A Bandit Approach to Indirect Inference

Felix Steinberger Eriksson and Erik Ildring

***Abstract*—We present a novel approach to the family of parameter estimation methods known as indirect inference (II), using results from bandit optimization, a sub-field of reinforcement learning concerned with stateless Markov decision processes (MDPs). First, we present the problem of indirect inference and show how it may be cast into the general framework of MDPs. We then discuss how this approach alleviates some limitations imposed by the randomness inherent to the optimization required for indirect inference. The bandit-based approach to indirect inference is implemented in code for two well-established bandit algorithms and subsequently validated experimentally. Our approach is demonstrated to work well in practice on the simple task of estimating the parameters of a Gaussian model, as well as on the non-trivial task of estimating the parameters of a mixture normal model. While the approach as a whole shows promise, the discretization required for analysis poses computational problems for some types of estimation tasks: Potential avenues to refining the method for such tasks are discussed.**

***Sammanfattning*—Vi presenterar ett nytt angreppssätt för den samling parameterskattningsmetoder som vanligtvis kallas indirekt inferens (II) som använder resultat från bandit-optimering, en underkategori av förstärkande inlärning som berör Markov-beslutsprocesser med triviala tillståndsrum. Först presenterar vi indirekt inferens generellt och visar hur problemet kan formuleras ekvivalent inom ramverket för Markov-beslutsprocesser. Därefter diskuterar vi hur vårt angreppssätt förbigår vissa begränsningar som påtvingas av den slumpkaraktär som är nödvändig att hantera i optimeringssteget av indirekt inferens. Den bandit-baserade metoden för indirekt inferens implementeras i programkod med två väletablerade algoritmer för bandit-optimering och valideras experimentellt. Vi åskådliggör att metoden fungerar väl i praktiken både på ett enkelt problem som skattning av parametrarna för en normalfördelning, och på ett icke-trivialt problem som skattning av parametrarna för en blandning av normalfördelningar. Även om angreppssättet som helhet är lovande observerar vi till följd av diskretiseringen som används problem med beräkningseffektiviteten för vissa typer av skattningsproblem: Vi diskuterar hur metoden kan sofistikeras för att hantera dessa effektivt.**

***Index Terms*—Parameter estimation, indirect inference, multi-armed bandits.**

Supervisor: Braghadeesh Lakshminarayanan

TRITA number: TRITA-EECS-EX-2023:140

I. Introduction

When tasked with the problem of understanding or controlling a physical system one may utilize mathematical models. To formulate these models, data are commonly gathered from the physical system. These data can then be used in tandem with methods from a variety of fields, such as statistics, machine learning and system identification, to build models of the system. One predominant group of models is parametric models [1]. A parametric model is a collection of probability distributions, each associated with some parameter θ belonging to a predetermined parameter set Θ. A standing assumption is that the physical system generating data is well-described by one of these probability distributions with some true parameter θ^*, and a common desire is to estimate the true parameter using the gathered data from the physical system.

The procedure of estimating the unknown true parameter of a probability distribution from data is well-studied and a variety of methods have been derived. Two examples of such methods are method of moments estimation and maximum likelihood estimation [2]. Both of these parameter estimation methods require the knowledge of the likelihood function, which in some cases may be intractable.

One family of parameter estimation methods that tries to tackle the problem of intractable likelihood functions is indirect inference [3]. Indirect inference takes advantage of the fact that given a fixed value of the model parameters, data can be sampled from the model. It then uses a set of tractable auxiliary models that may not fully explain the physical system to match the simulated data sequence with the gathered real-world data sequence by means of a convex optimization objective function. The user's choice of auxiliary models affects the estimate and one hopes to choose an auxiliary model to balance a good enough approximation of the physical system dynamics and computational tractability. However, there is one issue with how the objective function is commonly defined. As simulated data is sampled from a probability distribution, given a fixed value of the parameters, two sample sequences may (and most likely will) not coincide. The objective function, while constructed to capture summary properties of the system, may hence vary between evaluations and is not strictly a function of the parameter. This can pose a challenge in implementations when choosing optimization solver because many solvers rely on first or second order derivative information; With a function that is not deterministic this may be difficult or impossible.

