The prior covariance matrix* $\mathbf{\Sigma}_0$ *satisfies*

$$\begin{aligned} \mathbf{\Sigma}_0\left(\left(s,x\right),\left(s,x\right)\right) &> 0, \ \forall\ (s,x), \\ \left|\mathbf{\Sigma}_0\left(\left(s,x\right),\left(s',x'\right)\right)/\mathbf{\Sigma}_0\left(\left(s,x\right),\left(s,x\right)\right)\right| &\leq \sqrt{\delta}, \ \forall\ \ (s,x) \neq \left(s',x'\right), \end{aligned}$$

where $\delta \in [0,1)$ *is a constant.*

Given the state-action pair (s_n, x_n) visited in the nth stage, we propose the following update

$$\begin{aligned}
\bar{V}_{n+1}\left(s_n, x_n\right) &= \bar{V}_n\left(s_n, x_n\right) - \frac{1}{n+2}\left(\bar{V}_n\left(s_n, x_n\right) - v_{n+1}\right), && (2.55)\\
\bar{V}_{n+1}(s, x) &= \bar{V}_n(s, x)\\
&\quad - \frac{1}{n+2}\frac{\mathbf{\Sigma}_n\left(\left(s, x\right), \left(s_n, x_n\right)\right)}{\mathbf{\Sigma}_n\left(\left(s_n, x_n\right), \left(s_n, x_n\right)\right)}\\
&\qquad \times \left(\bar{V}_n\left(s_n, x_n\right) - v_{n+1}\right), && (2.56)\\
\mathbf{\Sigma}_{n+1}\left(\left(s_n, x_n\right), \left(s_n, x_n\right)\right) &= \frac{n+1}{n+2}\mathbf{\Sigma}_n\left(\left(s_n, x_n\right), \left(s_n, x_n\right)\right), && (2.57)\\
\mathbf{\Sigma}_{n+1}\left(\left(s, x\right), \left(s_n, x_n\right)\right) &= \frac{n+1}{n+2}\mathbf{\Sigma}_n\left(\left(s, x\right), \left(s_n, x_n\right)\right), && (2.58)\\
\mathbf{\Sigma}_{n+1}\left(\left(s, x\right), \left(s, x\right)\right) &= \mathbf{\Sigma}_n\left(\left(s, x\right), \left(s, x\right)\right)\\
&\quad - \frac{\mathbf{\Sigma}_n\left(\left(s, x\right), \left(s_n, x_n\right)\right)\mathbf{\Sigma}_n\left(\left(s_n, x_n\right), \left(s, x\right)\right)}{(n+2)\mathbf{\Sigma}_n\left(\left(s_n, x_n\right), \left(s_n, x_n\right)\right)}, && (2.59)\\
\tilde{\mathbf{\Sigma}}_{n+1}\left(\left(s, x\right), \left(s', x'\right)\right) &= \mathbf{\Sigma}_n\left(\left(s, x\right), \left(s', x'\right)\right)\\
&\quad - \frac{\mathbf{\Sigma}_n\left(\left(s, x\right), \left(s_n, x_n\right)\right)\mathbf{\Sigma}_n\left(\left(s_n, x_n\right), \left(s', x'\right)\right)}{(n+2)\mathbf{\Sigma}_n\left(\left(s_n, x_n\right), \left(s_n, x_n\right)\right)},\\
\mathbf{\Sigma}_{n+1}\left(\left(s, x\right), \left(s', x'\right)\right) &= \mathrm{sgn}\left(\tilde{\mathbf{\Sigma}}_{n+1}\left(\left(s, x\right), \left(s', x'\right)\right)\right)\\
&\quad \times \min\left\{\left|\frac{\mathbf{\Sigma}_{n+1}\left(\left(s', x'\right), \left(s', x'\right)\right)\mathbf{\Sigma}_n\left(\left(s, x\right), \left(s', x'\right)\right)}{\mathbf{\Sigma}_n\left(\left(s', x'\right), \left(s', x'\right)\right)}\right|,\right.\\
&\qquad \left|\frac{\mathbf{\Sigma}_{n+1}\left(\left(s, x\right), \left(s, x\right)\right)\mathbf{\Sigma}_n\left(\left(s, x\right), \left(s', x'\right)\right)}{\mathbf{\Sigma}_n\left(\left(s, x\right), \left(s, x\right)\right)}\right|,\\
&\qquad \left.\left|\tilde{\mathbf{\Sigma}}_{n+1}\left(\left(s, x\right), \left(s', x'\right)\right)\right|\right\}, && (2.60)
\end{aligned}$$

for $(s, x) \neq (s', x') \neq (s_n, x_n)$, with $\mathrm{sgn}(x)$ being the sign function that equals zero if x equals zero and $x/|x|$ otherwise. These equations are mostly identical to (2.52)-(2.54), with the exception of (2.60), which is slightly modified to ensure that the absolute values of the ratios of the off-diagonal entries to the diagonal entries of $\mathbf{\Sigma}_n$

are decreasing in n and satisfy

$$\sup_{n\in\mathbb{N},\forall(s,x)\neq(s',x')}\left|\frac{\mathbf{\Sigma}_n\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_n\left((s,x),(s,x)\right)}\right|\leq\sqrt{\delta}. \tag{2.61}$$

This modification is needed to handle some technical issues in the convergence proof. We note, however, that the modified update is not much harder to implement than the original one, and there would be little difference to a practitioner looking to use an approximate Bayesian method for its practical benefits.

Let $I_n(s,x)$ be a binary variable that equals 1 if $(s_n,x_n)=(s,x)$ and zero otherwise, and define $T_n(s,x)\triangleq\sum_{t=0}^{n}I_t(s,x)$ to be the number of visits to (s,x) by time n. Two more assumptions are imposed: Assumption 2.5.4 is trivially satisfied for a finite state and action space, while Assumption 2.5.5 is a regularity condition requiring sufficient exploration of each state-action pair.

Assumption 2.5.4.

$$\sup_{\forall(s,x)}|C(s,x)|\leq C^*. \tag{2.62}$$

Assumption 2.5.5. *For every state-action pair* (s,x)*,*

$$\frac{T_n(s,x)+1}{n+1}\geq\lambda,\ \forall\ n\in\mathbb{N}, \tag{2.63}$$

where $\lambda\in(0,1)$ *is a constant.*

Finally, we use a projected version of (2.55)-(2.56) given by

$$\bar{V}_{n+1}(s,x)=\Pi_H\left(\bar{V}_n(s,x)-\frac{1}{n+2}\frac{\mathbf{\Sigma}_n\left((s,x),(s_n,x_n)\right)}{\mathbf{\Sigma}_n\left((s_n,x_n),(s_n,x_n)\right)}\left(\bar{V}_n\left(s_n,x_n\right)-v_{n+1}\right)\right), \tag{2.64}$$

where K is the cardinality of $\mathcal{S} \times \mathcal{X}$ and $H = [-M, M]^K$ with M taken to be large enough such that $\bar{V}_0, V \in H$. We prove that (2.64) is consistent. The proof integrates Theorem 2.3.3 with the theoretical approach of [65].

Theorem 2.5.2. *Suppose Assumptions 2.5.3-2.5.5 hold, and $\bar{V}_n$ is updated using (2.64), while $\mathbf{\Sigma}_n$ is updated using (2.57)-(2.60). Then $\bar{V}_n \to V$ almost surely.*

Proof. Without loss of generality, let $V = 0$. Define

$$\mathcal{F}_n \triangleq \mathcal{B}(v_1, ..., v_n, \bar{V}_0, ..., \bar{V}_n, \mathbf{\Sigma}_0, ..., \mathbf{\Sigma}_n),$$

fix an arbitrary state-action pair (s, x) and define

$$\begin{aligned} Q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) &= \left(\bar{V}_n(s,x) - \mathbb{E}\left(v_{n+1} \,|\, \mathcal{F}_n\right)\right) I_n(s,x), \\ q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) &= \left(\mathbb{E}\left(v_{n+1} \,|\, \mathcal{F}_n\right) - v_{n+1}\right) I_n(s,x), \\ \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) &= \sum_{(s',x') \neq (s,x)} \left(\frac{\mathbf{\Sigma}_n((s,x),(s',x'))}{\mathbf{\Sigma}_n((s',x'),(s',x'))} \left(\bar{V}_n(s',x') - v_{n+1}\right) I_n(s',x') \right). \end{aligned}$$

Then, keeping (s, x) fixed, (2.64) can be rewritten as

$$\bar{V}_{n+1}(s,x) = \Pi_H \left(\bar{V}_n(s,x) - \frac{Q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \right).$$

We first show the convergence of

$$W_{n+1}(s,x) \triangleq W_n(s,x) - \frac{W_n(s,x) I_n(s,x) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2}$$

to zero, which will be needed for the convergence of $\bar{V}_n$. This auxiliary technical lemma is proved right after the current proof.

Lemma 2.5.2. *Suppose Assumptions 2.5.3-2.5.5 hold, and $\bar{V}_n$ is updated using (2.64), while $\mathbf{\Sigma}_n$ is updated using (2.57)-(2.60). Then, $W_n(s,x) \to 0$ a.s.*

We will now use Lemma 2.5.2 to show the convergence of $\bar{V}_n$. For any $n_0 \geq 0$, define $W_{n_0;n_0}(s,x) \triangleq 0$ and

$$W_{n+1;n_0}(s,x) \triangleq W_{n;n_0}(s,x) - \frac{W_{n;n_0}(s,x) I_n(s,x) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2}$$

for $n \geq n_0$. Then, combining Lemma 2 in [65] with Lemma 2.5.2 above, it follows that, for every $\mu > 0$, there exists some positive integer N_2 such that

$$|W_{n;n_0}(s,x)| \leq \mu \tag{2.65}$$

for all $n_0 \geq N_2$ and $n \geq n_0$.

We now use an induction argument resembling that of [65]. Since $\sup \left|\bar{V}_n(s,x)\right| \leq M$, there exists some positive constant D_0 such that $\|\bar{V}_n\|_\infty \leq D_0$ for all n. Because $\gamma \in (0,1)$, we can take some small enough $\rho \in (0,1)$ such that $\gamma(1+3\rho) < 1$. Let $D_{k+1} = \gamma(1+3\rho)D_k$. Then, it is obvious that $D_k \to 0$, as $k \to \infty$.

Now suppose there exists some positive integer n_k such that $\|\bar{V}_n\|_\infty \leq D_k$ for all $n \geq n_k$. By (2.65), we can choose $\tau_k \geq n_k$ such that

$$|W_{n;\tau_k}(s,x)| \leq \gamma\rho D_k$$

for all (s,x) and all $n \geq \tau_k$. For $n \geq \tau_k$, define $Y_{\tau_k}(s,x) \triangleq D_k$ and

$$Y_{n+1}(s,x) \triangleq Y_n(s,x) - \frac{1}{n+2}\left(Y_n(s,x) - \gamma D_k\right) I_n(s,x). \tag{2.66}$$

Since $\gamma \in (0,1)$, it is obvious that $(Y_n(s,x))$ is a decreasing sequence with respect to n and

$$\lim_{n\to\infty} Y_n(s,x) = \gamma D_k.$$

Furthermore, since we have $Y_n(s,x) \geq \gamma D_k$ and $|W_{n;\tau_k}(s,x)| \leq \gamma \rho D_k$ (with $0 < \rho < 1$) for all $n \geq \tau_k$, it follows that

$$Y_n(s,x) + W_{n;\tau_k}(s,x) \geq 0, \qquad -Y_n(s,x) + W_{n;\tau_k}(s,x) \leq 0 \tag{2.67}$$

for all $n \geq \tau_k$.

Let

$$F\left(\bar{V}_n, s_n, x_n\right) \triangleq \mathbb{E}\left(v_{n+1} \,|\, \mathcal{F}_n\right) = C\left(s_n, x_n\right) + \gamma \mathbb{E}\left(\max_x \bar{V}_n\left(s_{n+1}, x\right) \,|\, \mathcal{F}_n\right).$$

For any V', V'',

$$\begin{aligned} \left|F\left(V', s_n, x_n\right) - F\left(V'', s_n, x_n\right)\right| &\triangleq \gamma \left|\mathbb{E}\left(\max_x V'\left(s_{n+1}, x\right) - \max_x V''\left(s_{n+1}, x\right)\right)\right| \\ &\leq \gamma \max_{(s,x) \in \mathcal{S} \times \mathcal{X}} \left|V'(s,x) - V''(s,x)\right|, \end{aligned}$$

whence $F\left(V', s_n, x_n\right)$ is a contraction mapping of V' with respect to the maximum norm $\|\cdot\|_\infty$. By Banach's fixed-point theorem, F has a unique fixed point, and from the definition of F, the fixed point is the true value function V, which was assumed at the beginning to equal zero. Hence we have

$$\left|F\left(V', s_n, x_n\right)\right| = \left\|F\left(V', s_n, x_n\right)\right\|_\infty \leq \gamma \left\|V'\right\|_\infty, \qquad \forall\, V' \in \mathbb{R}^K. \tag{2.68}$$

Now, suppose that $-Y_n(s,x) + W_{n;\tau_k}(s,x) \leq \bar{V}_n(s,x) \leq Y_n(s,x) + W_{n;\tau_k}(s,x)$ holds for some $n \geq \tau_k$. Then,

$$\begin{aligned} &\bar{V}_n(s,x) - \frac{Q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \\ &\quad = \bar{V}_n(s,x) - \frac{\left(\bar{V}_n(s,x) - \mathbb{E}\left(v_{n+1} \,|\, \mathcal{F}_n\right)\right) I_n(s,x) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \\ &\quad = \bar{V}_n(s,x) - \frac{\bar{V}_n(s,x) - \mathbb{E}\left(v_{n+1} \,|\, \mathcal{F}_n\right)}{n+2} I_n(s,x) - \frac{q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \end{aligned}$$

$$
\begin{aligned}
&= \bar{V}_n(s,x) - \frac{\bar{V}_n(s,x)}{n+2} I_n(s,x) \\
&\quad + \frac{F(\bar{V}_n, s_n, x_n)}{n+2} I_n(s,x) - \frac{q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \\
&\leq \bar{V}_n(s,x) - \frac{\bar{V}_n(s,x)}{n+2} I_n(s,x) \\
&\quad + \frac{\gamma \left\|\bar{V}_n\right\|_\infty}{n+2} I_n(s,x) - \frac{q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \\
&\leq \bar{V}_n(s,x)\left(1 - \frac{1}{n+2} I_n(s,x)\right) \\
&\quad + \frac{\gamma D_k}{n+2} I_n(s,x) - \frac{q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \\
&\leq (Y_n(s,x) + W_{n;\tau_k}(s,x))\left(1 - \frac{1}{n+2} I_n(s,x)\right) \\
&\quad + \frac{\gamma D_k}{n+2} I_n(s,x) - \frac{q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \\
&= Y_n(s,x) - \frac{1}{n+2}\left(Y_n(s,x) - \gamma D_k\right) I_n(s,x) \\
&\quad + W_{n;\tau_k}(s,x) - \frac{W_{n;\tau_k}(s,x) I_n(s,x) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2} \\
&= Y_{n+1}(s,x) + W_{n+1;\tau_k}(s,x),
\end{aligned}
$$

where the first inequality holds because of (2.68). Together with (2.67), this implies that

$$
\begin{aligned}
\bar{V}_{n+1}(s,x) &= \Pi_H\left(\bar{V}_n(s,x) - \frac{Q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) + \beta_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1})}{n+2}\right) \\
&\leq Y_{n+1}(s,x) + W_{n+1;\tau_k}(s,x).
\end{aligned}
$$

Using a symmetrical argument, we can show that $\bar{V}_{n+1}(s,x) \geq -Y_{n+1}(s,x) + W_{n+1;\tau_k}(s,x)$. Thus, we have $-Y_{n+1}(s,x) + W_{n+1;\tau_k}(s,x) \leq \bar{V}_{n+1}(s,x) \leq Y_{n+1}(s,x) + W_{n+1;\tau_k}(s,x)$. When $n = \tau_k$, we have $Y_{\tau_k}(s,x) = D_k$ and $W_{\tau_k;\tau_k}(s,x) = 0$, hence

$$
-Y_n(s,x) + W_{n;\tau_k}(s,x) \leq \bar{V}_n(s,x) \leq Y_n(s,x) + W_{n;\tau_k}(s,x)
$$

holds for $n = \tau_k$. By induction, we have

$$-Y_n(s,x) + W_{n;\tau_k}(s,x) \leq \bar{V}_n(s,x) \leq Y_n(s,x) + W_{n;\tau_k}(s,x) \tag{2.69}$$

for all $n \geq \tau_k$.

Since $Y_n(s,x) \to \gamma D_k$ and $|W_{n;\tau_k}(s,x)| \leq \gamma\rho D_k$ for all $n \geq \tau_k$, (2.69) implies that

$$\limsup_{n\to\infty} \left|\bar{V}_n(s,x)\right| \leq \gamma(1+2\rho)D_k < D_{k+1}$$

for every state-action pair (s,x). Hence, there exists some positive integer n_{k+1} such that $\|\bar{V}_n\|_\infty \leq D_{k+1}$ for all $n \geq n_{k+1}$. Thus by induction, we conclude that for every k, there exists some positive integer n_k such that

$$\|\bar{V}_n\|_\infty \leq D_k$$

for all $n \geq n_k$. Since $D_k \to 0$ as $k \to \infty$, we have $V_n \to 0$, as required. □

Proof of Lemma 2.5.2. We introduce the additional notation

$$\mathcal{A}(s,x) \triangleq \{n : I_n(s,x) = 1\}$$

and rank the elements of $\mathcal{A}(s,x)$ in ascending order to get an increasing sequence $(\zeta_n(s,x))$. That is, $(\zeta_n(s,x))$ is the sequence of time indices for which we are in state s and choose action x.

Let

$$\begin{aligned} R_n\left(W_n(s,x)\right) &\triangleq \mathbb{E}(W_n(s,x)I_n(s,x) + q_n(\bar{V}_n, \mathbf{\Sigma}_n, v_{n+1}) \,|\, \mathcal{F}_n) \\ &= \mathbb{E}(W_n(s,x)I_n(s,x) \,|\, \mathcal{F}_n) + \mathbb{E}((\mathbb{E}\left(v_{n+1} \,|\, \mathcal{F}_n\right) - v_{n+1})\, I_n(s,x) \,|\, \mathcal{F}_n) \\ &= W_n(s,x)I_n(s,x), \end{aligned}$$

where the last equality holds since $W_n(s,x)$ is also $\mathcal{F}_n$-measurable. Then, for all $n \in \mathcal{A}(s,x)$, we have $R_n(z) = z$, hence $R_n(z) = 0$ if and only if $z = 0$, whence Assumption 2.3.1 is verified. Assumption 2.3.2 is verified straightforwardly.

From (2.64), we know that $\bar{V}_n$ is uniformly bounded in n. Together with Assumption 2.5.4, this implies the existence of a positive constant C_1 satisfying

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left(\left(zI_n(s,x) + q_n(\bar{V}_n, \boldsymbol{\Sigma}_n, v_{n+1})\right)^2 + (\beta_n(\bar{V}_n, \boldsymbol{\Sigma}_n, v_{n+1}))^2 \,|\, \mathcal{F}_n\right) \;\leq\; C_1\left(1+z^2\right)$$

for all z. Therefore, in order to apply Theorem 2.3.3, it only remains to show that the condition (2.18) in Assumption 2.3.4 is satisfied. Due to the boundedness of $\bar{V}_n$ and C, it is sufficient to show that

$$\sum_{n=0}^{\infty}\left(\frac{1}{n+2}\sum_{(s',x')\neq(s,x)}\left(\left|\frac{\boldsymbol{\Sigma}_n((s,x),(s',x'))}{\boldsymbol{\Sigma}_n((s',x'),(s',x'))}\right| I_n(s',x')\right)\right) < \infty. \tag{2.70}$$

Define

$$\xi_n \;\triangleq\; \frac{1}{n+2}\sum_{(s',x')\neq(s,x)}\left(\left|\frac{\boldsymbol{\Sigma}_n((s,x),(s',x'))}{\boldsymbol{\Sigma}_n((s',x'),(s',x'))}\right| I_n(s',x')\right).$$

Then, by Kolmogorov's three-series theorem [61], it is sufficient to show the convergence of the three series

$$\sum P(|\xi_n| \geq c \,|\, \mathcal{F}_{n-1}), \quad \sum \mathbb{E}(\xi_n 1_{\{|\xi_n|\leq c\}} \,|\, \mathcal{F}_{n-1}), \quad \sum Var(\xi_n 1_{\{|\xi_n|\leq c\}} \,|\, \mathcal{F}_{n-1}),$$

where c is some positive constant.

From (2.61), by Chebyshev's inequality, we have

$$\sum P(|\xi_n| \geq c \,|\, \mathcal{F}_{n-1}) \leq \sum \frac{\delta}{c^2(n+2)^2} < \infty,$$

so the first series converges. Similarly, we can see that

$$\sum Var(\xi_n 1_{\{|\xi_n|\leq c\}} \,|\, \mathcal{F}_{n-1}) \leq \sum \mathbb{E}\left((\xi_n)^2 \,|\, \mathcal{F}_{n-1}\right) \leq \sum \frac{\delta}{(n+2)^2} < \infty,$$

so the last series converges. It remains to show that the second series also converges.

Finally, we show the convergence of the second series. From (2.57)-(2.60), first we can see that for all $(s',x') \neq (s,x)$, the ratio $\left|\frac{\mathbf{\Sigma}_n((s,x),(s',x'))}{\mathbf{\Sigma}_n((s',x'),(s',x'))}\right|$ is decreasing in n. Now, if (s,x) is the state-action pair observed at the nth stage, we have

$$\begin{aligned}
\mathbb{E}(\xi_{n+1}1_{\{|\xi_{n+1}|\leq c\}} \,|\, \mathcal{F}_n) &\leq \mathbb{E}(\xi_{n+1} \,|\, \mathcal{F}_n) \\
&= \frac{1}{n+3} \sum_{(s',x')\neq(s,x)} \mathbb{E}\left(\left|\frac{\mathbf{\Sigma}_{n+1}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{n+1}\left((s',x'),(s',x')\right)}\right| I_n\left(s',x'\right) \,|\, \mathcal{F}_n\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left(\left|\frac{\mathbf{\Sigma}_{n+1}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{n+1}\left((s',x'),(s',x')\right)}\right| \,|\, \mathcal{F}_n\right) &\leq \mathbb{E}\left(\frac{\left|\mathbf{\Sigma}_n\left((s,x),(s',x')\right)\right|(1-1/(n+2))}{\mathbf{\Sigma}_n\left((s',x'),(s',x')\right)(1-\delta/(n+2))} \,|\, \mathcal{F}_n\right) \\
&= \left|\frac{\mathbf{\Sigma}_n\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_n\left((s',x'),(s',x')\right)}\right| \frac{1-1/(n+2)}{1-\delta/(n+2)}, \qquad (2.71)
\end{aligned}$$

where the inequality holds because of (2.58), (2.59) and (2.61). Since $0 \leq \delta < 1$, there exists a large enough integer N such that

$$N \geq \frac{1}{\lambda(1-\delta)}.$$

Then, for any n, we have

$$\begin{aligned}
1 + \frac{1}{T_n(s,x)+1} &\leq 1 + \frac{1}{\lambda(n+1)} \\
&\leq 1 + N\frac{1-\delta}{n+1} \\
&\leq \left(1 + \frac{1-\delta}{n+1}\right)^N \\
&= \left(\frac{1-\frac{\delta}{n+2}}{1-\frac{1}{n+2}}\right)^N,
\end{aligned}$$

where the first inequality holds because of (2.63). Consequently,

$$\frac{1-\frac{1}{n+2}}{1-\frac{\delta}{n+2}} \leq \left(\frac{T_n(s,x)+1}{T_n(s,x)+2}\right)^{\frac{1}{N}},$$

whence (2.71) becomes

$$\mathbb{E}\left(\left|\frac{\mathbf{\Sigma}_{n+1}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{n+1}\left((s',x'),(s',x')\right)}\right| \mid \mathcal{F}_n\right) \leq \left|\frac{\mathbf{\Sigma}_{n}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{n}\left((s',x'),(s',x')\right)}\right| \left(\frac{T_n(s,x)+1}{T_n(s,x)+2}\right)^{\frac{1}{N}}.$$

We can take a large enough integer N_1 such that (s,x) is the state-action pair observed at stage N_1 and, for all $n \geq N_1$,

$$\begin{aligned}
&\mathbb{E}\left(\left|\frac{\mathbf{\Sigma}_{n+1}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{n+1}\left((s',x'),(s',x')\right)}\right| \mid \mathcal{F}_n\right)\\
\leq\ & \left|\frac{\mathbf{\Sigma}_{n}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{n}\left((s',x'),(s',x')\right)}\right| \left(\frac{T_n(s,x)+1}{T_n(s,x)+2}\right)^{\frac{1}{N}}\\
\leq\ & \left|\frac{\mathbf{\Sigma}_{N_1}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{N_1}\left((s',x'),(s',x')\right)}\right| \left(\frac{T_{N_1}(s,x)+1}{T_{N_1}(s,x)+2}\right)^{\frac{1}{N}} \cdots \left(\frac{T_n(s,x)+1}{T_n(s,x)+2}\right)^{\frac{1}{N}}\\
=\ & \left|\frac{\mathbf{\Sigma}_{N_1}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{N_1}\left((s',x'),(s',x')\right)}\right| \left(\frac{T_{N_1}(s,x)+1}{T_n(s,x)+2}\right)^{\frac{1}{N}}\\
\leq\ & \left|\frac{\mathbf{\Sigma}_{N_1}\left((s,x),(s',x')\right)}{\mathbf{\Sigma}_{N_1}\left((s',x'),(s',x')\right)}\right| \left(\frac{N_1+2}{\lambda(n+1)+1}\right)^{\frac{1}{N}}\\
\leq\ & \sqrt{\delta}\left(\frac{N_1+2}{\lambda(n+1)+1}\right)^{\frac{1}{N}}.
\end{aligned}$$

It follows that

$$\sum_{n\geq N_1} \mathbb{E}(\xi_{n+1}1_{\{|\xi_{n+1}|\leq c\}} \mid \mathcal{F}_n) \leq \sum_{n\geq N_1} \frac{\sqrt{\delta}}{n+3}\left(\frac{N_1+2}{\lambda(n+1)+1}\right)^{\frac{1}{N}} < \infty,$$

proving the convergence of the second series. Therefore, (2.70) holds and, by Theorem 2.3.3, $\lim_{n\to\infty}\left(W_{\zeta_n(s,x)}(s,x)\right)^2$ exists and is finite, and

$$\begin{aligned}
\sum_{n=1}^{\infty}\frac{1}{n+2}W_n(s,x)R_n\left(W_n(s,x)\right) &= \sum_{n=1}^{\infty}\frac{1}{n+2}\left(W_n(s,x)\right)^2 I_n(s,x)\\
&= \sum_{n=1}^{\infty}\frac{1}{\zeta_n(s,x)+2}\left(W_{\zeta_n(s,x)}(s,x)\right)^2\\
&< \infty
\end{aligned}$$

almost surely. Then, from (2.63), this implies

$$\lim_{n\to\infty} W_{\zeta_n(s,x)}(s,x) = 0.$$

Furthermore, from $\sup \left|\bar{V}_n(s,x)\right| \leq M$ and Assumption 2.5.4, there must exist some positive constant C_2 such that

$$\sup_{n\in\mathbb{N},\forall(s,x)} \left|\bar{V}_n(s,x) - v_{n+1}\right| \leq C_2.$$

Together with (2.70), this implies that $W_n(s,x) \to 0$ for all (s,x), as required. □

To our knowledge, Theorem 2.5.2 is the first consistency result for a correlated Bayesian belief model in the setting of approximate value iteration, where statistical estimation takes place simultaneously with policy optimization, represented by the max operator in (2.50). While [72, 73] have studied Gaussian process priors in dynamic programming, this work dealt with the much simpler problem (from a statistical perspective) of learning the value of a *fixed* policy. Despite the richness of the dynamic programming literature, convergence results for approximate value iteration tend to be much more difficult to obtain.

2.5.3 Ranking and Selection with Unknown Correlation Structures

Ranking and selection is a fundamental problem class in the simulation literature [74] that provides a mathematical framework for the study of information collection. We suppose that there are K design alternatives with unknown values $\theta^{(1)}, ..., \theta^{(K)}$, and that our goal is to identify $\arg\max_i \theta^{(i)}$ based on information collected from a limited number of simulation experiments with individual alternatives. Bayesian statistical models are widely used in this literature [75] because they offer a way to express our uncertainty about the unknown values and quantify how this uncertainty evolves as more information is collected. Much of the research in this area uses simple, conjugate Bayesian models and focuses on the development of procedures for efficient allocation of the budget [76, 77].

Suppose that we use a multivariate distribution to model our beliefs about $\theta = \left(\theta^{(1)}, ..., \theta^{(K)}\right)$. If, for two designs $i \neq j$, the prior includes correlations between $\theta^{(i)}$ and $\theta^{(j)}$, a single simulation experiment with design i will also provide information about design j. With sufficient correlation in the prior, we will be able to learn about many alternatives from a much smaller number of simulations. For this reason, correlated beliefs have a great deal of practical potential [11]; however, the drawback is that prior correlations are even more difficult to specify accurately than prior means. Approximate Bayesian models become useful here as a possible tool for learning both the means and the correlations [5]. In the following, we give a new analysis of a modified version of the approximate Bayesian procedure proposed by [38, 78].

Let $(Y_n)_{n=1}^{\infty}$ be a sequence of i.i.d. samples from the K-dimensional multivariate normal distribution $\mathcal{N}_K(\theta, \mathbf{\Sigma})$, where both θ and $\mathbf{\Sigma}$ are unknown. We impose the prior

$$\theta | \mathbf{\Sigma} \sim \mathcal{N}_K\left(\mu_0, q_0^{-1}\mathbf{\Sigma}\right), \qquad \mathbf{\Sigma} \sim \mathcal{W}_K^{-1}\left(\mathbf{B}_0, b_0\right).$$

Here, $\mathbf{\Sigma}$ follows an inverse Wishart distribution [79] with b_0 degrees of freedom and scale matrix $\mathbf{B}_0$. The conditional distribution of θ given $\mathbf{\Sigma}$ is multivariate normal with mean vector μ_0 and covariance matrix $q_0^{-1}\mathbf{\Sigma}$. It is well-known that, if the complete vectors (Y_n) can be observed, the above model is conjugate [49]. However, suppose that we can only observe one element of Y_n during the nth stage of sampling, for instance the kth element $Y_n^{(k)}$. In this case, the normal-inverse-Wishart prior is not conjugate with the scalar normal observation, an issue that we address using approximate Bayesian inference.

The sequence $(q_n, b_n, \mu_n, \mathbf{B}_n)$ of approximate posterior parameters is constructed as follows. First, we let

$$\begin{aligned} q_n &= n+1, && (2.72) \\ b_n &= n+K+1. && (2.73) \end{aligned}$$

Suppose that $Y_{n+1}^{(k)}$ is the observation collected in the $(n+1)$st stage of sampling (i.e., only the kth component of Y_{n+1} is observable). Then, we use the update

$$\mu_{n+1} = \mu_n - \frac{\mathbf{B}_n^{(\cdot,k)}}{\mathbf{B}_n^{(k,k)}} \frac{\mu_n^{(k)} - Y_{n+1}^{(k)}}{n+2}. \qquad (2.74)$$

Equations (2.72)-(2.74) are taken from [38]. In (2.74), we have already substituted (2.72) for q_n to simplify the computation.

It remains to set an update for $\mathbf{B}_n$. We first impose some assumptions on the starting prior $\mathbf{B}_0$, as in Section 2.5.2.

Assumption 2.5.6. *The prior scale matrix* $\mathbf{B}_0$ *satisfies*

$$\begin{aligned} \mathbf{B}_0^{(k,k)} &\geq L, \ \forall\ 1 \leq k \leq K, \\ \left| \mathbf{B}_0^{(j,k)} / \mathbf{B}_0^{(k,k)} \right| &\leq \sqrt{1-\delta}, \ \forall\ 1 \leq j \neq k \leq K, \end{aligned}$$

where $L > 0$ *and* $\delta \in \left(\frac{1}{2}, 1\right]$ *are constants.*

If only the kth element of Y_{n+1} is observable in the $(n+1)$st stage, we propose the update

$$\mathbf{B}_{n+1}^{(k,k)} = \max\left\{ \mathbf{B}_n^{(k,k)} + \frac{n+1}{n+2}\left(\mu_n^{(k)} - Y_{n+1}^{(k)}\right)^2, L(n+2) \right\}, \qquad (2.75)$$

$$\begin{aligned}
\mathbf{B}^{(j,j)}_{n+1} &= \mathbf{B}^{(j,j)}_{n} + \frac{2}{n+1}\left(\mathbf{B}^{(j,j)}_{n} - \frac{\mathbf{B}^{(j,k)}_{n}\mathbf{B}^{(k,j)}_{n}}{\mathbf{B}^{(k,k)}_{n}}\right) \\
&\quad + \frac{\left(\mu^{(k)}_{n} - Y^{(k)}_{n+1}\right)^2}{(n+2)\mathbf{B}^{(k,k)}_{n}}\left(\mathbf{B}^{(j,j)}_{n} + n\frac{\mathbf{B}^{(j,k)}_{n}\mathbf{B}^{(k,j)}_{n}}{\mathbf{B}^{(k,k)}_{n}}\right), && (2.76) \\
\mathbf{B}^{(j,k)}_{n+1} &= \operatorname{sgn}\left(\mathbf{B}^{(j,k)}_{n}\right) \cdot \min\left\{\left|\frac{\mathbf{B}^{(j,k)}_{n}}{\mathbf{B}^{(k,k)}_{n}}\mathbf{B}^{(k,k)}_{n+1}\right|, \left|\frac{\mathbf{B}^{(j,k)}_{n}}{\mathbf{B}^{(j,j)}_{n}}\mathbf{B}^{(j,j)}_{n+1}\right|\right\}, && (2.77) \\
\tilde{\mathbf{B}}^{(j,i)}_{n+1} &= \mathbf{B}^{(j,i)}_{n} + \frac{2}{n+1}\left(\mathbf{B}^{(j,i)}_{n} - \frac{\mathbf{B}^{(j,k)}_{n}\mathbf{B}^{(k,i)}_{n}}{\mathbf{B}^{(k,k)}_{n}}\right) \\
&\quad + \frac{\left(\mu^{(k)}_{n} - Y^{(k)}_{n+1}\right)^2}{(n+2)\mathbf{B}^{(k,k)}_{n}}\left(\mathbf{B}^{(j,i)}_{n} + n\frac{\mathbf{B}^{(j,k)}_{n}\mathbf{B}^{(k,i)}_{n}}{\mathbf{B}^{(k,k)}_{n}}\right), && (2.78) \\
\mathbf{B}^{(j,i)}_{n+1} &= \operatorname{sgn}\left(\tilde{\mathbf{B}}^{(j,i)}_{n+1}\right) \cdot \min\left\{\left|\tilde{\mathbf{B}}^{(j,i)}_{n+1}\right|, \left|\frac{\mathbf{B}^{(j,i)}_{n}}{\mathbf{B}^{(j,j)}_{n}}\mathbf{B}^{(j,j)}_{n+1}\right|, \left|\frac{\mathbf{B}^{(j,i)}_{n}}{\mathbf{B}^{(i,i)}_{n}}\mathbf{B}^{(i,i)}_{n+1}\right|\right\}, && (2.79)
\end{aligned}$$

for $i \neq j \neq k$. This update is based on the moment-matching mechanism from [38]; in particular, (2.76) is taken directly from that work (substituting (2.72) and (2.73) for q_n and b_n), while (2.78) is the moment-matching update for $\mathbf{B}^{(j,i)}_{n+1}$, and the first term inside the maximum in (2.75) is the moment-matching update for $\mathbf{B}^{(k,k)}_{n+1}$. The additional modifications that we have introduced are intended to handle technical issues, as in Section 2.5.2: note that, from (2.77) and (2.79), it follows that the absolute values of the ratios of the off-diagonal entries to the diagonal entries of $\mathbf{B}_n$ are decreasing in n. From Assumption 2.5.6, it also follows that

$$\sup_{n\in\mathbb{N},\forall j\neq k}\left|\frac{\mathbf{B}^{(j,k)}_{n}}{\mathbf{B}^{(k,k)}_{n}}\right| \leq \sqrt{1-\delta}. \tag{2.80}$$

Furthermore, from (2.76), we can see that

$$\begin{aligned}
\mathbf{B}^{(j,j)}_{n+1} &\geq \mathbf{B}^{(j,j)}_{n}\left(1 + \frac{2}{n+1}\left(1 - \frac{\mathbf{B}^{(j,k)}_{n}\mathbf{B}^{(k,j)}_{n}}{\mathbf{B}^{(j,j)}_{n}\mathbf{B}^{(k,k)}_{n}}\right)\right) \\
&\geq \mathbf{B}^{(j,j)}_{n}\left(1 + \frac{2}{n+1}\delta\right) \geq \mathbf{B}^{(j,j)}_{n}\left(1 + \frac{1}{n+1}\right),
\end{aligned}$$

which ensures, together with (2.75), that $\mathbf{B}_n^{(k,k)} \geq L\,(n+1)$ for all k. These modifications do not make the update much harder to implement, and would make little difference to a practitioner.

We now present a convergence result for the approximate Bayesian update in (2.72)-(2.79). Some last preliminary notation and assumptions are needed. Define $I_n^{(k)}$ to be a binary variable that equals 1 if the kth element is simulated at the nth stage and zero otherwise. Define $S_n^{(k)} \triangleq \sum_{t=0}^{n} I_t^{(k)}$ to be the number of simulations assigned to k up to time n.

Assumption 2.5.7.

$$\frac{S_n^{(k)}+1}{n+1} \geq \gamma,\ \forall\ n \in \mathbb{N},$$

where $\gamma \in (0,1]$ is a constant.

Assumption 2.5.7 essentially requires every alternative to receive a non-zero proportion of the simulation budget asymptotically. Many allocation policies satisfy this condition, including optimal computing budget allocation [76] and knowledge gradients [13].

Finally, we use a projected version of (2.74) given by

$$\mu_{n+1} = \Pi_H\left(\mu_n - \frac{\mathbf{B}_n^{(\cdot,k)}}{\mathbf{B}_n^{(k,k)}}\frac{\mu_n^{(k)} - Y_{n+1}^{(k)}}{n+2}\right), \tag{2.81}$$

where $H = [-M,M]^K$ with M taken to be large enough such that $\mu_0, \theta \in H$ (again interpreting θ as a fixed vector, as in previous examples).

Theorem 2.5.3. *Let $\sigma^2 = \sup_k Var\left(Y^{(k)}\right)$. Suppose Assumptions 2.5.6 and 2.5.7 hold with δ, M, L chosen to satisfy $2\delta L > 4M^2 + \sigma^2$. Under the projected update*

(2.81) for the posterior mean, and the update (2.75)-(2.79) for the scale matrix, we have $\mu_n \to \theta$ a.s.

Proof. Without loss of generality, let $\theta = 0$. We introduce the additional notation $\mathcal{A}^{(k)} \triangleq \left\{n : I_n^{(k)} = 1\right\}$ and rank the elements of $\mathcal{A}^{(k)}$ in ascending order to get an increasing sequence $\left(\zeta_n^{(k)}\right)$. That is, $\left(\zeta_n^{(k)}\right)$ is the sequence of time stages for which the kth element is simulated.

Define

$$\begin{aligned} Q_n(\mu_n, \mathbf{B}_n, Y_{n+1}) &= \left(\mu_n^{(k)} - Y_{n+1}^{(k)}\right) I_n^{(k)}, \\ \beta_n(\mu_n, \mathbf{B}_n, Y_{n+1}) &= \sum_{j \neq k} \left(\frac{\mathbf{B}_n^{(k,j)}}{\mathbf{B}_n^{(j,j)}} \left(\mu_n^{(j)} - Y_{n+1}^{(j)}\right) I_n^{(j)} \right), \end{aligned}$$

and rewrite (2.81) as

$$\mu_{n+1}^{(k)} = \Pi_H \left(\mu_n^{(k)} - \frac{1}{n+2} \left(Q_n(\mu_n, \mathbf{B}_n, Y_{n+1}) + \beta_n(\mu_n, \mathbf{B}_n, Y_{n+1}) \right) \right).$$

In words, if the kth alternative is simulated in the nth stage, we update our beliefs about k through Q_n. Otherwise, k is updated through the "bias" term.

Now define

$$\begin{aligned} \mathcal{F}_n &\triangleq \mathcal{B}(Y_1, ..., Y_n, \mu_0, ..., \mu_n, \mathbf{B}_0, ..., \mathbf{B}_n), \\ R_n\left(\mu_n^{(k)}\right) &\triangleq \mathbb{E}(Q_n(\mu_n, \mathbf{B}_n, Y_{n+1}) \,|\, \mathcal{F}_n) \\ &= \mu_n^{(k)} I_n^{(k)}. \end{aligned}$$

For all $n \in \mathcal{A}^{(k)}$, we have $R_n(x) = x$, whence $R_n(x) = 0$ if and only if $x = 0$, thus verifying Assumption 2.3.1. Assumption 2.3.2 is straightforward to verify.

By (2.81), we have

$$\sup_{n \in \mathbb{N}, 1 \leq k \leq K} \left|\mu_n^{(k)}\right| \leq M.$$

This, together with (2.80), implies that there exists a positive constant C_1 such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left((Q_n(x,\mathbf{B}_n,Y_{n+1}))^2 + (\beta_n(x,\mathbf{B}_n,Y_{n+1}))^2 \,|\, \mathcal{F}_n\right) \quad \leq \quad C_1$$

for all x satisfying $\sup_k \left|x^{(k)}\right| \leq M$. Therefore, in order to apply Theorem 2.3.3, it remains only to show that the condition (2.18) in Assumption 2.3.4 is satisfied. Also, since $\sup_{n,k} \left|\mu_n^{(k)}\right| \leq M$, it is sufficient to show that

$$\sum_{n=0}^{\infty} \left(\frac{1}{n+2} \sum_{j\neq k} \left(\left| \frac{\mathbf{B}_n^{(k,j)}}{\mathbf{B}_n^{(j,j)}} \right| I_n^{(j)} \right) \right) < \infty. \tag{2.82}$$

The remainder of the proof will establish (2.82).

Define

$$\xi_n \triangleq \frac{1}{n+2} \sum_{j\neq k} \left(\left| \frac{\mathbf{B}_n^{(k,j)}}{\mathbf{B}_n^{(j,j)}} \right| I_n^{(j)} \right).$$

By Kolmogorov's three-series theorem [61], it is sufficient to show the convergence of the three series

$$\sum P(|\xi_n| \geq c \,|\, \mathcal{F}_{n-1}), \quad \sum \mathbb{E}(\xi_n 1_{\{|\xi_n|\leq c\}} \,|\, \mathcal{F}_{n-1}), \quad \sum Var(\xi_n 1_{\{|\xi_n|\leq c\}} \,|\, \mathcal{F}_{n-1}),$$

where c is some positive constant. From (2.80), by Chebyshev's inequality, we have

$$\sum P(|\xi_n| \geq c \,|\, \mathcal{F}_{n-1}) \leq \sum \frac{1-\delta}{c^2(n+2)^2} < \infty,$$

so the first series converges. Similarly, we can see that

$$\sum Var(\xi_n 1_{\{|\xi_n|\leq c\}} \,|\, \mathcal{F}_{n-1}) \leq \sum \mathbb{E}\left((\xi_n)^2 \,|\, \mathcal{F}_{n-1}\right) \leq \sum \frac{1-\delta}{(n+2)^2} < \infty,$$

so the last series converges. It remains to show that the second series also converges.