To handle the problem of stochastic objective functions in indirect inference, the method can be cast into a stochastic bandit problem. One may then use existing theory and algorithms from this field, adapted to a optimization of a stochastic objective. For instance, bandit optimization does not need first or second order derivative information.

The main contributions of this paper are:

- We provide to our knowledge the first formulation of the general method of indirect inference as a stochastic bandit problem.
- We demonstrate on two example problems that bandit optimization provides indirect inference estimators that work well in practice.

The rest of this report is organized as follows. In Section II, preliminary theory is presented. Section III describes the main issue with indirect inference as typically presented and formulates the problem. In Section IV our proposed approach of indirect inference as a stochastic bandit is presented, and in Section V this approach is experimentally validated on two example problems. Section VI discusses our approach and the results of the experimental validation. Finally, in Section VII conclusions are drawn.

II. Preliminaries

A. Parameter Estimation

A family of parametric models is a set $\mathcal{M} = \{M(\theta) \mid \theta \in \Theta\}$ of probability distributions (also referred to as models) $M(\theta)$ defined on a common sample space, indexed by some parameter θ belonging to a set $\Theta \subseteq \mathbb{R}^n$ for some positive integer n, commonly called the parameter space [1]. The standing assumption is that one of the probability distributions in the model—say, the distribution $M(\theta^*)$ corresponding to the parameter $\theta^* \in \Theta$—corresponds to the physical process at hand. The parametric model $M(\theta^*)$ describes the data-generating physical process in the sense that data are assumed to be sampled independently and identically according to the probability distribution $M(\theta^*)$. Parametric models provide a convenient way to encode assumptions and knowledge about the qualitative behaviour of a data-generating process into a mathematical description of it while allowing for variability in the precise description.

Given a family of parametric models $\mathcal{M}$ of some physical process and some sequence $D_N = \big((x_i, y_i)\big)_{i=1}^N$ of data generated by the process, a common task is to determine the true distribution according to which the data were generated, in order to make qualitative predictions about future data, or to describe or draw conclusions about the mechanism underlying the data generation. As each distribution in the model is indexed by a parameter in Θ, one commonly refers to this as the problem of parameter estimation, i.e. determining the parameter θ^* of the true probability distribution $M(\theta^*)$. A systematic method for determining an estimate θ_N of the true parameter θ^* given the data D_N is called a parameter estimation method. Parameter estimation methods are commonly presented in terms of mappings $D_N \mapsto \theta_N \in \Theta$; In a more general setting, a parameter estimation method may not necessarily be a mapping but rather a procedure that explains how to calculate the estimate in terms of the data.

B. Indirect Inference

One family of parameter estimation methods introduced by Smith [3] and expanded on by Gourieaux *et al.* [4, 5] is indirect inference (II). Indirect inference is useful when the likelihood function or other conventional estimation starting points are intractable, as is the case in many modern econometric models [6] or when data may be missing [7, 8].

Consider as previously a sequence of data $D_N = \big((x_i, y_i)\big)_{i=1}^N$ for $N \in \mathbb{N}$ generated by a physical process assumed to be accurately described by a parametric model $\mathcal{M}$. Suppose further that $\mathcal{M}$ is intractable in the sense that a meaningful mapping $D_N \mapsto \theta_N$ is difficult to describe analytically or computationally expensive to calculate. Indirect inference introduces an auxiliary model $\mathcal{M}_a = \{M_a(\beta) \mid \beta \in \mathcal{B} \subseteq \mathbb{R}^k\}$ which is computationally simpler, but does not in general capture the full dynamics of the physical process. One now defines a parameter estimation method for $\mathcal{M}$ in the following steps (described graphically in Fig. 1).

1) Fitting an auxiliary model to the true data: Select, by some method, the auxiliary parameter $\beta^* \in \mathcal{B}$ such that the resulting auxiliary model $M_a(\beta^*)$ describes the data $D_N = \big((x_i, y_i)\big)_{i=1}^N$ best (by some measure of goodness) among all models in $\mathcal{M}_a$.