Recall that, by (2.75)-(2.79), the ratios $\left|\mathbf{B}^{(k,j)}_n/\mathbf{B}^{(j,j)}_n\right|$ are decreasing in n for all $j \neq k$. If the kth element is chosen at the nth stage, we have

$$\mathbb{E}(\xi_{n+1}1_{\{|\xi_{n+1}|\leq c\}}\,|\,\mathcal{F}_n) \leq \mathbb{E}(\xi_{n+1}\,|\,\mathcal{F}_n) = \frac{1}{n+3}\sum_{j\neq k}\mathbb{E}\left(\left|\frac{\mathbf{B}^{(k,j)}_{n+1}}{\mathbf{B}^{(j,j)}_{n+1}}\right| I^{(j)}_n\,|\,\mathcal{F}_n\right),$$

and

$$\begin{aligned}\mathbb{E}\left(\left|\frac{\mathbf{B}^{(k,j)}_{n+1}}{\mathbf{B}^{(j,j)}_{n+1}}\right|\,|\,\mathcal{F}_n\right) &\leq \mathbb{E}\left(\frac{\left|\mathbf{B}^{(k,j)}_n\right|\left(1+\frac{\left(\mu^{(k)}_n-Y^{(k)}_{n+1}\right)^2}{\mathbf{B}^{(k,k)}_n}\right)}{\mathbf{B}^{(j,j)}_n\left(1+\frac{2\delta}{n+1}\right)}\,|\,\mathcal{F}_n\right)\\ &\leq \left|\frac{\mathbf{B}^{(k,j)}_n}{\mathbf{B}^{(j,j)}_n}\right|\frac{1+\frac{\mathbb{E}\left(\left(\mu^{(k)}_n-Y^{(k)}_{n+1}\right)^2|\,\mathcal{F}_n\right)}{L(n+1)}}{1+\frac{2\delta}{n+1}}\\ &\leq \left|\frac{\mathbf{B}^{(k,j)}_n}{\mathbf{B}^{(j,j)}_n}\right|\frac{1+\frac{4M^2+\sigma^2}{L(n+1)}}{1+\frac{2\delta}{n+1}},\end{aligned}\qquad(2.83)$$

where the first inequality is due to (2.76), (2.77) and (2.80), and the last inequality holds because $\left|\mu^{(k)}_n-\theta^{(k)}\right|\leq 2M$ and $\sup_k Var\left(Y^{(k)}\right)=\sigma^2$. Since $2\delta L > 4M^2+\sigma^2$, there exists a large enough integer N such that

$$N > \frac{1}{\gamma\left(2\delta-\frac{4M^2+\sigma^2}{L}\right)}.$$

Then, there exists some integer N_1 such that for all $n \geq N_1$,

$$n+1+\frac{4M^2+\sigma^2}{L} \leq N\gamma\left(2\delta-\frac{4M^2+\sigma^2}{L}\right)(n+1),$$

whence

$$\begin{aligned}1 &\leq N\gamma\frac{2\delta-\frac{4M^2+\sigma^2}{L}}{n+1+\frac{4M^2+\sigma^2}{L}}(n+1)\\ &\leq \gamma(n+1)\left(1+N\frac{2\delta-\frac{4M^2+\sigma^2}{L}}{n+1+\frac{4M^2+\sigma^2}{L}}\right)-\gamma(n+1)\end{aligned}$$

$$
\begin{aligned}
&\leq \gamma(n+1)\left(1+\frac{2\delta-\frac{4M^2+\sigma^2}{L}}{n+1+\frac{4M^2+\sigma^2}{L}}\right)^N-\gamma(n+1)\\
&= \gamma(n+1)\left(\left(\frac{n+1+2\delta}{n+1+\frac{4M^2+\sigma^2}{L}}\right)^N-1\right)\\
&\leq \left(S_n^{(k)}+1\right)\left(\left(\frac{n+1+2\delta}{n+1+\frac{4M^2+\sigma^2}{L}}\right)^N-1\right)\\
&= \left(S_n^{(k)}+1\right)\left(\left(\frac{1+\frac{2\delta}{n+1}}{1+\frac{(4M^2+\sigma^2)/L}{n+1}}\right)^N-1\right),
\end{aligned}
$$

where the last inequality holds because of Assumption 2.5.7. Hence we have

$$
\frac{1+\frac{(4M^2+\sigma^2)/L}{n+1}}{1+\frac{2\delta}{n+1}} \leq \left(\frac{S_n^{(k)}+1}{S_n^{(k)}+2}\right)^{\frac{1}{N}},
$$

and (2.83) becomes

$$
\mathbb{E}\left(\left|\frac{\mathbf{B}_{n+1}^{(k,j)}}{\mathbf{B}_{n+1}^{(j,j)}}\right| \,|\, \mathcal{F}_n\right) \leq \left|\frac{\mathbf{B}_{n}^{(k,j)}}{\mathbf{B}_{n}^{(j,j)}}\right| \left(\frac{S_n^{(k)}+1}{S_n^{(k)}+2}\right)^{\frac{1}{N}}.
$$

We can take a large enough integer N_2 such that $N_2 \geq N_1$ and the kth element is chosen at stage N_2. Then, for all $n \geq N_2$,

$$
\begin{aligned}
\mathbb{E}\left(\left|\frac{\mathbf{B}_{n+1}^{(k,j)}}{\mathbf{B}_{n+1}^{(j,j)}}\right| \,|\, \mathcal{F}_n\right) &\leq \left|\frac{\mathbf{B}_{n}^{(k,j)}}{\mathbf{B}_{n}^{(j,j)}}\right| \left(\frac{S_n^{(k)}+1}{S_n^{(k)}+2}\right)^{\frac{1}{N}}\\
&\leq \left|\frac{\mathbf{B}_{N_2}^{(k,j)}}{\mathbf{B}_{N_2}^{(j,j)}}\right| \left(\frac{S_{N_2}^{(k)}+1}{S_{N_2}^{(k)}+2}\right)^{\frac{1}{N}} \cdots \left(\frac{S_n^{(k)}+1}{S_n^{(k)}+2}\right)^{\frac{1}{N}}\\
&= \left|\frac{\mathbf{B}_{N_2}^{(k,j)}}{\mathbf{B}_{N_2}^{(j,j)}}\right| \left(\frac{S_{N_2}^{(k)}+1}{S_n^{(k)}+2}\right)^{\frac{1}{N}}\\
&\leq \left|\frac{\mathbf{B}_{N_2}^{(k,j)}}{\mathbf{B}_{N_2}^{(j,j)}}\right| \left(\frac{N_2+2}{\gamma(n+1)+1}\right)^{\frac{1}{N}}\\
&\leq \sqrt{1-\delta}\left(\frac{N_2+2}{\gamma(n+1)+1}\right)^{\frac{1}{N}},
\end{aligned}
$$

whence

$$
\sum_{n\geq N_2}\mathbb{E}(\xi_{n+1}1_{\{|\xi_{n+1}|\leq c\}} \,|\, \mathcal{F}_n) \leq \sum_{n\geq N_2}\frac{\sqrt{1-\delta}}{n+3}\left(\frac{N_2+2}{\gamma(n+1)+1}\right)^{\frac{1}{N}} < \infty,
$$

so the second series converges and (2.82) holds.

Therefore, by Theorem 2.3.3, the limit $\lim_{n\to\infty}\left(\theta_n^{(k)}\right)^2$ exists and is finite. Furthermore,

$$\sum_{n=1}^{\infty}\frac{1}{n+2}\mu_n^{(k)}R_n\left(\mu_n^{(k)}\right)=\sum_{n=1}^{\infty}\frac{1}{n+2}\left(\mu_n^{(k)}\right)^2 I_n^{(k)}=\sum_{n=1}^{\infty}\frac{1}{\zeta_n^{(k)}+2}\left(\mu_{\zeta_n^{(k)}}^{(k)}\right)^2<\infty$$

almost surely. Then, from Assumption 2.5.7, there must exist a subsequence $\left(\mu_{\zeta_{n_t}^{(k)}}^{(k)}\right)_{t=1}^{\infty}$ of $\left(\mu_{\zeta_n^{(k)}}^{(k)}\right)_{n=1}^{\infty}$ such that $\left(\mu_{\zeta_{n_t}^{(k)}}^{(k)}\right)^2\to 0$. Since $\left(\mu_{\zeta_{n_t}^{(k)}}^{(k)}\right)_{t=1}^{\infty}$ is also a subsequence of $\left(\mu_n^{(k)}\right)_{n=1}^{\infty}$, and $\lim_{n\to\infty}\left(\mu_n^{(k)}\right)^2$ exists, we conclude that

$$\lim_{n\to\infty}\left(\mu_n^{(k)}\right)^2=0,$$

which concludes the proof. □

2.5.4 Censored Binary Observations with Unknown Mean and Variance

In this section, we present an extension of the motivating example from Section 2.2 in which both the mean *and* the variance of the underlying distribution are unknown and have to be learned from censored binary signals. Because our prior is now a bivariate distribution, the learning model in Section 2.2 cannot be easily extended and the moment-matching method no longer yields a tractable algorithm. Instead, we use a variational bound technique (similar to [6] or [47]) to create a new tractable approximate Bayesian model for this setting. Section 2.5.4.1 presents this model and proves its consistency using our theoretical framework from Section 2.3. Section 2.5.4.2 explains how the model was derived.

2.5.4.1 Learning Model and Consistency Proof

Consider the normal distribution $\mathcal{N}\left(\theta, \tau^{-1}\right)$, and suppose that both the mean θ and precision τ are unknown. A standard Bayesian model for this setting is the normal-gamma prior [49]; under this model, we assume that $\tau \sim \mathit{Gamma}\left(\alpha_0, \beta_0\right)$ and that the conditional distribution of θ, given τ, is $\mathcal{N}\left(\mu_0, \left(\kappa_0 \tau\right)^{-1}\right)$. These assumptions characterize the joint prior distribution of (θ, τ) using four belief parameters $(\alpha_0, \beta_0, \kappa_0, \mu_0)$.

As in Section 2.2, we will assume that only censored samples from the normal distribution are available. However, since there are now two unknown parameters, we will need to observe two samples per time period, rather than just one; thus, suppose that $\left(Y_n^{(1)}, Y_n^{(2)}\right)_{n=1}^{\infty}$ is a sequence of i.i.d. pairs, with both components of each pair drawn independently from $\mathcal{N}\left(\theta, \tau^{-1}\right)$, and let

$$B_{n+1} = \left(B_{n+1}^{(1)}, B_{n+1}^{(2)}\right) = \left(1_{\left\{Y_{n+1}^{(1)} < b_n^{(1)}\right\}}, 1_{\left\{Y_{n+1}^{(2)} < b_n^{(2)}\right\}}\right)$$

be a pair of censored binary signals observed at time n. We now require two thresholds $b_n^{(1)}, b_n^{(2)}$ per time period, with conditions on these two sequences to be specified further down.

Essentially, the model in Section 2.2 allows us to learn the *likelihood* of the censored signals. When there is only one unknown parameter (e.g., unknown mean and known variance, as in Section 2.2, or known mean and unknown variance), this is sufficient to learn its exact value. Now that there are two parameters to be learned, we require two sequences of observations in order to learn both parameters exactly.

Since moment-matching does not yield a computationally tractable solution in this model, we propose a new approximate Bayesian updating scheme in which the conditional distribution of (θ, τ) at time n is assumed to be normal-gamma with four recursively updated parameters $(\alpha_n, \beta_n, \kappa_n, \mu_n)$. We state the updating equations here; see Section 2.5.4.2 for the details of how they were derived. First, we let

$$\alpha_{n+1} = \frac{1}{2}(n+1), \tag{2.84}$$

$$\kappa_{n+1} = n+1, \tag{2.85}$$

identically to the conjugate model in Section 9.6 of [49]. These two parameters essentially count the number of observations, and we leave their role unchanged. For the remaining two parameters, we first apply a transformation

$$\xi_n \triangleq \mu_n \sqrt{\frac{\alpha_n}{\beta_n}}, \tag{2.86}$$

$$\eta_n \triangleq \sqrt{\frac{\alpha_n}{\beta_n}}, \tag{2.87}$$

and update

$$\xi_{n+1} = \xi_n - \frac{1}{n+1} \sum_{i=1,2} \left(B^{(i)}_{n+1} \frac{\phi\left(p^{(i)}_n\right)}{\Phi\left(p^{(i)}_n\right)} - \left(1 - B^{(i)}_{n+1}\right) \frac{\phi\left(p^{(i)}_n\right)}{1 - \Phi\left(p^{(i)}_n\right)} \right), \tag{2.88}$$

$$\begin{aligned}\eta_{n+1} &= \eta_n \\ &+ \frac{1}{n+1} \sum_{i=1,2} b^{(i)}_n \left(B^{(i)}_{n+1} \frac{\phi\left(p^{(i)}_n\right)}{\Phi\left(p^{(i)}_n\right)} - \left(1 - B^{(i)}_{n+1}\right) \frac{\phi\left(p^{(i)}_n\right)}{1 - \Phi\left(p^{(i)}_n\right)} \right),\end{aligned} \tag{2.89}$$

where $p^{(i)}_n = b^{(i)}_n \eta_n - \xi_n$ for $i = 1, 2$. The resulting scheme is statistically consistent, as shown in the following result.

Proposition 2.5.1. *Suppose that $(\alpha_n, \kappa_n, \xi_n, \eta_n)$ are updated using (2.84)-(2.85)* *and (2.88)-(2.89). Suppose, furthermore, that both sequences* $\left(b_n^{(1)}\right)_{n=0}^{\infty}$ *and* $\left(b_n^{(2)}\right)_{n=0}^{\infty}$ *are bounded, and* $\inf_n \left|b_n^{(1)} - b_n^{(2)}\right| > 0$. *Then,* $(\xi_n, \eta_n) \to (\theta\sqrt{\tau}, \sqrt{\tau})$ *almost surely.*

Proof. Let

$$
\begin{aligned}
t_n &\triangleq (\xi_n, \eta_n)^T, \\
\gamma &\triangleq (\theta\sqrt{\tau}, \sqrt{\tau})^T, \\
\mathcal{F}_n &\triangleq \mathcal{B}(B_1, ..., B_n, t_0, ..., t_n, b_0, ..., b_n), \\
q_n^{(i)} &\triangleq b_n^{(i)}\sqrt{\tau} - \theta\sqrt{\tau}, \\
Q_n(B_{n+1}, t_n) &\triangleq \sum_{i=1,2} \left(B_{n+1}^{(i)} \frac{\phi\left(p_n^{(i)}\right)}{\Phi\left(p_n^{(i)}\right)} - \left(1 - B_{n+1}^{(i)}\right) \frac{\phi\left(p_n^{(i)}\right)}{1 - \Phi\left(p_n^{(i)}\right)} \right) \left(1, -b_n^{(i)}\right)^T, \\
R_n(t_n) &\triangleq \mathbb{E}(Q_n(B_{n+1}, t_n)|\mathcal{F}_n) \\
&= \sum_{i=1,2} \left(\Phi\left(q_n^{(i)}\right) \frac{\phi\left(p_n^{(i)}\right)}{\Phi\left(p_n^{(i)}\right)} - \left(1 - \Phi\left(q_n^{(i)}\right)\right) \frac{\phi\left(p_n^{(i)}\right)}{1 - \Phi\left(p_n^{(i)}\right)} \right) \left(1, -b_n^{(i)}\right)^T,
\end{aligned}
$$

then (2.88) and (2.89) are equivalent to

$$t_{n+1} = t_n - \frac{1}{n+1} Q_n(B_{n+1}, t_n).$$

First, as argued in the proof of Proposition 2.4.1, since $\left(b_n^{(1)}\right)_{n=0}^{\infty}$ and $\left(b_n^{(2)}\right)_{n=0}^{\infty}$ are bounded, there exists a positive constant C_1 such that

$$\sup_n \mathbb{E}\left(\|Q_n(B_{n+1}, x)\|_2^2 \,|\mathcal{F}_n\right) \leq C_1\left(1 + \|x - \gamma\|_2^2\right),$$

thus Assumption 2.3.3 is satisfied. Then we have

$$
\begin{aligned}
\mathbb{E}\left(\|t_{n+1} - \gamma\|_2^2 \,|\mathcal{F}_n\right) \;\leq\; & \|t_n - \gamma\|_2^2 \left(1 + \frac{C_1}{(n+1)^2}\right) + \frac{C_1}{(n+1)^2} \\
& - \frac{2}{n+1}(t_n - \gamma)^T R_n(t_n), \qquad (2.90)
\end{aligned}
$$

where

$$\begin{aligned}
& (t_n-\gamma)^T R_n(t_n) \\
= & \sum_{i=1,2}\left(q_n^{(i)}-p_n^{(i)}\right)\left(\Phi\left(q_n^{(i)}\right)\frac{\phi\left(p_n^{(i)}\right)}{\Phi\left(p_n^{(i)}\right)}-\left(1-\Phi\left(q_n^{(i)}\right)\right)\frac{\phi\left(p_n^{(i)}\right)}{1-\Phi\left(p_n^{(i)}\right)}\right) \\
\geq & \ 0, \qquad (2.91)
\end{aligned}$$

because, for both $i=1$ and $i=2$,

$$\left(q_n^{(i)}-p_n^{(i)}\right)\left(\Phi\left(q_n^{(i)}\right)\frac{\phi\left(p_n^{(i)}\right)}{\Phi\left(p_n^{(i)}\right)}-\left(1-\Phi\left(q_n^{(i)}\right)\right)\frac{\phi\left(p_n^{(i)}\right)}{1-\Phi\left(p_n^{(i)}\right)}\right)\geq 0. \qquad (2.92)$$

Then, from the proof of Theorem 2.3.1, (2.90) together with (2.91) implies that $\lim_{n\to\infty}\|t_n-\gamma\|_2^2$ exists and is finite, and that

$$\sum_{n=1}^{\infty}\frac{2}{n+1}(t_n-\gamma)^T R_n(t_n)<\infty$$

almost surely. From (2.92), since $\left(b_n^{(1)}\right)_{n=0}^{\infty}$ and $\left(b_n^{(2)}\right)_{n=0}^{\infty}$ are bounded, there must be a subsequence $\left(q_{n_k}^{(1)}-p_{n_k}^{(1)}, q_{n_k}^{(2)}-p_{n_k}^{(2)}\right)_{k=0}^{\infty}$ of $\left(q_n^{(1)}-p_n^{(1)}, q_n^{(2)}-p_n^{(2)}\right)_{n=0}^{\infty}$ that converges to 0. The subsequence can be written as

$$\begin{pmatrix} q_{n_k}^{(1)}-p_{n_k}^{(1)} \\ q_{n_k}^{(2)}-p_{n_k}^{(2)} \end{pmatrix}=\begin{pmatrix} 1 & -b_{n_k}^{(1)} \\ 1 & -b_{n_k}^{(2)} \end{pmatrix}(t_{n_k}-\gamma),$$

since $\inf_n\left|b_n^{(1)}-b_n^{(2)}\right|$ is positive, we have

$$t_{n_k}-\gamma=\frac{1}{b_{n_k}^{(1)}-b_{n_k}^{(2)}}\begin{pmatrix} -b_{n_k}^{(2)} & b_{n_k}^{(1)} \\ -1 & 1 \end{pmatrix}\begin{pmatrix} q_{n_k}^{(1)}-p_{n_k}^{(1)} \\ q_{n_k}^{(2)}-p_{n_k}^{(2)} \end{pmatrix},$$

then since $\left(b_n^{(1)}\right)_{n=0}^{\infty}$ and $\left(b_n^{(2)}\right)_{n=0}^{\infty}$ are bounded, and $\inf_n\left|b_n^{(1)}-b_n^{(2)}\right|$ is positive, the subsequence $(t_{n_k}-\gamma)_{k=0}^{\infty}$ also converges to 0, and we know that $\lim_{n\to\infty}\|t_n-\gamma\|_2^2$ exists and is finite, thus we have $\lim_{n\to\infty}\|t_n-\gamma\|_2^2=0$. Therefore, $t_n\to\gamma$ a.s., as required. □

2.5.4.2 Derivation of the Learning Model

Suppose that, at time n, (θ, τ) follows a normal-gamma density, denoted by

$$f_n\left(\theta, \tau \mid \alpha_n, \beta_n, \mu_n, \kappa_n\right) = \frac{\beta_n^{\alpha_n}\sqrt{\kappa_n}}{\Gamma(\alpha_n)\sqrt{2\pi}}\tau^{\alpha_n - 1/2}\exp\left(-\beta_n\tau - \frac{\kappa_n\tau(\theta-\mu_n)^2}{2}\right).$$

Then, the posterior density given B_{n+1} can be written as

$$\begin{aligned} & g_n\left(\theta, \tau | \alpha_n, \beta_n, \mu_n, \kappa_n, b_n, B_{n+1}\right) \\ = \quad & \frac{f_n\left(\theta, \tau | \alpha_n, \beta_n, \mu_n, \kappa_n\right) w\left(\theta, \tau, b_n, B_{n+1}\right)}{\displaystyle\int\int f_n\left(\theta, \tau | \alpha_n, \beta_n, \mu_n, \kappa_n\right) w\left(\theta, \tau, b_n, B_{n+1}\right) d\theta d\tau}, \end{aligned}$$

where

$$w\left(\theta, \tau, b_n, B_{n+1}\right) = \prod_{i=1,2}\left(\Phi\left(\sqrt{\tau}(b_n^{(i)} - \theta)\right)\right)^{B_{n+1}^{(i)}}\left(1 - \Phi\left(\sqrt{\tau}(b_n^{(i)} - \theta)\right)\right)^{1-B_{n+1}^{(i)}}.$$

Obviously it is difficult to characterize this posterior density g_n directly. Therefore, we would like to approximate the posterior density g_n by a normal-gamma density $f_{n+1}\left(\theta, \tau | \alpha_{n+1}, \beta_{n+1}, \mu_{n+1}, \kappa_{n+1}\right)$, through minimizing the Kullback-Leibler divergence $D_n = D(f_{n+1}||g_n) = \mathbb{E}_{f_{n+1}}\left(\log\frac{f_{n+1}}{g_n}\right)$, where $\mathbb{E}_{f_{n+1}}\left(\cdot\right)$ denotes the expectation taken with respect to the density f_{n+1}.

We work through the derivation for the case where $B_{n+1} = (1, 1)$; the other three cases can be obtained similarly. In this case, we write

$$\begin{aligned} \log\frac{f_{n+1}}{g_n} &= \log f_{n+1} - \log g_n \\ &= -\frac{1}{2}\left(\left(\theta - \mu_{n+1}\right)^2\kappa_{n+1} - \left(\theta - \mu_n\right)^2\kappa_n\right)\tau \\ &\quad + \left(\alpha_{n+1} - \alpha_n\right)\log\tau - (\beta_{n+1} - \beta_n)\tau \\ &\quad + (\alpha_{n+1}\log\beta_{n+1} - \alpha_n\log\beta_n) + \frac{1}{2}(\log\kappa_{n+1} - \log\kappa_n) \end{aligned}$$

$$+ \left(\log\Gamma(\alpha_{n+1}) - \log\Gamma(\alpha_n)\right)$$

$$- \sum_{i=1,2} \log\left(\Phi\left(\sqrt{\tau}(b_n^{(i)} - \theta)\right)\right) + C_1, \qquad (2.93)$$

where C_1 is a constant that does not depend on $(\alpha_{n+1}, \beta_{n+1}, \mu_{n+1}, \kappa_{n+1})$. The expectation $\mathbb{E}_{f_{n+1}}\left(\log\left(\Phi\left(\sqrt{\tau}(b_n^{(i)} - \theta)\right)\right)\right)$ is still difficult to evaluate, so we approximate $\log\left(\Phi\left(b_n^{(i)}\sqrt{\tau} - \theta\sqrt{\tau}\right)\right)$ by its first-order Taylor expansion with respect to $(\theta\sqrt{\tau}, \sqrt{\tau})$ around $\left(\mu_n\sqrt{r_n}, \sqrt{r_n}\right)$, where $r_n = \alpha_n/\beta_n$. This is analogous to the technique used in [6], where a Taylor expansion is also used to "linearize" a difficult posterior. We will use additional simplifications of the various expressions in order to obtain a tractable scheme.

By using the first-order Taylor expansion, we have

$$\begin{aligned}
\log\left(\Phi\left(\sqrt{\tau}(b_n^{(i)} - \theta)\right)\right) \;&\approx\; \log\left(\Phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)\right) \\
&\quad - \frac{\phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)}{\Phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)}(\theta\sqrt{\tau} - \mu_n\sqrt{r_n}) \\
&\quad + \frac{\phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)}{\Phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)} b_n^{(i)}(\sqrt{\tau} - \sqrt{r_n}).
\end{aligned}$$

By replacing $\log\left(\Phi\left(\sqrt{\tau}(b_n^{(i)} - \theta)\right)\right)$ in (2.93) by the above expression, we obtain an approximation $\tilde{D}_n$ of the KL divergence, given by

$$\begin{aligned}
\tilde{D}_n \;=\; & \mathbb{E}_{f_{n+1}}\Bigg(-\frac{1}{2}\left((\theta - \mu_{n+1})^2\kappa_{n+1} - (\theta - \mu_n)^2\kappa_n\right)\tau \\
& + (\alpha_{n+1} - \alpha_n)\log\tau - (\beta_{n+1} - \beta_n)\tau \\
& + \alpha_{n+1}\log\beta_{n+1} + \frac{1}{2}\log\kappa_{n+1} + \log\Gamma(\alpha_{n+1}) \\
& + \sum_{i=1,2}\left(\frac{\phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)}{\Phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)}\theta\sqrt{\tau} - \frac{\phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)}{\Phi\left(\sqrt{r_n}(b_n^{(i)} - \mu_n)\right)} b_n^{(i)}\sqrt{\tau}\right)\Bigg) + C_2
\end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{1}{2}\left(\frac{\kappa_n}{\kappa_{n+1}} + (\mu_{n+1}-\mu_n)^2\kappa_n\frac{\alpha_{n+1}}{\beta_{n+1}}\right) + (\alpha_{n+1}-\alpha_n)(\psi(\alpha_{n+1}) - \log\beta_{n+1}) \\
& -(\beta_{n+1}-\beta_n)\frac{\alpha_{n+1}}{\beta_{n+1}} + \alpha_{n+1}\log\beta_{n+1} + \frac{1}{2}\log\kappa_{n+1} + \log\Gamma(\alpha_{n+1}) \\
& + \sum_{i=1,2}\left(\frac{\phi\left(\sqrt{r_n}(b_n^{(i)}-\mu_n)\right)}{\Phi\left(\sqrt{r_n}(b_n^{(i)}-\mu_n)\right)}\mu_{n+1}\sqrt{r_{n+1}} - \frac{\phi\left(\sqrt{r_n}(b_n^{(i)}-\mu_n)\right)}{\Phi\left(\sqrt{r_n}(b_n^{(i)}-\mu_n)\right)}b_n^{(i)}\sqrt{r_{n+1}}\right) \\
& + C_3, \qquad (2.94)
\end{aligned}
$$

where ψ is the digamma function, and C_2, C_3 are two constants that do not depend on $(\alpha_{n+1}, \beta_{n+1}, \mu_{n+1}, \kappa_{n+1})$. The expectation $\mathbb{E}_{f_{n+1}}(\sqrt{\tau})$ is also replaced by its point estimate $\sqrt{r_{n+1}}$.

By applying the transformation (2.86)-(2.87), we can simplify (2.94) as

$$
\begin{aligned}
\tilde{D}_n \quad = \quad & \frac{1}{2}\kappa_n\left(\xi_{n+1} - \xi_n\frac{\eta_{n+1}}{\eta_n}\right)^2 - 2\alpha_n\log\eta_{n+1} + \alpha_n\frac{\eta_{n+1}^2}{\eta_n^2} \\
& + \sum_{i=1,2}\left(\frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)}\xi_{n+1} - \frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)}b_n^{(i)}\eta_{n+1}\right) + C_4, \quad (2.95)
\end{aligned}
$$

where C_4 is a constant that does not depend on (ξ_{n+1}, η_{n+1}). We further approximate (2.95) as

$$
\begin{aligned}
\hat{D}_n \quad \approx \quad & \frac{1}{2}\kappa_n\left(\xi_{n+1} - \xi_n\right)^2 - 2\alpha_n\log\eta_{n+1} + \alpha_n\frac{\eta_{n+1}^2}{\eta_n^2} \\
& + \sum_{i=1,2}\left(\frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)}\xi_{n+1} - \frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)}b_n^{(i)}\eta_{n+1}\right) + C_4. \quad (2.96)
\end{aligned}
$$

Now, instead of updating (β_n, μ_n), we will update (ξ_n, η_n) through taking ξ_{n+1} and η_{n+1} such that the partial derivatives of $\hat{D}_n$ with respect to ξ_{n+1} and η_{n+1} are both equal to zero. From (2.96),

$$
\frac{\partial\hat{D}_n}{\partial\xi_{n+1}} \quad = \quad \kappa_n\left(\xi_{n+1} - \xi_n\right) + \sum_{i=1,2}\frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)},
$$

$$\frac{\partial \hat{D}_n}{\partial \eta_{n+1}} = -2\alpha_n/\eta_{n+1} + 2\alpha_n \frac{\eta_{n+1}}{\eta_n^2} - \sum_{i=1,2} \frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)} b_n^{(i)},$$

thus we have

$$\begin{aligned} \xi_{n+1} &= \xi_n - \frac{1}{\kappa_n} \sum_{i=1,2} \frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)}, \\ \eta_{n+1}^2 &= \eta_n^2 + \frac{\eta_n^2 \eta_{n+1}}{2\alpha_n} \sum_{i=1,2} \frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)} b_n^{(i)}. \end{aligned} \tag{2.97}$$

However, we can see that (2.97) is not linear, so we will instead use the update

$$\eta_{n+1} = \eta_n + \frac{1}{2\alpha_n} \sum_{i=1,2} \frac{\phi\left(b_n^{(i)}\eta_n - \xi_n\right)}{\Phi\left(b_n^{(i)}\eta_n - \xi_n\right)} b_n^{(i)}.$$

From (2.84) and (2.85), we know that $\kappa_n = 2\alpha_n = n+1$. Repeating the above analysis symmetrically for $B_{n+1} = (1,0), (0,1)$ and $(0,0)$, we obtain the updates in (2.88)-(2.89).

2.6 Conclusion

We have presented the first theoretical framework for proving the consistency of estimators constructed using approximate Bayesian inference. Our approach interprets many of these estimators as stochastic approximation procedures with the addition of an extra "bias" term. We have proposed a convergent SA algorithm of this form and demonstrated its versatility in creating entirely new consistency proofs for a suite of previously-studied approximate Bayesian schemes that have proven themselves in practical applications, but were previously unamenable to theoretical analysis. Notably, this includes three multivariate procedures with broad methodological applications in analytics, simulation and stochastic optimization. We believe that our work offers new theoretical support for the use of approximate Bayesian inference in complex learning problems, and that it provides researchers with a set of tools for developing consistency proofs in other application areas.

Chapter 3: Complete Expected Improvement Converges to an Optimal Budget Allocation

3.1 Introduction

In the ranking and selection (R&S) problem, there are M "alternatives" (or "systems"), and each alternative $j \in \{1, ..., M\}$ has an unknown value $\mu^{(j)} \in \mathbb{R}$ (for simplicity, suppose that $\mu^{(i)} \neq \mu^{(j)}$ for $i \neq j$). We wish to identify the unique best alternative $j^* = \arg\max_j \mu^{(j)}$. For any j, we have the ability to collect noisy samples of the form $W^{(j)} \sim \mathcal{N}\left(\mu^{(j)}, \left(\lambda^{(j)}\right)^2\right)$, but we are limited to a total of N samples that have to be allocated among the alternatives, under independence assumptions ensuring that samples of j do not provide any information about $i \neq j$. After the sampling budget has been consumed, we select the alternative with the highest sample mean. We say that "correct selection" occurs if the selected alternative is identical to j^*. We seek to allocate the budget in a way that maximizes the probability of correct selection.

R&S has a long history dating back to [80], and continues to be an active area of research; see the tutorials by [24] and [25]. Most modern research on this problem considers *sequential* allocation strategies, in which the decision-maker may spend part of the sampling budget, observe the results, and adjust the allocation of the remaining samples accordingly. The literature has developed various algorithmic approaches, including indifference-zone methods [81], optimal computing budget allocation (or OCBA; see [82]), and expected improvement [10]. The related literature on multi-armed bandits [83] has contributed other approaches such as Thompson sampling [34], although the bandit problem uses a different objective function from R&S and thus a good method for one problem may work poorly in the other [15]. Reference [9] gave a rigorous foundation for the notion of *optimal budget allocation* with regard to probability of correct selection. Denote by $0 \leq N^{(j)} \leq N$ the number of samples assigned to alternative j (thus, $\sum_j N^{(j)} = N$), and take $N \to \infty$ while keeping the proportion $\alpha^{(j)} = N^{(j)}/N$ constant. The optimal proportions $\alpha_*^{(j)}$ (among all possible vectors $\alpha \in \mathbb{R}_{++}^M$ satisfying $\sum_j \alpha^{(j)} = 1$) satisfy two conditions:

- Proportion assigned to alternative j^*:

$$\left(\frac{\alpha_*^{(j^*)}}{\lambda^{(j^*)}}\right)^2 = \sum_{j \neq j^*} \left(\frac{\alpha_*^{(j)}}{\lambda^{(j)}}\right)^2 \tag{3.1}$$

- Proportions assigned to arbitrary $i, j \neq j^*$:

$$\frac{\left(\mu^{(i)} - \mu^{(j^*)}\right)^2}{\left(\lambda^{(i)}\right)^2/\alpha_*^{(i)} + \left(\lambda^{(j^*)}\right)^2/\alpha_*^{(j^*)}} = \frac{\left(\mu^{(j)} - \mu^{(j^*)}\right)^2}{\left(\lambda^{(j)}\right)^2/\alpha_*^{(j)} + \left(\lambda^{(j^*)}\right)^2/\alpha_*^{(j^*)}} \tag{3.2}$$

Under this allocation, the probability of incorrect selection will converge to zero at the fastest possible rate (exponential with the best possible exponent). Of course, (3.1)-(3.2) themselves depend on the unknown performance values. A common workaround is to replace these values with plug-in estimators and repeatedly solve for the optimal proportions in a sequential manner. Even then, the optimality conditions are cumbersome to solve, which may explain why researchers and practitioners prefer suboptimal heuristics that are easier to implement. To give a recent example, [84] uses large deviations theory to derive optimality conditions, analogous to (3.1)-(3.2), for a general class of simulation-based optimization problems, but advocates approximating the conditions to obtain a more tractable solution.

In this paper, we focus on one particular class of heuristics, namely expected improvement (EI) methods, which have consistently demonstrated computational and practical advantages in a wide variety of problem classes [85–87] ever since their introduction in [10]. EI is a Bayesian approach to R&S that allocates samples in a purely sequential manner: each successive sample is used to update the posterior distributions of the values $\mu^{(j)}$, and the next sample is adaptively assigned using the so-called "value of information" criterion. This notion will be formalized in Section 3.2; here, we simply note that there are many competing definitions, such as the classic EI criterion of [10], the knowledge gradient criterion [11], or the LL_1 criterion of [12]. Reference [13] showed that the seemingly minor differences between these variants produce very different asymptotic allocations, but also that all of these allocations are suboptimal.

Recently, however, [16] proposed a new criterion called "complete expected improvement" or CEI. The formal definition of CEI is given in Section 3.3, but the main idea is that, when we evaluate the potential of a seemingly-suboptimal alternative to improve over the current-best value, we treat both of the values in this comparison as random variables (unlike classic EI, which only uses a plug-in estimate of the best value). This idea was created and implemented in [16] in the context of Gaussian Markov random fields, a more sophisticated Bayesian learning model than the version of R&S with independent normal samples that we consider here. Although the Gaussian Markov model is far more scalable and practical, it also presents greater difficulties for theoretical analysis: for example, no analog of (3.1)-(3.2) is available for statistical models with Gaussian Markov structure. In the present paper, we translate the CEI criterion to our simpler model, which enables us to study its theoretical convergence rate, and ultimately leads to strong new theoretical arguments in support of the CEI method.

Our main contribution in this paper is to prove that, with a slight modification to the method as laid out in [16], this modified version of CEI achieves both (3.1) and (3.2) asymptotically as $N \to \infty$. Not only is this a new result for EI-type methods, it is also one of the strongest guarantees for any R&S heuristic to date. To compare it with the state of the art, [15] presents a class of heuristics, called "top-two methods," which can also achieve optimal allocations, but only when a tuning parameter is set optimally. A more recent work by [88], which appeared while the present paper was under review, extended the top-two approach to use CEI calculations, but kept the requirement of a tunable parameter. By contrast, our approach requires no tuning whatsoever. A different work by [14] finds a way to reverse-engineer the EI calculations to optimize the rate, but this approach requires one to first solve (3.1)-(3.2) with plug-in estimators, and the procedure does not have a natural interpretation as an EI criterion. By contrast, CEI requires no additional computational effort compared to classic EI, and has a very simple and intuitive interpretation. In this way, our paper bridges the gap between theoretical notions of rate-optimality and the more practical concerns that motivate EI methods.

3.2 Preliminaries

We first provide some formal background for the optimality conditions (3.1)-(3.2) derived in [9], and then give an overview of EI-type methods. It is important to note that the theoretical framework of [9], as well as the theoretical analysis developed in the present paper, relies on a frequentist interpretation of R&S, in which the value of alternative i is treated as a fixed (though unknown) constant. On the other hand, EI methods are derived using Bayesian arguments; however, once the derivation is complete, one is free to apply and study the resulting algorithm in a frequentist setting (as we do in this paper). To avoid confusion, we first describe the frequentist model, then introduce details of the Bayesian model where necessary.

In the frequentist model, the values $\mu^{(i)}$ are fixed for $i = 1, ..., M$. Let $\{j_n\}_{n=0}^{\infty}$ be a sequence of alternatives chosen for sampling. For each j_n, we observe $W_{n+1}^{(j_n)} \sim \mathcal{N}\left(\mu^{(j_n)}, \left(\lambda^{(j_n)}\right)^2\right)$ where $\lambda^{(j)} > 0$ is assumed to be known for all j. We let $\mathcal{F}_n$ be the sigma-algebra generated by $j_0, W_1^{(j_0)}, ..., j_{n-1}, W_n^{(j_{n-1})}$. The allocation $\{j_n\}_{n=0}^{\infty}$ is said to be *adaptive* if each j_n is $\mathcal{F}_n$-measurable, and *static* if all j_n are $\mathcal{F}_0$-measurable. We define $I_n^{(j)} = 1_{\{j_n = j\}}$ and let $N_n^{(j)} = \sum_{m=0}^{n-1} I_m^{(j)}$ be the number of times that alternative j is sampled up to time index $n = 1, 2, ...$

At time n, we can calculate the statistics

$$\theta_n^{(j)} = \frac{1}{N_n^{(j)}} \sum_{m=0}^{n-1} I_m^{(j)} W_{m+1}^{(j)}, \tag{3.3}$$

$$\left(\sigma_n^{(j)}\right)^2 = \frac{\left(\lambda^{(j)}\right)^2}{N_n^{(j)}}. \tag{3.4}$$

If our sampling budget is limited to n samples, then $j^*_n = \arg\max_j \theta^{(j)}_n$ will be the final selected alternative. Correct selection occurs at time index n if $j^*_n = j^*$. The probability of correct selection (PCS), written as $P\left(j^*_n = j^*\right)$, depends on the rule used to allocate the samples. Reference [9] proves that, for any static allocation that assigns a proportion $\alpha^{(j)} > 0$ of the budget to each alternative j, the convergence rate of PCS can be expressed in terms of the limit

$$\Gamma^{\alpha} = -\lim_{n\to\infty} \frac{1}{n} \log P\left(j^*_n \neq j^*\right). \tag{3.5}$$

That is, the probability of *incorrect* selection converges to zero at an exponential rate where the exponent includes a constant Γ^{α} that depends on the vector α of proportions. Equations (3.1)-(3.2) characterize the proportions that optimize the rate (maximize Γ^{α}) under the assumption of independent normal samples. Although [9] only considers static allocations, nonetheless, to date, (3.5) continues to be one of the strongest rate results for R&S. Optimal static allocations derived through this framework can be used as guidance for the design of dynamic allocations; see, for example, [84] and [89].

We now describe EI, a prominent class of adaptive methods. EI uses a Bayesian model of the learning process, which is very similar to the model presented above, but makes the additional assumption that $\mu^{(j)} \sim \mathcal{N}\left(\theta^{(j)}_0, \left(\sigma^{(j)}_0\right)^2\right)$, where $\theta^{(j)}_0$ and $\sigma^{(j)}_0$ are pre-specified prior parameters. It is also assumed that $\mu^{(i)}, \mu^{(j)}$ are independent for all $i \neq j$. Under these assumptions, it is well-known [49] that the posterior distribution of $\mu^{(j)}$ given $\mathcal{F}_n$ is $\mathcal{N}\left(\theta^{(j)}_n, \left(\sigma^{(j)}_n\right)^2\right)$ where the posterior mean and variance can be computed recursively. Under the non-informative prior $\sigma^{(j)}_0 = \infty$, the Bayesian posterior parameters $\theta_n^{(j)}, \sigma_n^{(j)}$ are identical to the frequentist statistics defined in (3.3)-(3.4), and so we can use the same notation for both settings.

One of the first (and probably the best-known) EI algorithms was introduced by [10]. In this version of EI, as applied to our R&S model, we take $j_n = \arg\max_j v_n^{(j)}$ where

$$\begin{aligned} v_n^{(j)} &= \mathbb{E}\left(\max\left\{\mu^{(j)} - \theta_n^{(j_n^*)}, 0\right\} \mid \mathcal{F}_n\right) \\ &= \sigma_n^{(j)} f\left(-\frac{\left|\theta_n^{(j)} - \theta_n^{(j_n^*)}\right|}{\sigma_n^{(j)}}\right), \end{aligned} \tag{3.6}$$

and $f(z) = z\Phi(z) + \phi(z)$ with ϕ, Φ being the standard Gaussian pdf and cdf, respectively. We can view (3.6) as a measure of the potential that the true value of j will improve upon the current-best estimate $\theta_n^{(j_n^*)}$. The EI criterion $v_n^{(j)}$ may be recomputed at each time stage n based on the most recent posterior parameters.

Reference [13] gave the first convergence rate analysis of this algorithm. Under EI, we have

$$\lim_{n\to\infty} \frac{N_n^{(j^*)}}{n} = 1, \tag{3.7}$$

$$\lim_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(j)}} = \left(\frac{\lambda^{(i)}\left|\mu^{(j)} - \mu^{(j^*)}\right|}{\lambda^{(j)}\left|\mu^{(i)} - \mu^{(j^*)}\right|}\right)^2, \qquad i, j \neq j^*, \tag{3.8}$$

where the limits hold almost surely. Clearly, (3.7)-(3.8) do not match (3.1)-(3.2) except in the limiting case where $\alpha_*^{(j^*)} \to 1$. Because $N^{(j)}/n \to 0$ for $j \neq j^*$, EI will not achieve an exponential convergence rate for any finite M. The limiting allocations for two other variants of EI are also derived in [13], but they do not recover (3.1)-(3.2) either.

3.3 Algorithm and Main Results

Reference [16] proposed to replace (3.6) with

$$v_n^{(j)} = \mathbb{E}\left(\max\left\{\mu^{(j)} - \mu^{(j_n^*)}, 0\right\} \mid \mathcal{F}_n\right), \tag{3.9}$$

which can be written in closed form as

$$v_n^{(j)} = \sqrt{\left(\sigma_n^{(j)}\right)^2 + \left(\sigma_n^{(j_n^*)}\right)^2} f\left(-\frac{\left|\theta_n^{(j)} - \theta_n^{(j_n^*)}\right|}{\sqrt{\left(\sigma_n^{(j)}\right)^2 + \left(\sigma_n^{(j_n^*)}\right)^2}}\right) \tag{3.10}$$

for any $j \neq j_n^*$. In this way, the value of collecting information about j depends, not only on our uncertainty about j, but also on our uncertainty about j_n^*. [16] considers a more general Gaussian Markov model with correlated beliefs, so the original presentation of CEI included a term representing the posterior covariance between $\mu^{(j)}$ and $\mu^{(j_n^*)}$. In this paper we only consider independent priors, so we work with (3.10), which translates the CEI concept to our R&S model.

From (3.9), it follows that $v_n^{(j_n^*)} = 0$ for all n. Thus, we cannot simply assign $j_n = \arg\max_j v_n^{(j)}$ because, in that case, j_n^* would never be chosen. It is necessary to modify the procedure by introducing some additional logic to handle samples assigned to j_n^*. To the best of our knowledge, this issue is not explicitly discussed in [16]. In fact, many adaptive methods are unable to efficiently identify when j_n^* should be measured; thus, both the classic EI method of [10], and the popular Thompson sampling algorithm [34], will sample j_n^* too often. The class of top-two methods, first introduced in [15], addresses this problem by essentially assigning a fixed proportion β of samples to j_n^*, while using Thompson sampling or other means to choose between the other alternatives. Optimal allocations can be attained if β is tuned correctly, but the optimal choice of β is problem-dependent and generally difficult to find.

```
Let $n = 0$ and repeat the following:

1: Check whether
       $\left(\frac{N_n^{(j_n^*)}}{\lambda^{(j_n^*)}}\right)^2 < \sum_{j \neq j_n^*} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2.$     (3.11)
   If (3.11) holds, assign $j_n = j_n^*$. If (3.11) does not hold, assign $j_n = \arg\max_{j \neq j_n^*} v_n^{(j)}$, where $v_n^{(j)}$ is given by (3.10).

2: Observe $W_{n+1}^{(j_n)}$, update posterior parameters, and increment $n$ by 1.
```

Figure 3.1: Modified CEI (mCEI) algorithm for R&S.

Based on these considerations, we give a modified CEI procedure in Figure 3.1. The modification adds condition (3.11), which mimics (3.1) to decide whether j_n^* should be sampled. This condition is trivial to implement, and the mCEI algorithm is completely free of tunable parameters. It is shown in [90] that mCEI samples every alternative infinitely often as $n \to \infty$.

We now state our main results on the asymptotic rate-optimality of mCEI. Essentially, these theorems state that conditions (3.1) and (3.2) will hold in the limit as $n \to \infty$. Both theorems should be interpreted in the frequentist sense, that is, $\mu^{(j)}$ is a fixed but unknown constant for each j.

Theorem 3.3.1 (Optimal alternative)**.** *Let $\alpha_n^{(j)} = N_n^{(j)}/n$. Under the mCEI algorithm,*

$$\lim_{n\to\infty} \left(\frac{\alpha_n^{(j^*)}}{\lambda^{(j^*)}}\right)^2 - \sum_{j\neq j_n^*} \left(\frac{\alpha_n^{(j)}}{\lambda^{(j)}}\right)^2 = 0$$

almost surely.

Theorem 3.3.2 (Suboptimal alternatives)**.** *For $j \neq j^*$, define*

$$\tau_n^{(j)} = \frac{\left(\mu^{(j)} - \mu^{(j^*)}\right)^2}{\left(\lambda^{(j)}\right)^2/\alpha_n^{(j)} + \left(\lambda^{(j^*)}\right)^2/\alpha_n^{(j^*)}}.$$

where $\alpha_n^{(j)} = N_n^{(j)}/n$. Under the mCEI algorithm,

$$\lim_{n\to\infty} \frac{\tau_n^{(i)}}{\tau_n^{(j)}} = 1$$

almost surely, for any $i, j \neq j^*$.

3.4 Proofs of Main Results

For notational convenience, we assume that $j^* = 1$ is the unique optimal alternative. Since, under mCEI, $N_n^{(j)} \to \infty$ for all j, on almost every sample path we will always have $j_n^* = 1$ for all large enough n. It is therefore sufficient to prove Theorems 3.3.1 and 3.3.2 for a simplified version of mCEI with (3.10) replaced by

$$v_n^{(j)} = \sqrt{\frac{\left(\lambda^{(j)}\right)^2}{N_n^{(j)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}\, f\left(-\frac{\left|\theta_n^{(j)} - \theta_n^{(1)}\right|}{\sqrt{\frac{\left(\lambda^{(j)}\right)^2}{N_n^{(j)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}}\right). \tag{3.12}$$

and (3.11) replaced by

$$\left(\frac{N_n^{(1)}}{\lambda^{(1)}}\right)^2 < \sum_{j>1} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2. \tag{3.13}$$

To simplify the presentation of the key arguments, we treat the noise parameters $\lambda^{(j)}$ as being known. If, in (3.4), we replace $\lambda^{(j)}$ by the standard sample deviation (as recommended, e.g., by both [10] and [16]), then simply plug the resulting approximation into (3.10), the limiting allocation will not be affected. Because the rate-optimality framework of [9] is frequentist and assumes that selection is based only on sample means, it does not make any distinction between known and unknown variance in terms of characterizing an optimal allocation.

3.4.1 Proof of Theorem 3.3.1

First, we define the quantity

$$\Delta_n \triangleq \left(\frac{N_n^{(1)}/\lambda^{(1)}}{n}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n}\right)^2$$

and prove the following technical lemma. We remind the reader that, in this and all subsequent proofs, we assume that sampling decisions are made by mCEI with (3.12)-(3.13) replacing (3.10)-(3.11).