2) Generation of synthetic data: Select some parameter $\theta \in \Theta$ and generate a sequence $\tilde{D}_N = \big((\tilde{x}_i, \tilde{y}_i)\big)_{i=1}^N$ of N synthetic data points by sampling independently from $M(\theta)$.

3) Fitting an auxiliary model to the synthetic data: Select, by the same method as in *1)*, the parameter $\beta(\theta) \in \mathcal{B}$ such that the resulting auxiliary model $M_a(\beta(\theta))$ describes the synthetic data $\tilde{D}_N$ best among all models in $\mathcal{M}_a$ by the same measure of goodness as in *1)*. We emphasize the dependence of this parameter on θ through the synthetic data generated according to the distribution $M(\theta)$.

4) Minimizing a score function: Select some positive definite matrix W and define the score (or loss) function $J : \Theta \to \mathbb{R}$ by

$$J(\theta) = (\beta(\theta) - \beta^*)^T W (\beta(\theta) - \beta^*). \tag{1}$$

In practice, one often takes $W = I$ to simplify analysis. Determine some $\theta_N \in \arg\min_{\theta \in \Theta} J(\theta)$. This θ_N is the produced estimate of the true parameter θ^*. The positive-definiteness of W ensures a unique minimizing $\beta(\theta)$, but several values of θ may correspond to that value of $\beta(\theta)$ as $\mathcal{B}$ is typically a lower-dimensional space than Θ.

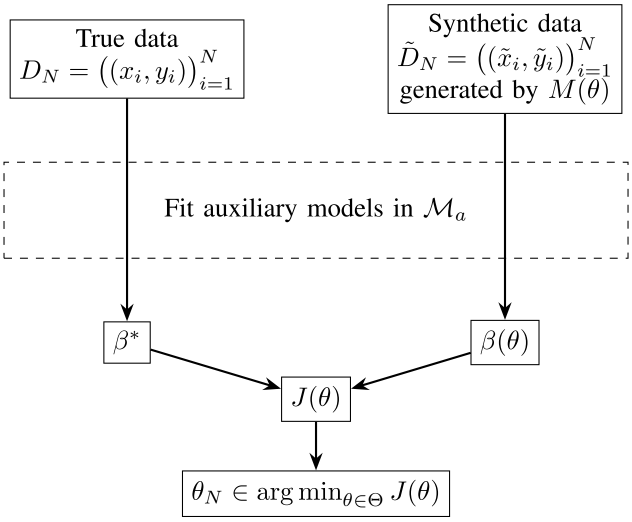
True data
$D_N = \big((x_i, y_i)\big)_{i=1}^N$
Synthetic data
$\tilde{D}_N = \big((\tilde{x}_i, \tilde{y}_i)\big)_{i=1}^N$
generated by $M(\theta)$
Fit auxiliary models in $\mathcal{M}_a$
β^*
$\beta(\theta)$
$J(\theta)$
$\theta_N \in \arg\min_{\theta \in \Theta} J(\theta)$

Fig. 1. Indirect inference, adapted from [5]. On selection of some $\theta \in \Theta$ synthetic data $\tilde{D}_N$ are generated according to the corresponding model and auxiliary models are fitted to the true data D_N and the synthetic data $\tilde{D}_N$. The goodness of fit in terms of θ is evaluated using the loss function $J(\theta)$.