Lemma 3.4.1. *If alternative 1 is sampled at time n, then $\Delta_{n+1} - \Delta_n > 0$. If any other alternative is sampled at time n, then $\Delta_{n+1} - \Delta_n < 0$.*

Proof. Suppose that alternative 1 is sampled at time n. Then,

$$\begin{aligned}&\Delta_{n+1} - \Delta_n \\ = &\left(\frac{\left(N_n^{(1)}+1\right)/\lambda^{(1)}}{n+1}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n+1}\right)^2 \\ &-\left(\left(\frac{N_n^{(1)}/\lambda^{(1)}}{n}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n}\right)^2\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\left(\lambda^{(1)}\right)^2}\left(\left(\frac{\left(N_n^{(1)}+1\right)}{n+1}\right)^2 - \left(\frac{N_n^{(1)}}{n}\right)^2\right) \\
&\quad + \left(\sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n+1}\right)^2\right) \\
&> 0.
\end{aligned}
$$

If some alternative $j' > 1$ is sampled, then $\Delta_n \geq 0$ and

$$
\begin{aligned}
\Delta_{n+1} - \Delta_n &= \left(\frac{N_n^{(1)}/\lambda^{(1)}}{n+1}\right)^2 - \sum_{j\neq j'}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n+1}\right)^2 - \left(\frac{\left(N_n^{(j')}+1\right)/\lambda^{(j')}}{n+1}\right)^2 \\
&\quad - \left(\left(\frac{N_n^{(1)}/\lambda^{(1)}}{n}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n}\right)^2\right) \\
&= \left(\frac{N_n^{(1)}/\lambda^{(1)}}{n+1}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n+1}\right)^2 - \frac{2N_n^{(j')}+1}{\left(\lambda^{(j')}(n+1)\right)^2} \\
&\quad - \left(\left(\frac{N_n^{(1)}/\lambda^{(1)}}{n}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n}\right)^2\right) \\
&= \left(\frac{n^2}{(n+1)^2} - 1\right)\Delta_n - \frac{2N_n^{(j')}+1}{\left(\lambda^{(j')}(n+1)\right)^2} \\
&< 0,
\end{aligned}
$$

which completes the proof. □

Let $\ell = \min_j \lambda^{(j)}$ and recall that $\ell > 0$ by assumption. Now, for all $\varepsilon > 0$, there exists a large enough n_1 such that $n_1 > \frac{2}{\ell^2\varepsilon} - 1$. Consider arbitrary $n \geq n_1$ and suppose that $\Delta_n < 0$. This means that alternative 1 is sampled at time n, whence $\Delta_{n+1} - \Delta_n > 0$ by Lemma 3.4.1. Furthermore,

$$
\begin{aligned}
\Delta_{n+1} &= \left(\frac{\left(N_n^{(1)}+1\right)/\lambda^{(1)}}{n+1}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n+1}\right)^2 \\
&= \Delta_n + \frac{2N_n^{(1)}+1}{\left(\lambda^{(1)}(n+1)\right)^2}
\end{aligned}
$$

$$\begin{aligned}
&< \frac{2n+2}{\left(\lambda^{(1)}(n+1)\right)^2} \\
&\leq \frac{2}{\left(\lambda^{(1)}\right)^2(n_1+1)} \\
&< \frac{\ell^2}{\left(\lambda^{(1)}\right)^2}\varepsilon \\
&\leq \varepsilon.
\end{aligned}$$

Similarly, suppose that $\Delta_n \geq 0$. This means that some $j' > 1$ is sampled, whence $\Delta_{n+1} - \Delta_n < 0$ by Lemma 3.4.1. Using similar arguments as before, we find

$$\begin{aligned}
\Delta_{n+1} &= \left(\frac{N_n^{(1)}/\lambda^{(1)}}{n+1}\right)^2 - \sum_{j=2}^{M}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{n+1}\right)^2 - \frac{2N_n^{(j')}+1}{\left(\lambda^{(j')}(n+1)\right)^2} \\
&= \Delta_n - \frac{2N_n^{(j')}+1}{\left(\lambda^{(j')}(n+1)\right)^2} \\
&\geq -\frac{2n+2}{\left(\lambda^{(j')}(n+1)\right)^2} \\
&\geq -\varepsilon.
\end{aligned}$$

Thus, if there exists some large enough n_2 satisfying $n_2 \geq n_1$ and $-\varepsilon < \Delta_{n_2} < \varepsilon$, then it follows that, for all $n \geq n_2$, we have $\Delta_n \in (-\varepsilon, \varepsilon)$, which implies $\lim_{n\to\infty} \Delta_n = 0$ and completes the proof of Theorem 3.3.1. It only remains to show the existence of such n_2.

Again, we consider two cases. First, suppose that $\Delta_{n_1} < 0$. Since mCEI samples every alternative infinitely often, we can let $n_2 = \inf\{n > n_1 : \Delta_n \geq 0\}$. Since n_2 will be the first time after n_1 that any $j' > 1$ is sampled, we have $\Delta_{n_2-1} < 0$ and $n_2 - 1 \geq n_1$. From the previous arguments, we have $0 \leq \Delta_{n_2} < \varepsilon$. Similarly, in the second case where $\Delta_{n_1} \geq 0$, we let $n_2 = \inf\{n > n_1 : \Delta_n < 0\}$, whence $\Delta_{n_2-1} \geq 0$ and $n_2 - 1 \geq n_1$. The previous arguments imply $-\varepsilon < \Delta_{n_2} < 0$. Thus, we can always find $n_2 \geq n_1$ satisfying $-\varepsilon < \Delta_{n_2} < \varepsilon$, as required.

3.4.2 Proof of Theorem 3.3.2

The proof relies on several technical lemmas. For notational convenience, we define $d_n^{(j)} \triangleq \left|\theta_n^{(j)} - \theta_n^{(1)}\right|$ and $\delta_n^{(j)} = \left(d_n^{(j)}\right)^2$ for all $j > 1$. Furthermore, for any j and any positive integer m, we define

$$k_{(n,n+m)}^{(j)} \triangleq N_{n+m}^{(j)} - N_n^{(j)}$$

to be the number of samples allocated to alternative j from stage n to stage $n+m-1$.

The first technical lemma implies that, for any two alternatives i and j, $N_n^{(i)} = \Theta\left(N_n^{(j)}\right)$[1] and $N_n^{(i)} = \Theta\left(n\right)$.

Lemma 3.4.2. *For any two alternatives i and j,* $\limsup_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(j)}} < \infty$.

Proof. We proceed by contradiction. Suppose that $i, j > 1$ satisfy $\limsup_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(j)}} = \infty$. Let $c = \lim_{n\to\infty} \frac{\delta_n^{(j)}}{\delta_n^{(i)}} + 1 = \frac{\left(\mu^{(j)}-\mu^{(1)}\right)^2}{\left(\mu^{(i)}-\mu^{(1)}\right)^2} + 1$. Then, there must exist a large enough stage m such that

$$\frac{N_m^{(i)}}{N_m^{(j)}} > \max\left\{c, 1\right\} \frac{\left(\lambda^{(i)}\right)^2 + \lambda^{(1)}\lambda^{(i)}}{\left(\lambda^{(j)}\right)^2},$$

and we will sample alternative i to make $\frac{N_{m+1}^{(i)}}{N_{m+1}^{(j)}} > \frac{N_m^{(i)}}{N_m^{(j)}}$. But, at this stage m,

$$\begin{aligned} v_m^{(i)} &= \sqrt{\frac{\left(\lambda^{(i)}\right)^2}{N_m^{(i)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_m^{(1)}}} f\left(-\frac{d_m^{(i)}}{\sqrt{\frac{\left(\lambda^{(i)}\right)^2}{N_m^{(i)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_m^{(1)}}}}\right) \\ &\leq \sqrt{\frac{\left(\lambda^{(i)}\right)^2}{N_m^{(i)}} + \frac{\lambda^{(1)}\lambda^{(i)}}{N_m^{(i)}}} f\left(-\frac{d_m^{(i)}}{\sqrt{\frac{\left(\lambda^{(i)}\right)^2}{N_m^{(i)}} + \frac{\lambda^{(1)}\lambda^{(i)}}{N_m^{(i)}}}}\right) \qquad (3.14) \end{aligned}$$

[1] For two positive sequences (a_n) and (b_n), we say $a_n = \Theta(b_n)$ if and only if $a_n = O(b_n)$ and $b_n = O(a_n)$.

$$\begin{aligned}
&= \sqrt{\frac{\left(\lambda^{(i)}\right)^2 + \lambda^{(1)}\lambda^{(i)}}{N_m^{(i)}}} f\left(-\frac{d_m^{(i)}}{\sqrt{\frac{\left(\lambda^{(i)}\right)^2+\lambda^{(1)}\lambda^{(i)}}{N_m^{(i)}}}}\right) \\
&< \sqrt{\frac{\left(\lambda^{(j)}\right)^2}{N_m^{(j)}}} f\left(-\frac{d_m^{(j)}}{\sqrt{\frac{\left(\lambda^{(j)}\right)^2}{N_m^{(j)}}}}\right) \qquad (3.15) \\
&< \sqrt{\frac{\left(\lambda^{(j)}\right)^2}{N_m^{(j)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_m^{(1)}}} f\left(-\frac{d_m^{(j)}}{\sqrt{\frac{\left(\lambda^{(j)}\right)^2}{N_m^{(j)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_m^{(1)}}}}\right) \\
&= v_m^{(j)}, \qquad (3.16)
\end{aligned}$$

where (3.14) holds because a suboptimal alternative is sampled at stage m, and (3.15) holds because $\lim_{m\to\infty} \frac{d_m^{(j)}}{d_m^{(i)}} = \frac{\left|\mu^{(j)}-\mu^{(1)}\right|}{\left|\mu^{(i)}-\mu^{(1)}\right|}$. From the definition of the mCEI algorithm, (3.16) implies that we cannot sample i at stage m. We conclude that $\limsup_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(j)}} < \infty$ for any two suboptimal alternatives i and j.

From this result, we can see that, for $i, j > 1$, we have

$$0 < \liminf_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(j)}} \leq \limsup_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(j)}} < \infty.$$

Together with Theorem 3.3.1, this implies that, for any $i > 1$, we have

$$0 < \liminf_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(1)}} \leq \limsup_{n\to\infty} \frac{N_n^{(i)}}{N_n^{(1)}} < \infty,$$

completing the proof. □

Now let

$$\begin{aligned}
z_n^{(j)} &\triangleq \frac{d_n^{(j)}}{\sqrt{\frac{\left(\lambda^{(j)}\right)^2}{N_n^{(j)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}}, \\
t_n^{(j)} &\triangleq \left(z_n^{(j)}\right)^2 = \frac{\delta_n^{(j)}}{\frac{\left(\lambda^{(j)}\right)^2}{N_n^{(j)}} + \frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}.
\end{aligned}$$

For any j, both $z_n^{(j)}$ and $t_n^{(j)}$ go to infinity as $n \to \infty$. We apply an expansion of the Mills ratio [91] to $v_n^{(j)}$. For all large enough n,

$$\begin{aligned} v_n^{(j)} &= \frac{d_n^{(j)}}{z_n^{(j)}} f\left(-z_n^{(j)}\right) \\ &= \frac{d_n^{(j)}}{z_n^{(j)}} \phi\left(z_n^{(j)}\right) \left(-z_n^{(j)} \frac{1-\Phi\left(z_n^{(j)}\right)}{\phi\left(z_n^{(j)}\right)} + 1\right) \\ &= \frac{d_n^{(j)}}{z_n^{(j)}} \phi\left(z_n^{(j)}\right) \left(-z_n^{(j)} \frac{1}{z_n^{(j)}} \left(1 - \frac{1}{\left(z_n^{(j)}\right)^2} + O\left(\frac{1}{\left(z_n^{(j)}\right)^4}\right)\right) + 1\right) \qquad (3.17) \\ &= \frac{d_n^{(j)}}{\left(z_n^{(j)}\right)^3} \phi\left(z_n^{(j)}\right) \left(1 + O\left(\frac{1}{\left(z_n^{(j)}\right)^2}\right)\right), \end{aligned}$$

where (3.17) comes from the Mills ratio. Then,

$$\begin{aligned} 2\log\left(v_n^{(j)}\right) &= 2\log d_n^{(j)} - 6\log z_n^{(j)} + 2\log\phi\left(z_n^{(j)}\right) + 2\log\left(1 + O\left(\frac{1}{\left(z_n^{(j)}\right)^2}\right)\right) \\ &= \log\delta_n^{(j)} - 3\log t_n^{(j)} - \log(2\pi) - t_n^{(j)} + 2\log\left(1 + O\left(\frac{1}{t_n^{(j)}}\right)\right) \\ &= -t_n^{(j)}\left(1 + O\left(\frac{\log t_n^{(j)}}{t_n^{(j)}}\right)\right). \end{aligned}$$

For any two suboptimal alternatives i and j, define

$$\begin{aligned} r_n^{(i,j)} &\triangleq \frac{2\log\left(v_n^{(i)}\right)}{2\log\left(v_n^{(j)}\right)} \\ &= \frac{t_n^{(i)}}{t_n^{(j)}} \frac{1 + O\left(\frac{\log t_n^{(i)}}{t_n^{(i)}}\right)}{1 + O\left(\frac{\log t_n^{(j)}}{t_n^{(j)}}\right)}, \qquad (3.18) \end{aligned}$$

and note that both $1 + O\left(\frac{\log t_n^{(i)}}{t_n^{(i)}}\right)$ and $1 + O\left(\frac{\log t_n^{(j)}}{t_n^{(j)}}\right)$ converge to 1 as $n \to \infty$. We will show that $r_n^{(i,j)} \to 1$ for any suboptimal i and j; then, (3.18) will yield $\frac{t_n^{(i)}}{t_n^{(j)}} \to 1$, completing the proof of Theorem 3.3.2.

Note that, for any j, the CEI quantity $v_n^{(j)}$ can change when either j or the optimal alternative is sampled. Thus, it is necessary to characterize the relative frequency of such samples. This requires three other technical lemmas. First, Lemma 3.4.3 shows that the number of samples that could be allocated to the optimal alternative between two samples of any suboptimal alternatives (not necessarily the same one) is $O(1)$ and vice versa; next, Lemma 3.4.4 shows that $k_{(n,n+m)}^{(1)}$ is $O\left(\sqrt{n \log\log n}\right)$; finally, Lemma 3.4.6 bounds $n^{3/4}\left|\delta_{n+1}^{(i)} - \delta_n^{(i)}\right|$.

Lemma 3.4.3. *Between two samples assigned to any suboptimal alternatives (i.e., two time stages when condition (3.13) fails), the number of samples that could be allocated to the optimal alternative is at most equal to some fixed constant B_1; symmetrically, between two samples of alternative 1, the number of samples that could be allocated to any suboptimal alternatives is at most equal to some fixed constant B_2.*

Proof. Define $Q_n \triangleq \left(N_n^{(1)}/\lambda^{(1)}\right)^2 - \sum_{j=2}^M \left(N_n^{(j)}/\lambda^{(j)}\right)^2$. Suppose that, at some stage n, $Q_n < 0$ and $Q_{n+1} \geq 0$, which means that the optimal alternative is sampled at time n and then a suboptimal alternative is sampled at time $n+1$. Let $m \triangleq \inf\{l > 0 : Q_{n+l} < 0\}$, i.e., stage $n+m$ is the first time that alternative 1 is sampled after stage n. Then, in order to show that between two samples of alternative 1, the number of samples that could be allocated to suboptimal alternatives is $O(1)$, it is sufficient to show that $m = O(1)$.

To show this, first we can see that

$$Q_{n+1} = \left(\frac{N_n^{(1)}+1}{\lambda^{(1)}}\right)^2 - \sum_{j=2}^M \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2$$

$$
\begin{aligned}
&= \left(\frac{N_n^{(1)}}{\lambda^{(1)}}\right)^2 - \sum_{j=2}^{M} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2 + \frac{2N_n^{(1)}+1}{\left(\lambda^{(1)}\right)^2} \\
&= Q_n + \frac{2N_n^{(1)}+1}{\left(\lambda^{(1)}\right)^2} \\
&< \frac{2N_n^{(1)}+1}{\left(\lambda^{(1)}\right)^2} \\
&\leq C_1 N_n^{(1)},
\end{aligned} \tag{3.19}
$$

where C_1 is a suitable fixed positive constant and the first inequality holds because $Q_n < 0$. Then, for any stage $n+s$, where $0 < s < m$, we have

$$
\begin{aligned}
Q_{n+s} &= \left(\frac{N_{n+s}^{(1)}}{\lambda^{(1)}}\right)^2 - \sum_{j=2}^{M} \left(\frac{N_{n+s}^{(j)}}{\lambda^{(j)}}\right)^2 \\
&= \left(\frac{N_{n+1}^{(1)}}{\lambda^{(1)}}\right)^2 - \sum_{j=2}^{M} \left(\frac{N_{n+s}^{(j)}}{\lambda^{(j)}}\right)^2 \\
&= \left(\frac{N_n^{(1)}+1}{\lambda^{(1)}}\right)^2 - \sum_{j=2}^{M} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2 - \left(\sum_{j=2}^{M} \left(\frac{N_{n+s}^{(j)}}{\lambda^{(j)}}\right)^2 - \sum_{j=2}^{M} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2\right) \\
&< C_1 N_n^{(1)} - \left(\sum_{j=2}^{M} \left(\frac{N_{n+s}^{(j)}}{\lambda^{(j)}}\right)^2 - \sum_{j=2}^{M} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2\right),
\end{aligned}
$$

where the inequality holds because of (3.19). We can also see that, after stage n, the increment of $\sum_{j=2}^{M} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2$ obtained by allocating a sample to alternative j is at least $\frac{2N_n^{(j)}}{\left(\lambda^{(j)}\right)^2}$. Then, for all large enough n,

$$
\sum_{j=2}^{M} \left(\frac{N_{n+s}^{(j)}}{\lambda^{(j)}}\right)^2 - \sum_{j=2}^{M} \left(\frac{N_n^{(j)}}{\lambda^{(j)}}\right)^2 \geq 2s \frac{\min_{\{j>1\}} N_n^{(j)}}{\max_{\{j>1\}} \left(\lambda^{(j)}\right)^2} \geq C_2 s N_n^{(1)},
$$

where C_2 is a suitable positive constant and the last inequality follows by Lemma 3.4.2. Therefore, for any $0 < s < m$, we have $Q_{n+s} < \left(C_1 - C_2 s\right) N_n^{(1)}$. But, from the definition of m, for any $0 < s < m$, $Q_{n+s} \geq 0$ must hold. Thus, any $0 < s < m$ cannot be greater than C_1/C_2; in other words, we must have $m \leq C_1/C_2 + 1$, which implies $m = O(1)$ for all large enough n. This proves the second claim of the lemma. The first claim of the lemma can be proved in a similar way due to symmetry. □

Lemma 3.4.4. *If some suboptimal alternative $i > 1$ is sampled at stage $n \geq 3$, then $k^{(1)}_{(n,n+m)} = O\left(\sqrt{n \log\log n}\right)$ for*

$$m \triangleq \inf\left\{l > 0 : I^{(i)}_{n+l} = 1\right\}.$$

Proof. We first introduce a technical lemma, which establishes a relationship between $k^{(1)}_{(n,n+m)}$ and samples assigned to suboptimal alternatives. The lemma is proved right after the current proof.

Lemma 3.4.5. *Let C_1 be any positive constant, and take a large enough n such that some suboptimal alternative $i > 1$ is sampled at stage n. Define*

$$m \triangleq \inf\left\{l > 0 : I^{(i)}_{n+l} = 1\right\}, \qquad s \triangleq \sup\left\{l < m : I^{(1)}_{n+l} = 0\right\}.$$

Suppose that there exists a sufficiently large positive constant C_2 (dependent on C_1, but independent of n) for which

$$C_2\sqrt{n \log\log n} \leq k^{(1)}_{(n,n+s)} \leq n$$

holds. Then, there exists a suboptimal alternative $j \neq i$ and a time stage $n+u$, where $u \leq s$, such that j is sampled at stage $n+u$ and

$$\left(1 + C_1\frac{\sqrt{n\log\log n}}{n}\right)\frac{N^{(j)}_n}{N^{(1)}_n} < \frac{N^{(j)}_n + k^{(j)}_{(n,n+u)}}{N^{(1)}_n + k^{(1)}_{(n,n+s)}} \leq \frac{N^{(j)}_n + k^{(j)}_{(n,n+u)}}{N^{(1)}_n + k^{(1)}_{(n,n+u)}}, \tag{3.20}$$

holds.

Essentially, Lemma 3.4.5 will be used to prove the desired result by contradiction; we will show that (3.20) cannot arise, and therefore $k^{(j)}_{(n,n+m)}$ must be $O\left(\sqrt{n\log\log n}\right)$.

For convenience, we abbreviate $k^{(j)}_{(n,n+m)}$ by the notation $k^{(j)}_{l}$. We will prove the lemma by contradiction. Suppose that the conclusion of the lemma does not hold, that is, $\frac{k^{(1)}_m}{\sqrt{n\log\log n}}$ can be arbitrarily large. Since we sample $i > 1$ at stage n, then for any other suboptimal alternative $j \neq i$, we have

$$r^{(i,j)}_n = \frac{t^{(i)}_n}{t^{(j)}_n}\frac{1+O\left(\frac{\log t^{(i)}_n}{t^{(i)}_n}\right)}{1+O\left(\frac{\log t^{(j)}_n}{t^{(j)}_n}\right)} \leq 1.$$

Then, by Lemma 3.4.2, there must exist positive constants C_1 and C_2 such that, for all large enough n,

$$\frac{t^{(i)}_n}{t^{(j)}_n} \leq 1 + C_1\left(\frac{\log t^{(j)}_n}{t^{(j)}_n} + \frac{\log t^{(i)}_n}{t^{(i)}_n}\right) \leq 1 + C_2\frac{\log n}{n},$$

that is, equivalently,

$$\frac{\delta^{(i)}_n\left(\lambda^{(j)}\right)^2}{N^{(j)}_n} + \frac{\delta^{(i)}_n\left(\lambda^{(1)}\right)^2}{N^{(1)}_n} \leq \frac{\delta^{(j)}_n\left(\lambda^{(i)}\right)^2\left(1+C_2\frac{\log n}{n}\right)}{N^{(i)}_n} + \frac{\delta^{(j)}_n\left(\lambda^{(1)}\right)^2\left(1+C_2\frac{\log n}{n}\right)}{N^{(1)}_n} \quad (3.21)$$

Then, at stage $n+u$, where $0 < u < m$, there must exist positive constants C_3 and C_4 such that, for all large enough n,

$$r^{(i,j)}_{n+u} = \frac{t^{(i)}_{n+u}}{t^{(j)}_{n+u}}\frac{1+O\left(\frac{\log t^{(i)}_{n+u}}{t^{(i)}_{n+u}}\right)}{1+O\left(\frac{\log t^{(j)}_{n+u}}{t^{(j)}_{n+u}}\right)} \leq \frac{t^{(i)}_{n+u}}{t^{(j)}_{n+u}}\frac{1}{1-C_3\left(\frac{\log t^{(i)}_{n+u}}{t^{(i)}_{n+u}}+\frac{\log t^{(j)}_{n+u}}{t^{(j)}_{n+u}}\right)} < \frac{t^{(i)}_{n+u}}{t^{(j)}_{n+u}}\frac{1}{1-C_4\frac{\log n}{n}}.$$

Thus, for all large enough n, in order to have $r^{(i,j)}_{n+u} < 1$, it is sufficient to require

$$\frac{t^{(i)}_{n+u}}{t^{(j)}_{n+u}} \leq 1 - C_4\frac{\log n}{n},$$

or, equivalently,

$$\begin{aligned}&\frac{\delta_{n+u}^{(i)}\left(\lambda^{(j)}\right)^2}{N_n^{(j)}+k_u^{(j)}}+\frac{\delta_{n+u}^{(i)}\left(\lambda^{(1)}\right)^2}{N_n^{(1)}+k_u^{(1)}}\\ &\quad\le\ \frac{\delta_{n+u}^{(j)}\left(\lambda^{(i)}\right)^2\left(1-C_4\frac{\log n}{n}\right)}{N_n^{(i)}+k_u^{(i)}}+\frac{\delta_{n+u}^{(j)}\left(\lambda^{(1)}\right)^2\left(1-C_4\frac{\log n}{n}\right)}{N_n^{(1)}+k_u^{(1)}}.\end{aligned}\tag{3.22}$$

Note that $k_u^{(i)}=1$. By the convergence of $\delta_n^{(i)}$ and $\delta_n^{(j)}$, for all large enough n, we have

$$\begin{aligned}\left(\delta_n^{(j)}-\delta_n^{(i)}\right)\left(\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)-\delta_n^{(i)}\right)&>0,\\ \left(\delta_n^{(j)}-\delta_n^{(i)}\right)\left(\delta_n^{(j)}\left(1-C_4\frac{\log n}{n}\right)-\delta_n^{(i)}\right)&>0.\end{aligned}$$

If $\lim_{n\to\infty}\frac{\delta_n^{(j)}}{\delta_n^{(i)}}>1$, i.e., $\mu^{(j)}<\mu^{(i)}$, then by (3.21) we have

$$\begin{aligned}\frac{\delta_{n+u}^{(i)}\left(\lambda^{(j)}\right)^2}{N_n^{(j)}+k_u^{(j)}} &= \frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(i)}\left(\lambda^{(j)}\right)^2}{N_n^{(j)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}\\ &\le \frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}\left(\lambda^{(i)}\right)^2\left(1+C_2\frac{\log n}{n}\right)}{N_n^{(i)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}\\ &\quad+\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}\left(\lambda^{(1)}\right)^2\left(1+C_2\frac{\log n}{n}\right)-\delta_n^{(i)}\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}\\ &= \frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}\left(\lambda^{(i)}\right)^2\left(1-C_4\frac{\log n}{n}\right)}{N_n^{(i)}+1}\frac{\left(1+C_2\frac{\log n}{n}\right)}{\left(1-C_4\frac{\log n}{n}\right)}\frac{N_n^{(i)}+1}{N_n^{(i)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}\\ &\quad+\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}\left(\lambda^{(1)}\right)^2\left(1+C_2\frac{\log n}{n}\right)-\delta_n^{(i)}\left(\lambda^{(1)}\right)^2}{N_n^{(1)}+k_u^{(1)}}\frac{N_n^{(1)}+k_u^{(1)}}{N_n^{(1)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}.\end{aligned}$$

It follows that there must exist a positive constant C_5 such that

$$\begin{aligned}\frac{\delta_{n+u}^{(i)}\left(\lambda^{(j)}\right)^2}{N_n^{(j)}+k_u^{(j)}} &\le \frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}\left(\lambda^{(i)}\right)^2\left(1-C_4\frac{\log n}{n}\right)}{N_n^{(i)}+1}\left(1+C_5\frac{\log n}{n}\right)\frac{N_n^{(i)}+1}{N_n^{(i)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}\\ &\quad+\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}\left(\lambda^{(1)}\right)^2\left(1+C_2\frac{\log n}{n}\right)-\delta_n^{(i)}\left(\lambda^{(1)}\right)^2}{N_n^{(1)}+k_u^{(1)}}\frac{N_n^{(1)}+k_u^{(1)}}{N_n^{(1)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}.\end{aligned}$$

Thus, to satisfy (3.22), it is sufficient to have

$$\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}}{\delta_{n+u}^{(j)}}\left(1+C_5\frac{\log n}{n}\right)\frac{N_n^{(i)}+1}{N_n^{(i)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}\ \le\ 1,\tag{3.23}$$

$$\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}} \frac{\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)-\delta_n^{(i)}}{\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)-\delta_{n+u}^{(i)}} \frac{N_n^{(1)}+k_u^{(1)}}{N_n^{(1)}} \frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}} \quad \leq \quad 1. \tag{3.24}$$

Note that for all large enough n and any alternative $i \neq 1$, by Lemma 3.4.2, we have

$$\begin{aligned}
\left|\delta_{n+u}^{(i)}-\delta_n^{(i)}\right| &= \left|\left(d_{n+u}^{(i)}\right)^2-\left(d_n^{(i)}\right)^2\right| \\
&= \left|\left(\theta_{n+u}^{(i)}-\theta_{n+u}^{(1)}\right)^2-\left(\theta_n^{(i)}-\theta_n^{(1)}\right)^2\right| \\
&= \left|\left(\theta_{n+u}^{(i)}-\theta_{n+u}^{(1)}\right)+\left(\theta_n^{(i)}-\theta_n^{(1)}\right)\right|\left|\left(\theta_{n+u}^{(i)}-\theta_n^{(i)}\right)-\left(\theta_{n+u}^{(1)}-\theta_n^{(1)}\right)\right| \\
&\leq \left|\left(\theta_{n+u}^{(i)}-\theta_{n+u}^{(1)}\right)+\left(\theta_n^{(i)}-\theta_n^{(1)}\right)\right| \\
&\quad \cdot\left(\left|\theta_{n+u}^{(i)}-\mu^{(i)}\right|+\left|\theta_n^{(i)}-\mu^{(i)}\right|+\left|\theta_{n+u}^{(1)}-\mu^{(1)}\right|+\left|\theta_n^{(1)}-\mu^{(1)}\right|\right) \\
&= O\left(\sqrt{\frac{\log\log N_n^{(i)}}{N_n^{(i)}}}\right)+O\left(\sqrt{\frac{\log\log N_n^{(1)}}{N_n^{(1)}}}\right) \\
&= O\left(\sqrt{\frac{\log\log n}{n}}\right),
\end{aligned}$$

where the fourth equality holds because of the law of the iterated logarithm, and the last equality holds by Lemma 3.4.2. Then for all large enough n, we have

$$\begin{aligned}
\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}} &= \frac{\delta_{n+u}^{(i)}-\delta_n^{(i)}+\delta_n^{(i)}}{\delta_n^{(i)}}=1+\frac{\delta_{n+u}^{(i)}-\delta_n^{(i)}}{\delta_n^{(i)}}=1+O\left(\sqrt{\frac{\log\log n}{n}}\right), \\
\frac{\delta_n^{(j)}}{\delta_{n+u}^{(j)}} &= 1+O\left(\sqrt{\frac{\log\log n}{n}}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)-\delta_n^{(i)}}{\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)-\delta_{n+u}^{(i)}} \\
&\quad= 1+\frac{\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)-\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)-\left(\delta_n^{(i)}-\delta_{n+u}^{(i)}\right)}{\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)-\delta_{n+u}^{(i)}} \\
&\quad\leq 1+\frac{\left|\delta_n^{(j)}-\delta_{n+u}^{(j)}\right|+\left|\delta_n^{(j)}C_2\frac{\log n}{n}\right|+\left|\delta_{n+u}^{(j)}C_4\frac{\log n}{n}\right|+\left|\delta_n^{(i)}-\delta_{n+u}^{(i)}\right|}{\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)-\delta_{n+u}^{(i)}} \\
&\quad= 1+O\left(\sqrt{\frac{\log\log n}{n}}\right).
\end{aligned}$$

Then together with Lemma 3.4.2, there exists a positive constant C_6 such that, for all large enough n, the LHS of (3.23) satisfies

$$\begin{aligned}
&\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}}{\delta_{n+u}^{(j)}}\left(1+C_5\frac{\log n}{n}\right)\frac{N_n^{(i)}+1}{N_n^{(i)}}N_n^{(j)}-N_n^{(j)} \\
&= \left(1+O\left(\sqrt{\frac{\log\log n}{n}}\right)\right)\left(1+C_5\frac{\log n}{n}\right)\left(1+O\left(\frac{1}{n}\right)\right)N_n^{(j)}-N_n^{(j)} \\
&\leq C_6\sqrt{n\log\log n},
\end{aligned}$$

while the LHS of (3.24) satisfies

$$\begin{aligned}
&\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\frac{\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)-\delta_n^{(i)}}{\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)-\delta_{n+u}^{(i)}}\frac{N_n^{(1)}+k_u^{(1)}}{N_n^{(1)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}} \\
&= \left(1+O\left(\sqrt{\frac{\log\log n}{n}}\right)\right)\frac{N_n^{(1)}+k_u^{(1)}}{N_n^{(1)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}} \\
&\leq \left(1+C_6\frac{\sqrt{n\log\log n}}{n}\right)\frac{N_n^{(1)}+k_u^{(1)}}{N_n^{(1)}}\frac{N_n^{(j)}}{N_n^{(j)}+k_u^{(j)}}.
\end{aligned}$$

Therefore, to satisfy (3.23), it is sufficient to have

$$C_6\sqrt{n\log\log n}\leq k_u^{(j)}. \tag{3.25}$$

Now define

$$s\triangleq\sup\left\{l<m: I_{n+l}^{(1)}=0\right\}. \tag{3.26}$$

Since $\frac{k_m^{(1)}}{\sqrt{n\log\log n}}$ can be arbitrarily large, we can suppose that $k_s^{(1)}>C_7\sqrt{n\log\log n}$, where C_7 is a positive constant to be specified. By Lemma 3.4.5, since C_6 is a fixed positive constant, there must exist a constant C_8 such that, if $C_7\geq C_8$, there exists a suboptimal $j\neq i$, and a stage $n+u$ with $u\leq s$, such that j is sampled at stage $n+u$ and

$$\left(1+C_6\frac{\sqrt{n\log\log n}}{n}\right)\frac{N_n^{(j)}}{N_n^{(1)}}<\frac{N_n^{(j)}+k_u^{(j)}}{N_n^{(1)}+k_s^{(1)}}\leq\frac{N_n^{(j)}+k_u^{(j)}}{N_n^{(1)}+k_u^{(1)}}.$$

Then, (3.24) holds at stage $n+u$. At the same time, since

$$\frac{N_n^{(j)} + k_u^{(j)}}{N_n^{(1)} + k_s^{(1)}} > \left(1 + C_6 \frac{\sqrt{n \log \log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}} \geq \frac{N_n^{(j)}}{N_n^{(1)}},$$

we have $\frac{k_u^{(j)}}{k_s^{(1)}} \geq \frac{N_n^{(j)}}{N_n^{(1)}}$. From Lemma 3.4.2, there must exist a positive constant C_9 such that, for all large enough n,

$$k_u^{(j)} \geq C_9 k_s^{(1)} \geq C_9 C_7 \sqrt{n \log \log n}.$$

Now let $C_7 = \max\left\{C_8, \frac{C_6}{C_9}\right\}$. Then, both (3.24) and (3.25) are satisfied at stage $n+u$, so (3.22) is satisfied, which means

$$r_{n+u}^{(i,j)} < 1 \qquad \Rightarrow \qquad v_{n+u}^{(i)} > v_{n+u}^{(j)}.$$

But the alternative j is sampled at stage $n+u$, which means $v_{n+u}^{(i)} \leq v_{n+u}^{(j)}$. The desired contradiction follows.

Now, consider the other case where $\lim_{n\to\infty} \frac{\delta_n^{(j)}}{\delta_n^{(i)}} < 1$, i.e., $\mu^{(j)} > \mu^{(i)}$. By (3.21), we have

$$\begin{aligned}
\frac{\delta_{n+u}^{(j)} \left(\lambda^{(i)}\right)^2 \left(1 - C_4 \frac{\log n}{n}\right)}{N_n^{(i)} + 1} &= \frac{\delta_{n+u}^{(j)}}{\delta_n^{(j)}} \frac{\delta_n^{(j)} \left(\lambda^{(i)}\right)^2 \left(1 + C_2 \frac{\log n}{n}\right)}{N_n^{(i)}} \frac{1 - C_4 \frac{\log n}{n}}{1 + C_2 \frac{\log n}{n}} \frac{N_n^{(i)}}{N_n^{(i)} + 1} \\
&\geq \frac{\delta_{n+u}^{(j)}}{\delta_n^{(j)}} \frac{\delta_n^{(i)} \left(\lambda^{(j)}\right)^2}{N_n^{(j)}} \frac{1 - C_4 \frac{\log n}{n}}{1 + C_2 \frac{\log n}{n}} \frac{N_n^{(i)}}{N_n^{(i)} + 1} \\
&\quad + \frac{\delta_{n+u}^{(j)}}{\delta_n^{(j)}} \frac{\delta_n^{(i)} \left(\lambda^{(1)}\right)^2 - \delta_n^{(j)} \left(\lambda^{(1)}\right)^2 \left(1 + C_2 \frac{\log n}{n}\right)}{N_n^{(1)}} \\
&\quad \cdot \frac{1 - C_4 \frac{\log n}{n}}{1 + C_2 \frac{\log n}{n}} \frac{N_n^{(i)}}{N_n^{(i)} + 1}.
\end{aligned}$$

Then, there must exist a positive constant C_{10} such that, for all large enough n,

$$\frac{\delta_{n+u}^{(j)} \left(\lambda^{(i)}\right)^2 \left(1 - C_4 \frac{\log n}{n}\right)}{N_n^{(i)} + 1} \geq \frac{\delta_{n+u}^{(j)}}{\delta_n^{(j)}} \frac{\delta_n^{(i)} \left(\lambda^{(j)}\right)^2}{N_n^{(j)}} \frac{1}{1 + C_{10} \frac{\log n}{n}} \frac{N_n^{(i)}}{N_n^{(i)} + 1}$$

$$
\begin{aligned}
&+\frac{\delta_{n+u}^{(j)}}{\delta_n^{(j)}}\frac{\delta_n^{(i)}\left(\lambda^{(1)}\right)^2-\delta_n^{(j)}\left(\lambda^{(1)}\right)^2\left(1+C_2\frac{\log n}{n}\right)}{N_n^{(1)}}\\
&\cdot\frac{1}{1+C_{10}\frac{\log n}{n}}\frac{N_n^{(i)}}{N_n^{(i)}+1}.
\end{aligned}
$$

Thus, to satisfy (3.22), for all large enough n, it is sufficient to have

$$
\begin{aligned}
\frac{\delta_{n+u}^{(j)}}{\delta_n^{(j)}}\frac{\delta_n^{(i)}}{\delta_{n+u}^{(i)}}\frac{1}{N_n^{(j)}}\frac{1}{1+C_{10}\frac{\log n}{n}}\frac{N_n^{(i)}}{N_n^{(i)}+1} &\geq \frac{1}{N_n^{(j)}+k_u^{(j)}},\\
\frac{\delta_{n+u}^{(j)}}{\delta_n^{(j)}}\frac{\delta_n^{(i)}-\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)}{N_n^{(1)}}\frac{1}{1+C_{10}\frac{\log n}{n}}\frac{N_n^{(i)}}{N_n^{(i)}+1} &\geq \frac{\delta_{n+u}^{(i)}-\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)}{N_n^{(1)}+k_u^{(1)}},
\end{aligned}
$$

which can equivalently be rewritten as

$$
\begin{aligned}
k_u^{(j)} &\geq \frac{\delta_n^{(j)}}{\delta_{n+u}^{(j)}}\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\left(1+C_{10}\frac{\log n}{n}\right)\frac{N_n^{(i)}+1}{N_n^{(i)}}N_n^{(j)}-N_n^{(j)}, && (3.27)\\
k_u^{(1)} &\geq \frac{\delta_n^{(j)}}{\delta_{n+u}^{(j)}}\left(1+C_{10}\frac{\log n}{n}\right)\frac{\delta_{n+u}^{(i)}-\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)}{\delta_n^{(i)}-\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)}\frac{N_n^{(i)}+1}{N_n^{(i)}}N_n^{(1)}\\
&\quad -N_n^{(1)}. && (3.28)
\end{aligned}
$$

Similarly as above, by Lemma 3.4.2, there exist positive constants C_{11}, C_{12}, C_{13} and C_{14} such that, for all large enough n,

$$
\frac{\delta_n^{(j)}}{\delta_{n+u}^{(j)}}\frac{\delta_{n+u}^{(i)}}{\delta_n^{(i)}}\left(1+C_{10}\frac{\log n}{n}\right)\frac{N_n^{(i)}+1}{N_n^{(i)}}N_n^{(j)}-N_n^{(j)} \leq C_{11}\sqrt{n\log\log n},
$$

and

$$
\begin{aligned}
&\frac{\delta_n^{(j)}}{\delta_{n+u}^{(j)}}\left(1+C_{10}\frac{\log n}{n}\right)\frac{\delta_{n+u}^{(i)}-\delta_{n+u}^{(j)}\left(1-C_4\frac{\log n}{n}\right)}{\delta_n^{(i)}-\delta_n^{(j)}\left(1+C_2\frac{\log n}{n}\right)}\frac{N_n^{(i)}+1}{N_n^{(i)}}N_n^{(1)}-N_n^{(1)}\\
&\leq \left(1+C_{10}\frac{\log n}{n}\right)\left(1+C_{12}\sqrt{\frac{\log\log n}{n}}\right)\frac{N_n^{(i)}+1}{N_n^{(i)}}N_n^{(1)}-N_n^{(1)}\\
&\leq \left(1+C_{13}\sqrt{\frac{\log\log n}{n}}\right)\frac{N_n^{(i)}+1}{N_n^{(i)}}N_n^{(1)}-N_n^{(1)}\\
&\leq C_{14}\sqrt{n\log\log n}.
\end{aligned}
$$

Therefore, to satisfy (3.27) and (3.28), it is sufficient to have

$$k_u^{(j)} \geq C_{11}\sqrt{n \log\log n}, \tag{3.29}$$

$$k_u^{(1)} \geq C_{14}\sqrt{n \log\log n}. \tag{3.30}$$

Again, define s as in (3.26). Since $\frac{k_m^{(1)}}{\sqrt{n \log\log n}}$ can be arbitrarily large, we can suppose that $k_s^{(1)} > C_{15}\sqrt{n \log\log n}$, where C_{15} is a positive constant to be specified. By Lemma 3.4.5, since C_{11} is a fixed positive constant, there must exist a constant C_{16} such that, if $C_{15} \geq C_{16}$, there exists a suboptimal alternative $j \neq i$, and a stage $n+u$ with $u \leq s$, such that j is sampled at stage $n+u$ and

$$\left(1 + C_{11}\frac{\sqrt{n \log\log n}}{n}\right)\frac{N_n^{(j)}}{N_n^{(1)}} < \frac{N_n^{(j)} + k_u^{(j)}}{N_n^{(1)} + k_s^{(1)}} \leq \frac{N_n^{(j)} + k_u^{(j)}}{N_n^{(1)} + k_u^{(1)}},$$

whence

$$\frac{N_n^{(j)} + k_u^{(j)}}{N_n^{(1)} + k_s^{(1)}} > \left(1 + C_{11}\frac{\sqrt{n \log\log n}}{n}\right)\frac{N_n^{(j)}}{N_n^{(1)}} \geq \frac{N_n^{(j)}}{N_n^{(1)}}.$$

Then, we have $\frac{k_u^{(j)}}{k_s^{(1)}} \geq \frac{N_n^{(j)}}{N_n^{(1)}}$. From Lemma 3.4.2, there must exist a positive constant C_{17} such that for all large enough n,

$$k_u^{(j)} \geq C_{17}k_s^{(1)} \geq C_{17}C_{15}\sqrt{n \log\log n}.$$

At the same time, by Lemma 3.4.3, for all large enough n, we also have

$$k_u^{(1)} \geq \frac{k_u^{(j)} + 1}{B_2} - 1 \geq \frac{C_{17}C_{15}\sqrt{n \log\log n} + 1}{B_2} - 1 \geq \frac{C_{17}C_{15}\sqrt{n \log\log n}}{2B_2}.$$

Now, let $C_{15} = \max\left\{C_{16}, \frac{C_{11}}{C_{17}}, \frac{2B_2C_{14}}{C_{17}}\right\}$. Then both (3.29) and (3.30) are satisfied at stage $n+u$, so (3.22) is satisfied, which means that

$$r_{n+u}^{(i,j)} < 1 \qquad \Rightarrow \qquad v_{n+u}^{(i)} > v_{n+u}^{(j)}.$$

But the alternative j is sampled at stage $n+u$, which means that $v_{n+u}^{(i)} \leq v_{n+u}^{(j)}$. Again, we have the desired contradiction. □

Proof of Lemma 3.4.5. For convenience, we abbreviate $k_{(n,n+m)}^{(j)}$ by the notation $k_m^{(j)}$ for all j. First, since C_2 is a constant and $\lim_{n\to\infty} \frac{\sqrt{n \log\log n}}{n} = 0$, it follows that, for all large enough n, we must have $C_2\sqrt{n\log\log n} \leq n$. Intuitively, from the definition of m and s, stage $n+m$ is the first time that alternative i is sampled after stage n, and stage $n+s$ is the last time that a suboptimal alternative is sampled before stage $n+m$. Recall that, by assumption, we must have $C_2\sqrt{n\log\log n} \leq k_s^{(1)} \leq n$ for some positive constant C_2 to be specified.