C. Optimizers

Several of the steps of the calculation of an indirect inference estimate involve optimization. The minimization of the score function J is an obvious example, but also the fitting of auxiliary models to data is usually achieved through the optimization of some goodness of fit criterion. To talk about the specifics of these steps more systematically, we introduce some notation in the form of the general optimizers $\mathcal{F}_a$ and $\mathcal{O}$. The auxiliary optimizer $\mathcal{F}_a$ takes a data sequence D_N and an auxiliary model $\mathcal{M}_a$ and produces, by some procedure, a parameter $\mathcal{F}_a(D_N, \mathcal{M}_a) \in \mathcal{B}$ corresponding to some model of best fit belonging to $\mathcal{M}_a$. The optimizer $\mathcal{O}$ takes some loss function J, a parameter space Θ and a hyperparameter set $\mathcal{H}$ specific to the optimizer used, and produces, by some procedure, a parameter $\mathcal{O}(J, \Theta, \mathcal{H}) \in \Theta$ minimizing $J(\theta)$. The specific measures of goodness of fit as well as procedures used to determine $\mathcal{F}_a(D_N, \mathcal{M}_a)$ and $\mathcal{O}(J, \Theta, \mathcal{H})$ are encoded in $\mathcal{F}_a$ and $\mathcal{O}$ (and $\mathcal{H}$), respectively: While of major importance in implementations, their specifics are not pertinent to a discussion of indirect inference in general.

D. Multi-armed bandits

Multi-armed bandits [9] is a family of Markov Decision Processes (MDPs) [10]. In a multi-armed bandit problem the state space is trivial in the sense that there is no state (or one state, and all actions result in deterministic transitions to the same state). The actions are numbered from 1 to K for some $K \in \mathbb{N}$ and are called *arms*. Each arm has a unknown reward distribution from which a reward is sampled upon choosing said arm. Formally, a *stochastic bandit*, a type of multi-armed bandit, is the MDP $\big(T, \mathcal{S}, \mathcal{A}, p_t(\cdot \mid s, a), r_t(s, a)\big)$ where the state space $\mathcal{S} = \emptyset$ is trivial, the action space $\mathcal{A} = \{1, \dots, K\}$, $p_t(\emptyset, a) = 1$ for all arms $a \in \mathcal{A}$. Lastly, let us denote the reward distribution of arm a with $\mathcal{D}_a$ so that $r_t(\emptyset, a) \sim \mathcal{D}_a$ (that is, $r_t(\emptyset, a)$ is a random sample from the distribution $\mathcal{D}_a$). The goal is to find the arm with greatest expected value. Assume that an algorithm A, known in this context as an *agent*, is to interact with a stochastic bandit problem. Then the following scheme describes the interaction.

Algorithm 1 Stochastic bandit [9]

```
Parameters: Time horizon $T$, $K$ arms, unknown reward distributions $\mathcal{D}_a$ for all arms.
for $t = 1$ to $T$ do
    Arm $a_t$ is chosen by agent $A$ according to some rule
    The reward $r_t$ is sampled from distribution $\mathcal{D}_{a_t}$
    The reward $r_t$ is given to the agent $A$
end for
```

The concept of *regret* will now be defined, which is a commonly used metric of performance for algorithms trying to solve bandit problems [11]. Let the expected return of arm a be denoted by $\mu(a) := \mathbb{E}[\mathcal{D}_a]$, and let the best expected return of any of the K arms be denoted by $\mu^* = \max_{a \in \mathcal{A}} \mathbb{E}[\mathcal{D}_a]$. For simplicity we assume that the expected rewards are stationary in time. Now suppose that for each time step $t = 1, \dots T$ an arm a_t is picked by some bandit algorithm A. The regret of A at time T (or simply the regret) is then defined as

$$R^A(T) = T\mu^* - \sum_{t=1}^{T} \mu(a_t). \tag{2}$$

The regret can be seen as the *cost of learning* since $T\mu^*$ is the best expected sum of rewards that can be collected up to time T (achieved by an oracle agent which knows the best arm ahead of time) and $\sum_{t=1}^{T} \mu(a_t)$ is the sum of average rewards collected up to time T by the algorithm.