At stage n, since we sample a suboptimal i by assumption, we must have

$$\left(N_n^{(1)}/\lambda^{(1)}\right)^2 \geq \sum_{j=2}^{M} \left(N_n^{(j)}/\lambda^{(j)}\right)^2. \tag{3.31}$$

At stage $n+s$, from the definition of s, it is also some suboptimal alternative that is sampled. Repeating the arguments in the proof of Theorem 3.3.1, we obtain

$$\left(\frac{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}{n+s}\right)^2 - \sum_{j=2}^{M}\left(\frac{\left(N_n^{(j)}+k_s^{(j)}\right)/\lambda^{(j)}}{n+s}\right)^2 \leq \frac{C_3}{n}$$

for some fixed positive constant C_3. Note that $k_s^{(i)} = 1$, whence

$$\sum_{j\geq 2, j\neq i}\left(\frac{\left(N_n^{(j)}+k_s^{(j)}\right)/\lambda^{(j)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2 + \left(\frac{\left(N_n^{(i)}+1\right)/\lambda^{(i)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2$$
$$+\frac{C_3}{n}\left(\frac{n+s}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2 \geq 1.$$

From Lemma 3.4.2, we know that $\liminf_{n\to\infty} \frac{N_n^{(1)}}{n} > 0$. Then, there must exist some constant C_4 such that

$$C_3\left(\frac{n+s}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2 \le C_4,$$

whence

$$\sum_{j\ge 2, j\ne i}\left(\frac{\left(N_n^{(j)}+k_s^{(j)}\right)/\lambda^{(j)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2 \ge 1-\left(\frac{\left(N_n^{(i)}+1\right)/\lambda^{(i)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2-\frac{C_4}{n},$$

and for all large enough n,

$$\begin{align*}
&\sum_{j\ge 2, j\ne i}\left[\left(\frac{\left(N_n^{(j)}+k_s^{(j)}\right)/\lambda^{(j)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2-\left(\frac{N_n^{(j)}/\lambda^{(j)}}{N_n^{(1)}/\lambda^{(1)}}\right)^2\right] \\
\ge\ & 1-\sum_{j\ge 2, j\ne i}\left(\frac{N_n^{(j)}/\lambda^{(j)}}{N_n^{(1)}/\lambda^{(1)}}\right)^2-\left(\frac{\left(N_n^{(i)}+1\right)/\lambda^{(i)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2-\frac{C_4}{n} \\
\ge\ & \left(\frac{N_n^{(i)}/\lambda^{(i)}}{N_n^{(1)}/\lambda^{(1)}}\right)^2-\left(\frac{\left(N_n^{(i)}+1\right)/\lambda^{(i)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2-\frac{C_4}{n} \tag{3.32} \\
=\ & \left(\frac{\lambda^{(1)}}{\lambda^{(i)}}\right)^2\frac{\left(N_n^{(i)}\right)^2\left(N_n^{(1)}+k_s^{(1)}\right)^2-\left(N_n^{(1)}\right)^2\left(N_n^{(i)}+1\right)^2}{\left(N_n^{(1)}+k_s^{(1)}\right)^2\left(N_n^{(1)}\right)^2}-\frac{C_4}{n} \\
=\ & \left(\frac{\lambda^{(1)}}{\lambda^{(i)}}\right)^2\frac{\left(N_n^{(i)}\right)^2\left(2N_n^{(1)}k_s^{(1)}+\left(k_s^{(1)}\right)^2\right)-\left(N_n^{(1)}\right)^2\left(2N_n^{(i)}+1\right)}{\left(N_n^{(1)}+k_s^{(1)}\right)^2\left(N_n^{(1)}\right)^2}-\frac{C_4}{n} \\
=\ & \left(\frac{\lambda^{(1)}}{\lambda^{(i)}}\right)^2\frac{\left(N_n^{(i)}\right)^2\left(2N_n^{(1)}\left(k_s^{(1)}-\frac{N_n^{(1)}}{N_n^{(i)}}\right)+\left(k_s^{(1)}\right)^2-\left(\frac{N_n^{(1)}}{N_n^{(i)}}\right)^2\right)}{\left(N_n^{(1)}+k_s^{(1)}\right)^2\left(N_n^{(1)}\right)^2}-\frac{C_4}{n} \\
>\ & \left(\frac{\lambda^{(1)}}{\lambda^{(i)}}\right)^2\frac{\left(N_n^{(i)}\right)^2\left(2N_n^{(1)}\left(\frac{k_s^{(1)}}{2}\right)+\frac{1}{2}\left(k_s^{(1)}\right)^2\right)}{\left(N_n^{(1)}+k_s^{(1)}\right)^2\left(N_n^{(1)}\right)^2}-\frac{C_4}{n} \tag{3.33} \\
=\ & \frac{1}{2}\left(\frac{\lambda^{(1)}}{\lambda^{(i)}}\right)^2\frac{1}{\left(N_n^{(1)}+k_s^{(1)}\right)^2}\left(\frac{N_n^{(i)}}{N_n^{(1)}}\right)^2\left(2N_n^{(1)}k_s^{(1)}+\left(k_s^{(1)}\right)^2\right)-\frac{C_4}{n},
\end{align*}$$

where (3.32) holds due to (3.31), while (3.33) holds since $\liminf_{n\to\infty}\frac{N_n^{(i)}}{N_n^{(1)}}>0$ and $k_s^{(1)}\ge C_2\sqrt{n\log\log n}$ for a positive constant C_2. Since $\liminf_{n\to\infty}\frac{N_n^{(i)}}{N_n^{(1)}}>0$ and

$\liminf_{n\to\infty} \frac{N_n^{(1)}}{n} > 0$, there must exist positive constants C_5, C_6, C_7, C_8 and C_9 such that, for all large enough n, we have

$$\begin{aligned}
&\frac{1}{2}\left(\frac{\lambda^{(1)}}{\lambda^{(i)}}\right)^2 \frac{1}{\left(N_n^{(1)}+k_s^{(1)}\right)^2}\left(\frac{N_n^{(i)}}{N_n^{(1)}}\right)^2\left(2N_n^{(1)}k_s^{(1)}+\left(k_s^{(1)}\right)^2\right)-\frac{C_4}{n} \\
\geq\ & C_5\frac{1}{\left(N_n^{(1)}+k_s^{(1)}\right)^2}\left(2N_n^{(1)}k_s^{(1)}+\left(k_s^{(1)}\right)^2\right)-\frac{C_4}{N_n^{(1)}} \\
=\ & \frac{C_5}{\left(N_n^{(1)}+k_s^{(1)}\right)^2}\left(2N_n^{(1)}k_s^{(1)}+\left(k_s^{(1)}\right)^2-C_6\frac{\left(N_n^{(1)}+k_s^{(1)}\right)^2}{N_n^{(1)}}\right) \\
\geq\ & \frac{C_5}{\left(N_n^{(1)}+k_s^{(1)}\right)^2}\left(2N_n^{(1)}k_s^{(1)}+\left(k_s^{(1)}\right)^2-2C_7N_n^{(1)}\right) && (3.34) \\
\geq\ & \frac{C_5}{\left(N_n^{(1)}+k_s^{(1)}\right)^2}2\left(k_s^{(1)}-C_7\right)N_n^{(1)} \\
\geq\ & \frac{C_8\left(k_s^{(1)}-C_7\right)}{N_n^{(1)}} && (3.35) \\
\geq\ & \frac{C_8\left(C_2\sqrt{n\log\log n}-C_7\right)}{n} \\
\geq\ & \frac{C_9C_2\sqrt{n\log\log n}}{n},
\end{aligned}$$

where (3.34) and (3.35) hold because $k_s^{(1)} \leq n$. Then,

$$\sum_{j\geq 2, j\neq i}\left[\left(\frac{\left(N_n^{(j)}+k_s^{(j)}\right)/\lambda^{(j)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2-\left(\frac{N_n^{(j)}/\lambda^{(j)}}{N_n^{(1)}/\lambda^{(1)}}\right)^2\right] > \frac{C_9C_2\sqrt{n\log\log n}}{n},$$

so there must be some suboptimal j such that

$$\left(\frac{\left(N_n^{(j)}+k_s^{(j)}\right)/\lambda^{(j)}}{\left(N_n^{(1)}+k_s^{(1)}\right)/\lambda^{(1)}}\right)^2-\left(\frac{N_n^{(j)}/\lambda^{(j)}}{N_n^{(1)}/\lambda^{(1)}}\right)^2 > \frac{1}{M-2}\frac{C_9C_2\sqrt{n\log\log n}}{n}.$$

Let $C_{10} = \frac{C_9}{M-2}$ and $C_{11} = \frac{C_{10}C_2}{4}$. Then,

$$\left(\frac{\left(N_n^{(j)}+k_s^{(j)}\right)/\left(N_n^{(1)}+k_s^{(1)}\right)}{N_n^{(j)}/N_n^{(1)}}\right)^2 > 1+\frac{C_{10}C_2\sqrt{n\log\log n}}{n},$$

and, for all large enough n, we have

$$\frac{\left(N_n^{(j)} + k_s^{(j)}\right) / \left(N_n^{(1)} + k_s^{(1)}\right)}{N_n^{(j)}/N_n^{(1)}} > 1 + \frac{C_{11}\sqrt{n \log\log n}}{n}, \tag{3.36}$$

whence

$$\frac{N_n^{(j)} + k_s^{(j)}}{N_n^{(1)} + k_s^{(1)}} > \left(1 + C_{11}\frac{\sqrt{n \log\log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}}. \tag{3.37}$$

For the alternative j that satisfies (3.37), let

$$u \triangleq \sup\left\{ l \le s : I_{n+l}^{(j)} = 1 \right\}.$$

Then, stage $n+u$ is the last time that alternative j is sampled before or at stage $n+m$. Since $k_s^{(j)}$ is monotonically increasing in s, we have

$$\begin{aligned}
\frac{N_n^{(j)} + k_u^{(j)}}{N_n^{(1)} + k_u^{(1)}} &\ge \frac{N_n^{(j)} + k_u^{(j)}}{N_n^{(1)} + k_s^{(1)}} \\
&\ge \frac{N_n^{(j)} + k_s^{(j)} - 1}{N_n^{(1)} + k_s^{(1)}} \\
&= \left(1 - \frac{1}{N_n^{(j)} + k_s^{(j)}}\right) \frac{N_n^{(j)} + k_s^{(j)}}{N_n^{(1)} + k_s^{(1)}} \\
&> \left(1 - \frac{1}{N_n^{(j)} + k_s^{(j)}}\right) \left(1 + C_{11}\frac{\sqrt{n \log\log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}},
\end{aligned}$$

where the last line follows from (3.37). By Lemma 3.4.2, there must exist a positive constant C_{12} such that, for all large enough n,

$$\begin{aligned}
&\left(1 - \frac{1}{N_n^{(j)} + k_s^{(j)}}\right) \left(1 + C_{11}\frac{\sqrt{n \log\log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}} \\
&\ge \left(1 - \frac{C_{12}}{n}\right) \left(1 + C_{11}\frac{\sqrt{n \log\log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}} \\
&= \left(1 + C_{11}\frac{\sqrt{n \log\log n}}{n} - \frac{C_{12}}{n} - C_{12}C_{11}\frac{\sqrt{n \log\log n}}{n^2}\right) \frac{N_n^{(j)}}{N_n^{(1)}} \\
&\ge \left(1 + \frac{C_{11}}{2}\frac{\sqrt{n \log\log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}} \\
&= \left(1 + C_{13}\frac{\sqrt{n \log\log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}},
\end{aligned}$$

where $C_{13} = \frac{C_{11}}{2} = \frac{C_{10}C_2}{8}$. Note that constants C_3 through C_{10} are fixed and do not depend on C_1 or C_2. Thus, for all large enough n, if we take C_2 to be sufficiently large, i.e., $C_2 \geq 8C_1/C_{10}$, to make $C_{13} \geq C_1$, then

$$\frac{N_n^{(j)} + k_u^{(j)}}{N_n^{(1)} + k_s^{(1)}} > \left(1 + C_1 \frac{\sqrt{n \log\log n}}{n}\right) \frac{N_n^{(j)}}{N_n^{(1)}},$$

which completes the proof. □

Lemma 3.4.6. *For any alternative i, $n^{3/4}\left|\delta_{n+1}^{(i)} - \delta_n^{(i)}\right| \to 0$ almost surely as $n \to \infty$.*

Proof. First, if an alternative j other than 1 or i is sampled at stage n, it is obvious that $n^{3/4}\left|\delta_{n+1}^{(i)} - \delta_n^{(i)}\right| = 0$.

Second, if alternative i is sampled at stage n, then for all large enough n, there exists a constant C_1 such that

$$\begin{aligned} n^{3/4}\left|\delta_{n+1}^{(i)} - \delta_n^{(i)}\right| &= n^{3/4}\left|\left(d_{n+1}^{(i)}\right)^2 - \left(d_n^{(i)}\right)^2\right| \\ &\leq C_1 n^{3/4}\left|d_{n+1}^{(i)} - d_n^{(i)}\right| \\ &= \frac{C_1}{n^{1/4}} n \left|\theta_{n+1}^{(i)} - \theta_n^{(i)}\right|, \end{aligned}$$

where

$$\begin{aligned} n\left|\theta_{n+1}^{(i)} - \theta_n^{(i)}\right| &= n\left|\frac{1}{N_n^{(i)} + 1}\left(N_n^{(i)}\theta_n^{(i)} + W_{n+1}^{(i)}\right) - \theta_n^{(i)}\right| \\ &\leq \frac{n}{N_n^{(i)} + 1}\left|\theta_n^{(i)}\right| + \frac{n}{N_n^{(i)} + 1}\left|W_{n+1}^{(i)}\right| \\ &= O(1)\left(1 + \left|W_{n+1}^{(i)}\right|\right), \end{aligned}$$

thus there exists a constant C_2 such that

$$n^{3/4}\left|\delta_{n+1}^{(i)} - \delta_n^{(i)}\right| \leq \frac{C_2}{n^{1/4}}\left(1 + \left|W_{n+1}^{(i)}\right|\right).$$

Finally, if alternative 1 is sampled at stage n, then similarly as above, for all large enough n, there exist constants C_3 and C_4 such that

$$\begin{aligned} n^{3/4}\left|\delta^{(i)}_{n+1}-\delta^{(i)}_{n}\right| &\leq \frac{C_3}{n^{1/4}}n\left|\theta^{(1)}_{n+1}-\theta^{(1)}_{n}\right| \\ &\leq \frac{C_4}{n^{1/4}}\left(1+\left|W^{(1)}_{n+1}\right|\right). \end{aligned}$$

Then it is sufficient to show $\frac{\left|W^{(i)}_{n+1}\right|}{n^{1/4}} \to 0$ and $\frac{\left|W^{(1)}_{n+1}\right|}{n^{1/4}} \to 0$ almost surely. By Markov's inequality, for all $\varepsilon > 0$,

$$P\left(\frac{\left|W^{(i)}_{n+1}\right|}{n^{1/4}} \geq \varepsilon\right) \leq \mathbb{E}\left(\frac{\left(W^{(i)}_{n+1}\right)^8}{n^2\varepsilon^8}\right) \leq \frac{C_5}{n^2\varepsilon^8},$$

where C_5 is a fixed constant, thus $\frac{\left|W^{(i)}_{n+1}\right|}{n^{1/4}} \to 0$ in probability. Furthermore, by the Borel-Cantelli lemma, since

$$\sum_n P\left(\frac{\left|W^{(i)}_{n+1}\right|}{n^{1/4}} \geq \varepsilon\right) \leq \sum_n \frac{C_5}{n^2\varepsilon^8} < \infty,$$

then we have $\frac{\left|W^{(i)}_{n+1}\right|}{n^{1/4}} \to 0$ almost surely. Using similar arguments, we also have $\frac{\left|W^{(1)}_{n+1}\right|}{n^{1/4}} \to 0$ almost surely, completing the proof. □

Let $i, j > 1$ and suppose that i is sampled at stage n. We will first place an $O\left(\frac{1}{n^{3/4}}\right)$ bound on the increment $r^{(i,j)}_{n+1} - r^{(i,j)}_{n}$. We will then place a bound of $O\left(\frac{\sqrt{n\log\log n}}{n^{3/4}}\right)$ on the growth of $\left(r^{(i,j)}_n\right)$ in between two samples of i (note that, by definition, $r^{(i,j)}_n \leq 1$ at any stage n when i is sampled). As this bound vanishes to zero as $n \to \infty$, it will then be shown to follow that $r^{(i,j)}_n \to 1$.

If i is sampled at stage n, then $r^{(i,j)}_n \leq 1$ and

$$r^{(i,j)}_{n+1} - r^{(i,j)}_{n} = \frac{\log\left(v^{(i)}_{n+1}\right)}{\log\left(v^{(j)}_{n+1}\right)} - \frac{\log\left(v^{(i)}_{n}\right)}{\log\left(v^{(j)}_{n}\right)}$$

$$
\begin{aligned}
&= \frac{\log\left(v^{(i)}_{n+1}\right) - \log\left(v^{(i)}_{n}\right)}{\log\left(v^{(j)}_{n}\right)} \\
&\leq \frac{\left|\log\left(v^{(i)}_{n+1}\right) - \log\left(v^{(i)}_{n}\right)\right|}{\left|\log\left(v^{(j)}_{n}\right)\right|} \\
&= \frac{n^{1/4}}{2\left|\log\left(v^{(j)}_{n}\right)\right|}\frac{1}{n^{1/4}}\left|\left(\frac{\delta^{(i)}_{n+1}}{\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n+1}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}} - \frac{\delta^{(i)}_{n}}{\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}}\right)\right. \\
&\quad +3\left(\log\frac{\delta^{(i)}_{n+1}}{\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n+1}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}} - \log\frac{\delta^{(i)}_{n}}{\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}}\right) \\
&\quad -2\left[\log\left(1+O\left(\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n+1}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}\right)\right)\right. \\
&\quad \left.-\log\left(1+O\left(\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}\right)\right)\right] \\
&\quad \left.-\left(\log\delta^{(i)}_{n+1} - \log\delta^{(i)}_{n}\right)\right|. \qquad (3.38)
\end{aligned}
$$

By Lemma 3.4.2, there exists a positive constant C_1 such that, for all large enough n,

$$
\begin{aligned}
\frac{2\left|\log\left(v^{(j)}_{n}\right)\right|}{n^{1/4}} &= \frac{t^{(j)}_n}{n^{1/4}}\left(1+O\left(\frac{\log t^{(j)}_n}{t^{(j)}_n}\right)\right) \\
&> \frac{1}{2n^{1/4}}\frac{\delta^{(j)}_n}{\frac{\left(\lambda^{(j)}\right)^2}{N^{(j)}_n}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}} \\
&\geq C_1\frac{n}{n^{1/4}} = C_1 n^{3/4}.
\end{aligned}
$$

On the other hand, for all large enough n, there also exists a positive constant C_2 such that

$$
\frac{1}{n^{1/4}}\left|\frac{\delta^{(i)}_{n+1}}{\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n+1}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}} - \frac{\delta^{(i)}_{n}}{\frac{\left(\lambda^{(i)}\right)^2}{N^{(i)}_n}+\frac{\left(\lambda^{(1)}\right)^2}{N^{(1)}_n}}\right|
$$

$$
\begin{aligned}
&\leq \frac{1}{n^{1/4}}\left(\left|\frac{\delta_{n+1}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}+1}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}-\frac{\delta_{n+1}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}\right|+\left|\frac{\delta_{n+1}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}-\frac{\delta_{n}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}\right|\right)\\
&= \frac{1}{n^{1/4}}\left(O\left(1\right)+O(n)\left|\delta_{n+1}^{(i)}-\delta_n^{(i)}\right|\right)\\
&= O\left(1\right)\left(\frac{1}{n^{1/4}}+n^{3/4}\left|\delta_{n+1}^{(i)}-\delta_n^{(i)}\right|\right)\\
&= O\left(1\right)\leq C_2,
\end{aligned}
$$

where the first equality holds from Lemma 3.4.2 and the last equality holds from Lemma 3.4.6. Then for all large enough n, we have

$$
\begin{aligned}
&\frac{3}{n^{1/4}}\left|\log\frac{\delta_{n+1}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}+1}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}-\log\frac{\delta_{n}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}\right|\\
&\qquad\leq \frac{3}{n^{1/4}}\left|\frac{\delta_{n+1}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}+1}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}-\frac{\delta_{n}^{(i)}}{\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}}\right|\\
&\qquad\leq 3C_2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\frac{2}{n^{1/4}}\left[\log\left(1+O\left(\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}+1}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}\right)\right)-\log\left(1+O\left(\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}\right)\right)\right]\right|\\
&+\frac{1}{n^{1/4}}\left|\log\delta_{n+1}^{(i)}-\log\delta_n^{(i)}\right|\\
&\leq \frac{2}{n^{1/4}}\left[\left|\log\left(1+O\left(\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}+1}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}\right)\right)\right|+\left|\log\left(1+O\left(\frac{\left(\lambda^{(i)}\right)^2}{N_n^{(i)}}+\frac{\left(\lambda^{(1)}\right)^2}{N_n^{(1)}}\right)\right)\right|\right]\\
&\quad+\frac{1}{n^{1/4}}\left|\log\delta_{n+1}^{(i)}-\log\delta_n^{(i)}\right|\\
&\leq C_2.
\end{aligned}
$$

We have now bounded all four terms in (3.38). Therefore, for all large enough n, we have

$$
r_{n+1}^{(i,j)}-r_n^{(i,j)}\leq\frac{5C_2/C_1}{n^{3/4}},
$$

and

$$r_{n+1}^{(i,j)} - 1 \leq r_n^{(i,j)} - 1 + \frac{5C_2/C_1}{n^{3/4}} \leq \frac{5C_2/C_1}{n^{3/4}}.$$

Thus, we have established a bound on the growth of $r_n^{(i,j)}$ that can occur as a result of sampling i at time n.

We now consider the growth of the ratio between stages n and $n+m$, where

$$m \triangleq \inf\left\{l > 0 : I_{n+l}^{(i)} = 1\right\}$$

as in the statement of Lemma 3.4.4. In words, $n+m$ is the index of the next time after n that we sample i. For any stage $n+s$ with $0 < s \leq m$, the inequality $r_{n+s+1}^{(i,j)} > r_{n+s}^{(i,j)}$ can only hold if alternative j or the optimal alternative is sampled at stage $n+s$.

If alternative j is sampled at stage $n+s$, then

$$\begin{aligned} r_{n+s+1}^{(i,j)} - r_{n+s}^{(i,j)} &= \frac{\log\left(v_{n+s+1}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)} - \frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s}^{(j)}\right)} \\ &= \frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)} - \frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s}^{(j)}\right)} \\ &\leq \left|\frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)}\right| \cdot \frac{\left|\log\left(v_{n+s+1}^{(j)}\right) - \log\left(v_{n+s}^{(j)}\right)\right|}{\left|\log\left(v_{n+s}^{(j)}\right)\right|}. \end{aligned}$$

Using similar arguments as above, we have

$$\frac{\left|\log\left(v_{n+s+1}^{(j)}\right) - \log\left(v_{n+s}^{(j)}\right)\right|}{\left|\log\left(v_{n+s}^{(j)}\right)\right|} = O\left((n+s)^{-3/4}\right) = O\left(n^{-3/4}\right),$$

and, by Lemma 3.4.2,

$$\left|\frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)}\right| = O(1).$$

Thus, there exists a constant C_3 such that

$$r_{n+s+1}^{(i,j)} - r_{n+s}^{(i,j)} \quad \leq \quad C_3 n^{-3/4}.$$

On the other hand, if alternative 1 is sampled at stage $n+s$, then

$$\begin{aligned}
r_{n+s+1}^{(i,j)} - r_{n+s}^{(i,j)} &= \frac{\log\left(v_{n+s+1}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)} - \frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s}^{(j)}\right)} \\
&\leq \left|\frac{\log\left(v_{n+s+1}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)} - \frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)}\right| + \left|\frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)} - \frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s}^{(j)}\right)}\right| \\
&= \left|\frac{\log\left(v_{n+s+1}^{(i)}\right) - \log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)}\right| + \left|\frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)}\right| \\
&\quad \cdot \frac{\left|\log\left(v_{n+s+1}^{(j)}\right) - \log\left(v_{n+s}^{(j)}\right)\right|}{\left|\log\left(v_{n+s}^{(j)}\right)\right|}.
\end{aligned}$$

Similarly as above, we have

$$\begin{aligned}
\left|\frac{\log\left(v_{n+s+1}^{(i)}\right) - \log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)}\right| &= O\left(n^{-3/4}\right), \\
\left|\frac{\log\left(v_{n+s}^{(i)}\right)}{\log\left(v_{n+s+1}^{(j)}\right)}\right| \frac{\left|\log\left(v_{n+s+1}^{(j)}\right) - \log\left(v_{n+s}^{(j)}\right)\right|}{\left|\log\left(v_{n+s}^{(j)}\right)\right|} &= O\left(n^{-3/4}\right).
\end{aligned}$$

Then, there exists a constant C_4 such that

$$r_{n+s+1}^{(i,j)} - r_{n+s}^{(i,j)} \quad \leq \quad C_4 n^{-3/4}.$$

Therefore, in all cases, for all large enough n, we have

$$r_{n+s+1}^{(i,j)} - r_{n+s}^{(i,j)} \leq \frac{C_5}{n^{3/4}},$$

where $C_5 = \max\{5C_2/C_1, C_3, C_4\}$. It follows that

$$r_{n+s+1}^{(i,j)} - 1 \quad \leq \quad r_{n+s}^{(i,j)} - 1 + \frac{C_5}{n^{3/4}}$$

$$\begin{aligned} &\leq \; r_n^{(i,j)} - 1 + \left(1 + k_s^{(j)} + k_s^{(1)}\right) \frac{C_5}{n^{3/4}} \\ &\leq \; \left(1 + k_s^{(j)} + k_s^{(1)}\right) \frac{C_5}{n^{3/4}}. \end{aligned}$$

However, from Lemma 3.4.4, we have $k_s^{(1)} \leq k_m^{(1)} = O\left(\sqrt{n \log\log n}\right)$ for all $0 < s \leq m$, and at the same time, from Lemma 3.4.3, we know that at most B_2 samples could be allocated to any suboptimal alternatives between two samples of alternative 1. Then we also have $k_s^{(j)} \leq k_m^{(j)} \leq B_2\left(k_m^{(1)} + 1\right)$, whence $k_m^{(j)} = O\left(\sqrt{n \log\log n}\right)$. It follows that

$$r_{n+s+1}^{(i,j)} - 1 \leq \left(1 + k_m^{(j)} + k_m^{(1)}\right) \frac{C_5}{n^{3/4}} = O\left(\frac{\sqrt{n \log\log n}}{n^{3/4}}\right),$$

whence $\limsup_{n\to\infty} r_n^{(i,j)} = 1$. By symmetry,

$$\liminf_{n\to\infty} r_n^{(i,j)} = \limsup_{n\to\infty} r_n^{(j,i)} = 1,$$

whence $\lim_{n\to\infty} r_n^{(i,j)} = 1$. This completes the proof.

3.5 Numerical Example

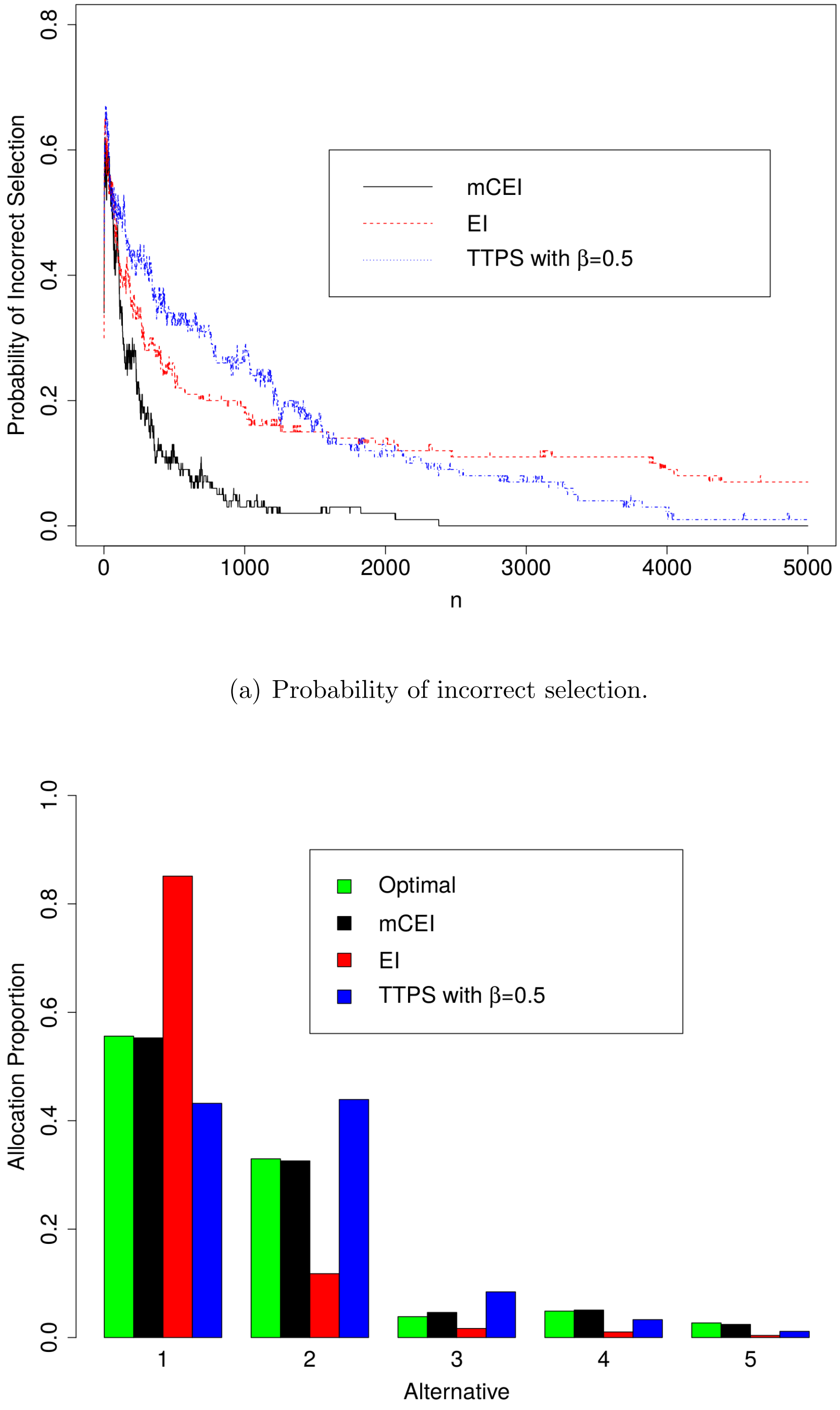
We present a numerical illustration of the mCEI method on a small synthetic problem. Two additional benchmarks were implemented. The first of these is the classic EI method from [10], given in (3.6). From (3.7)-(3.8), we do not expect this method to perform optimally in the long term; however, we include it because it is the fundamental procedure in the EI class of methods and thus a natural benchmark for mCEI. We also implemented the TTPS ("top-two probability sampling") method from [15]. This method assigns a fixed proportion β of the sampling budget to alternative j_n^* and allocates the rest based on a Thompson sampling-like criterion.

TTPS is an important benchmark since it can be made to achieve the optimal convergence rate if β is chosen correctly; however, since tuning β may be time-consuming in practice, [15] explicitly recommends setting $\beta = 0.5$ and derives a bound on the gap between the resulting convergence rate and the optimal one. We follow this recommendation in order to briefly comment on the tuning issue.

The synthetic example has five alternatives (systems) with true values $\mu = (0.5, 0.4, 0.3, 0.2, 0.1)$, standard deviations $\lambda = (1, 0.6, 0.6, 1, 1)$, the initial prior means $\theta_0 = 0$, and a budget of 5000 samples. Figure 2.1(a) shows the trajectory of the probability of incorrect selection, averaged over 100 macro-replications. Thus, the best alternative is $j^* = 1$, but the noise is greater for alternative 1 than for alternatives 2 and 3, which makes correct selection a bit more difficult.

By (3.7), we know that EI will not be able to achieve an exponential convergence rate, so it is unsurprising that it is eventually outperformed by TTPS; however, EI performs relatively well in the early stages. On the other hand, mCEI lags slightly behind EI during the first 200 replications, but subsequently discovers the best alternative very quickly. After 2500 samples, the empirical probability of incorrect selection is virtually zero under mCEI.

Figure 2.1(b) compares the allocations made by each method (also averaged over 100 macro-replications) to the optimal allocation, obtained by solving (3.1)-(3.2). As expected from (3.7), the EI allocation is far from optimal since it assigns most of the budget to the best alternative. The optimal proportion to assign to alternative 1 is slightly larger than 0.5; as a result, TTPS is not tuned optimally and thus consistently makes errors in all of the proportions. The allocation made



(a) Probability of incorrect selection.

(b) Simulation allocations after 5000 samples.

Figure 3.2: Comparison between mCEI and benchmark methods on the example problem.

by mCEI is very close to optimal.

Note that, even in this small problem, alternatives 3, 4 and 5 receive only about 10% of the budget under the optimal allocation. This suggests that, in some situations, the size of the problem may not necessarily determine its difficulty (aside from increasing the computational effort required to run a procedure), as many or even most of the alternatives may be similarly "irrelevant." Identifying characteristics that make problems more "difficult" may be an interesting subject for future work. At present, however, we only wish to illustrate the potential of mCEI to produce very close approximations of the optimal allocation, without any tuning, in a relatively small number of samples.

3.6 Conclusion

We have considered a ranking and selection problem with independent normal priors and samples, and shown that an EI-type method (a modified version of the CEI method of [16]) achieves the rate-optimality conditions of [9] asymptotically. This is the first such result available for any EI-type algorithm (previous rate results for other EI-type methods have shown that those methods achieve suboptimal allocations) that does not require any tuning.

This work strengthens the existing body of theoretical support for EI-type methods in general, and for the CEI method in particular. An interesting question is whether CEI would continue to perform optimally in, e.g., the more general Gaussian Markov framework of [16]. However, the current theoretical understanding of such models is quite limited, and more fundamental questions (for example, how correlated Bayesian models impact the rate of convergence) should be answered before any particular algorithm can be analyzed.

Chapter 4: Conclusion

In this thesis we focus on the interface between stochastic optimization and statistics. We apply statistical analysis to a suite of models that were established for stochastic optimization but only numerically studied in the literature, and show the theoretical validity of these models by showing the convergence of the algorithms. We also propose a new algorithm for solving the classic ranking & selection (R&S) problem, and show it is able to achieve the optimal budget allocation.

In Chapter 2, we propose a new general form of the stochastic approximation (SA) algorithm with "bias" terms included, and prove the convergence of this general algorithm. Then we apply this general framework to a suite of approximate Bayesian learning models including four univariate models and three multivariate models, all of which have proved their practical value for solving some realistic problem, and we show the convergence of each. On one hand, this work provides rigorous theoretical support for approximate Bayesian inference, as well as the inspiration for designing new approximate Bayesian models. For example, we propose a new approximate Bayesian model for learning through censored binary observations with unknown mean and variance and prove its consistency. On the other hand, it also gives us ideas about showing the convergence of other similar algorithms.

In Chapter 3, we propose a new algorithm based on the complete expected improvement (CEI) criterion for solving the R&S problem with finite alternatives under independent normality condition. We prove this algorithm recovers the optimal budget allocation asymptotically with respect to maximizing the probability of correct selection. This is the first EI-type algorithm that achieves the optimality condition, and it requires no extra computational effort or tuning work compared to the classic EI. This work bridges the gap between EI-type methods and the theoretical optimal budget allocation, and may inspire future work on designing algorithms that are able to recover the optimality condition in more general situations, for example, without the limitation of finite alternatives or without the normality assumption.

Bibliography

[1] M. Chhabra and S. Das. Learning the Demand Curve in Posted-Price Digital Goods Auctions. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 63–70, 2011.

[2] Sanmay Das and Malik Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 21, pages 361–368. 2009.

[3] W. B. Powell, A. George, H. Simão, W. R. Scott, A. Lamont, and J. Stewart. SMART: a stochastic multiscale model for the analysis of energy resources, technology, and policy. *INFORMS Journal on Computing*, 24(4):665–682, 2012.

[4] P. Dangauthier, R. Herbrich, T. P. Minka, and T. Graepel. TrueSkill Through Time: Revisiting the History of Chess. In J. C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, pages 337–344, 2007.

[5] H. Qu, I. O. Ryzhov, M. C. Fu, and Z. Ding. Sequential selection with unknown correlation structures. *Operations Research*, 63(4):931–948, 2015.

[6] T. S. Jaakkola and M. I. Jordan. Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10(1):25–37, 2000.

[7] H. Qu, I. O. Ryzhov, and M. C. Fu. Learning logistic demand curves in business-to-business pricing. In R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, editors, *Proceedings of the 2013 Winter Simulation Conference*, pages 29–40, 2013.

[8] H. J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications (2nd ed.)*. Springer, 2003.

[9] P. W. Glynn and S. Juneja. A large deviations perspective on ordinal optimization. In R. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 577–585, 2004.

[10] D.R. Jones, M. Schonlau, and W.J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

[11] W. B. Powell and I. O. Ryzhov. *Optimal Learning*. John Wiley and Sons, 2012.

[12] S. E. Chick, J. Branke, and C. Schmidt. Sequential Sampling to Myopically Maximize the Expected Value of Information. *INFORMS Journal on Computing*, 22(1):71–80, 2010.

[13] I. O. Ryzhov. On the convergence rates of expected improvement methods. *Operations Research*, 64(6):1515–1528, 2016.

[14] Y. Peng and M. C. Fu. Myopic allocation policy with asymptotically optimal sampling rate. *IEEE Transactions on Automatic Control*, 62(4):2041–2047, 2017.

[15] D. Russo. Simple Bayesian algorithms for best arm identification. *arXiv preprint arXiv:1602.08448*, 2017.

[16] P. Salemi, B. L. Nelson, and J. Staum. Discrete optimization via simulation using Gaussian Markov random fields. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, editors, *Proceedings of the 2014 Winter Simulation Conference*, pages 3809–3820. 2014.

[17] S. Y. Sohn and H. S. Kim. Random effects logistic regression model for default prediction of technology credit guarantee fund. *European Journal of Operational Research*, 183(1):472–478, 2007.

[18] C. K. Anderson and X. Xie. A choice-based dynamic programming approach for setting opaque prices. *Production and Operations Management*, 21(3):590–605, 2012.

[19] J. M. Keizers, J. W. M. Bertrand, and J. Wessels. Diagnosing order planning performance at a navy maintenance and repair organization, using logistic regression. *Production and Operations Management*, 12(4):445–463, 2003.

[20] D. J. Spiegelhalter and S. L. Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. *Networks*, 20(5):579–605, 1990.

[21] H. P. Simão, A. George, W. B. Powell, T. Gifford, J. Nienow, and J. Day. Approximate dynamic programming captures fleet operations for Schneider National. *Interfaces*, 40(5):342–352, 2010.

[22] W. B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality (2nd ed.).* John Wiley and Sons, New York, 2011.

[23] Z. Wen and B. Van Roy. Efficient exploration and value function generalization in deterministic systems. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26, pages 3021–3029, 2013.

[24] L. J. Hong and B. L. Nelson. A Brief Introduction To Optimization Via Simulation. In M.D. Rosetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference*, pages 75–85, 2009.

[25] M. Chau, M. C. Fu, H. Qu, and I. O. Ryzhov. Simulation optimization: a tutorial overview and recent developments in gradient-based methods. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, editors, *Proceedings of the 2014 Winter Simulation Conference*, pages 21–35, 2014.

[26] I. O. Ryzhov. Approximate Bayesian inference for simulation and optimization. In B. Defourny and T. Terlaky, editors, *Modeling and Optimization: Theory and Applications*, pages 1–28. Springer, 2015.

[27] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

[28] M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz. Approximate Bayesian computation. *PLoS Computational Biology*, 9(1):e1002803, 2013.

[29] V. Plagnol and S. Tavaré. Approximate Bayesian computation and MCMC. In H. Niederreiter, editor, *Monte Carlo and Quasi-Monte Carlo Methods*, pages 99–113. Springer, 2004.

[30] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[31] S. Asmussen and P. W. Glynn. A new proof of convergence of MCMC via the ergodic theorem. *Statistics & Probability Letters*, 81(10):1482–1485, 2011.

[32] D. T. Frazier, G. M. Martin, and C. P. Robert. On the consistency of approximate Bayesian computation. *arXiv preprint arXiv:1508.05178*, 2015.

[33] G. M. Martin, B. P. M. McCabe, D. T. Frazier, W. Maneesoonthorn, and C. P. Robert. Auxiliary likelihood-based approximate Bayesian computation in state space models. *arXiv preprint arXiv:1604.07949*, 2016.

[34] D. Russo and B. Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

[35] E. Gutin and V. Farias. Optimistic Gittins indices. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29, pages 3153–3161, 2016.

[36] A.B. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian data analysis (2nd ed.)*. CRC Press, 2004.

[37] T. P. Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Institute of Technology, 2001.

[38] Q. Zhang and Y. Song. Simulation selection for empirical model comparison. In L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, editors, *Proceedings of the 2015 Winter Simulation Conference*, pages 3777–3788, 2015.

[39] C. Wang and D. M. Blei. Variational inference in nonconjugate models. *Journal of Machine Learning Research*, 14:1005–1031, 2013.

[40] A. F. Garcia-Fernandez and L. Svensson. Gaussian MAP filtering using Kalman optimisation. *IEEE Transactions on Automatic Control (to appear)*, 2015.

[41] Y. Chen and I. O. Ryzhov. Approximate Bayesian inference as a form of stochastic approximation: a new consistency theory with applications. In T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, editors, *Proceedings of the 2016 Winter Simulation Conference*, pages 534–544, 2016.

[42] V. S. Borkar and S. P. Meyn. The ODE method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000.

[43] T. L. Lai. Stochastic approximation. *The Annals of Statistics*, 31(2):391–406, 2003.

[44] V. S. Borkar. *Stochastic approximation*. Cambridge University Press, 2008.

[45] M. Opper. A Bayesian approach to on-line learning. In D. Saad, editor, *On-line Learning in Neural Networks*, pages 363–378. 1998.

[46] I. O. Ryzhov and W. B. Powell. Information collection on a graph. *Operations Research*, 59(1):188–201, 2011.

[47] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303–1347, 2013.

[48] H. Qu, I. O. Ryzhov, M. C. Fu, E. Bergerson, and M. Kurka. Learning demand curves in B2B pricing: a new framework and case study. *Submitted for publication*, 2016.

[49] M. H. DeGroot. *Optimal Statistical Decisions*. John Wiley and Sons, 1970.

[50] Léon Bottou. Online learning and stochastic approximations. In D. Saad, editor, *On-line learning in neural networks*, pages 9–42. Cambridge, 1998.

[51] R. Pasupathy and S. Kim. The stochastic root-finding problem: Overview, solutions, and open questions. *ACM Transactions on Modeling and Computer Simulation*, 21(3):19:1–19:23, 2011.

[52] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Math. Stat.*, 22:400–407, 1951.

[53] F. Yousefian, A. Nedić, and U. V. Shanbhag. On stochastic gradient and subgradient methods with adaptive steplength sequences. *Automatica*, 48(1):56–67, 2012.

[54] H. Jiang and H. Xu. Stochastic approximation approaches to the stochastic variational inequality problem. *IEEE Transactions on Automatic Control*, 53(6):1462–1475, 2008.

[55] J. Koshal, A. Nedić, and U. V. Shanbhag. Regularized iterative stochastic approximation methods for stochastic variational inequality problems. *IEEE Transactions on Automatic Control*, 58(3):594–609, 2013.

[56] H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In T. L. Lai and D. Siegmund, editors, *Herbert Robbins Selected Papers*, pages 111–135. Springer, 1985.

[57] R. Herbrich, T. P. Minka, and T. Graepel. TrueSkill$^{\text{TM}}$: A Bayesian Skill Rating System. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 569–576, 2006.

[58] M. Chakraborty, S. Das, A. Lavoie, M. Magdon-Ismail, and Y. Naamad. Instructor rating markets. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 159–165. 2013.

[59] N.C. Petruzzi and M. Dada. Pricing and the newsvendor problem: A review with extensions. *Operations Research*, 47(2):183–194, 1999.

[60] D. Bertsimas and N. Kallus. From predictive to prescriptive analytics. *arXiv preprint arXiv:1402.5481*, 2015.

[61] Albert Nikolaevich Shiryaev. *Probability (2nd ed.)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1995.

[62] M. L. Puterman. *Markov Decision Processes.* John Wiley & Sons, New York, 1994.

[63] C. J .C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3):279–292, 1992.

[64] T. Jaakkola, M.I. Jordan, and S.P. Singh. Convergence of stochastic iterative dynamic programming algorithms. In J.D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems*, volume 6, pages 703–710, San Francisco, 1994. Morgan Kaufmann Publishers.

[65] John N. Tsitsiklis. Asynchronous stochastic approximation and Q-learning. *Machine Learning*, 16(3):185–202, 1994.

[66] J. Abounadi, D. P. Bertsekas, and V. S. Borkar. Stochastic approximation for nonexpansive maps: Application to Q-learning algorithms. *SIAM Journal on Control and Optimization*, 41(1):1–22, 2002.

[67] C. Szepesvári. The asymptotic convergence-rate of Q-learning. In M.I. Jordan, M.J. Kearns, and S.A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10, pages 1064–1070, Cambridge, MA, 1997. MIT Press.

[68] C. Ribeiro and C. Szepesvári. Q-learning combined with spreading: convergence and results. In *Proceedings of the ISRF-IEE International Conference on Intelligent and Cognitive Systems (Neural Networks Symposium)*, pages 32–36, 1996.

[69] I. O. Ryzhov, M. R. K. Mes, W. B. Powell, and G. A. van den Berg. Bayesian exploration for approximate dynamic programming. *Submitted for publication*, 2016.

[70] R. Dearden, N. Friedman, and S. Russell. Bayesian Q-learning. In *Proceedings of the 15th National Conference on Artificial Intelligence*, pages 761–768, 1998.

[71] A. D. Bull. Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12:2879–2904, 2011.

[72] Y. Engel, S. Mannor, and R. Meir. Bayes Meets Bellman: The Gaussian Process Approach to Temporal Difference Learning. In *Proceedings of the 20th International Conference on Machine Learning*, pages 154–161, 2003.

[73] Y. Engel, S. Mannor, and R. Meir. Reinforcement learning with Gaussian processes. In *Proceedings of the 22nd International Conference on Machine Learning*, pages 208–215, 2005.

[74] S.-H. Kim and B. L. Nelson. Selecting the best system. In S. G. Henderson and B. L. Nelson, editors, *Handbooks of Operations Research and Management Science, vol. 13: Simulation*, pages 501–534. North-Holland Publishing, Amsterdam, 2006.

[75] S. E. Chick. Subjective Probability and Bayesian Methodology. In S. G. Henderson and B. L. Nelson, editors, *Handbooks of Operations Research and Management Science, vol. 13: Simulation*, pages 225–258. North-Holland Publishing, Amsterdam, 2006.

[76] C.-H. Chen and L. H. Lee. *Stochastic simulation optimization: an optimal computing budget allocation.* World Scientific, 2010.

[77] C.-H. Chen, S. E. Chick, L. H. Lee, and N. A. Pujowidianto. Ranking and selection: efficient simulation budget allocation. In M. C. Fu, editor, *Handbook of Simulation Optimization*, pages 45–80. Springer, 2015.

[78] Q. Zhang and Y. Song. Moment-matching-based conjugacy approximation for Bayesian ranking and selection. *ACM Transactions on Modeling and Computer Simulation*, 27(4):26:1–26:23, 2017.

[79] A Gupta and D Nagar. *Matrix variate distributions.* Chapman & Hall/CRC, London, 2000.

[80] R.E. Bechhofer. A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics*, 25(1):16–39, 1954.

[81] Seong-Hee Kim and Barry L. Nelson. A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation*, 11(3):251–273, 2001.

[82] C.-H. Chen, J. Lin, E. Yücesan, and S. E. Chick. Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization. *Discrete Event Dynamic Systems*, 10(3):251–270, 2000.

[83] J. C. Gittins, K. D. Glazebrook, and R. Weber. *Multi-armed bandit allocation indices (2nd ed.).* John Wiley and Sons, 2011.

[84] R. Pasupathy, S. R. Hunter, N. A. Pujowidianto, L. H. Lee, and C.-H. Chen. Stochastically constrained ranking and selection via SCORE. *ACM Transactions on Modeling and Computer Simulation*, 25(1):1:1–1:26, 2014.

[85] J. Branke, S. E. Chick, and C. Schmidt. Selecting a selection procedure. *Management Science*, 53(12):1916–1932, 2007.

[86] W. R. Scott, W. B. Powell, and H. P. Simão. Calibrating simulation models using the knowledge gradient with continuous parameters. In B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, editors, *Proceedings of the 2010 Winter Simulation Conference*, pages 1099–1109, 2010.

[87] Bin Han, Ilya O. Ryzhov, and Boris Defourny. Optimal learning in linear regression with combinatorial feature selection. *INFORMS Journal on Computing*, 28(4):721–735, 2016.

[88] Chao Qin, Diego Klabjan, and Daniel Russo. Improving the expected improvement algorithm. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5381–5391. Curran Associates, Inc., 2017.

[89] S. R. Hunter and B. McClosky. Maximizing quantitative traits in the mating design problem via simulation-based Pareto estimation. *IIE Transactions*, 48(6):565–578, 2016.

[90] Y. Chen and I. O. Ryzhov. Rate-optimality of the complete expected improvement criterion. In W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, editors, *Proceedings of the 2017 Winter Simulation Conference*, pages 2173–2182, 2017.

[91] H. Ruben. A new asymptotic expansion for the normal probability integral and Mill's ratio. *Journal of the Royal Statistical Society*, B24(1):177–179, 1962.