In many scenarios one does not have access to the average rewards $\mu(a)$ for the arms $a \in \mathcal{A}$ and requires a method for estimating the regret. Assume that there is an oracle algorithm that knows which arm is best at each time step t, denote this arm a_t^*. Let r_t^O denote the reward gathered by the oracle at time t. Also for a bandit algorithm A, denote the reward acquired by A at time t by r_t^A. Using this we have that $\mathbb{E}\big[r_t^A\big] = \mathbb{E}\big[\mathcal{D}_{a_t}\big] = \mu(a_t)$ and $\mathbb{E}\big[r_t^O\big] = \mathbb{E}\big[\mathcal{D}_{a_t^*}\big] = \mu(a_t^*) = \mu^*$. It is therefore easy to see that the empirical regret

$$\tilde{R}^A(T) = \sum_{t=1}^{T} (r_t^O - r_t^A) \tag{3}$$

is an unbiased estimator of the regret $R^A(T)$ in eq. 2.

E. Bandit algorithms

We introduce two algorithms that seek to solve stochastic bandit problems.

One algorithm that provides a systematic solution to a stochastic bandit problem is the ϵ*-greedy* algorithm [11]. Let $n_a(t)$ be the number of times arm a has been selected up to and including time t. The empirical average reward of arm a up to time t can thus be defined as

$$\hat{\mu}_a(t) = \frac{1}{n_a(t)} \sum_{i=1}^{t} \chi_{\{a_i = a\}} r_t \tag{4}$$

where $\chi_{\{a_i = a\}}$ is the indicator function for the event $\{a_i = a\}$. The algorithm then decides which action a_t to select at time t by uniformly at random sampling an arm from all available arms $\{1, 2, \dots, K\}$ with probability (w.p.) ϵ, and selecting the arm with best empirical average reward up to time $t - 1$ w.p. $1 - \epsilon$.

Another algorithm for solving a stochastic bandit problem is the UCB1 algorithm [12]. This algorithm is based on finding an approximate upper confidence bound for the empirical average reward for each arm and at each time step and playing the arm with the largest upper bound value. This upper bound value (also referred to as index) of arm a at time is defined to be $b_a(t) = \hat{\mu}_a(t-1) + \sqrt{\frac{2\log(t)}{n_a(t-1)}}$.

Pseudo-code for the ϵ-greedy and UCB1 algorithms can be found in Appendix A.

III. Problem statement

In its original conception, indirect inference was used to estimate parameters of time series in econometric models. For instance, the seminal paper of Smith [3] demonstrates indirect inference on a time series model of business cycles considering consumption, investment, capital stock and production, which, while non-linear, has dynamics that are well-approximated by the linear dynamics provided by e.g. a vector autoregressive (VAR) model [13]. There is a standing and tacit assumption that, given synthetic data $\tilde{D}_N$ generated according to some model $M(\theta)$, the auxiliary parameter $\beta(\theta)$ is more or less independent of the actual realized synthetic data sequence and treated as a deterministic function of θ for optimization of the loss function $J(\theta)$. In practice, this works well, as the realized $\beta(\theta)$ are often unbiased estimates of the true quantity one purports to use in the optimization. Using a VAR model as an auxiliary model, for instance, the realized $\beta(\theta)$ are unbiased estimates of the true coefficients of best fit of the true data generation model. This still allows for some convergence guarantees to the true optimum by way of stochastic approximation [14, 15].

However, in general, the $\beta(\theta)$ calculated from some realized synthetic data sequence $\tilde{D}_N$ is not even an unbiased estimate of some interesting, model-dependent parameter that may be used for optimization. This may yield issues when $\beta(\theta)$, and in particular the resulting loss function $J(\theta)$, are treated as deterministic functions of θ for the sake of optimization. An option that preserves the purported performance guarantees of indirect inference is to satisfy oneself with utilizing auxiliary models that do in fact behave nicely in combination with optimization and exhibit unbiasedness, but this would limit the general applicability of indirect inference.

The main problem of this paper is to **find an approach to solve the optimization problem**

$$(P) \quad \arg\min_{\theta \in \Theta} J(\theta)$$

where the objective J is treated as a stochastic function.

IV. Proposed approach: Bandit optimization for Indirect Inference

Our main contribution is to cast the problem of indirect inference into the framework of bandit optimization. This is, to our knowledge, the first approach to the specific problem of indirect inference that deals explicitly with the randomness inherent in the loss function θ to solve (P).

A. Stochastic loss

Algorithm 2 describes a general procedure for how to calculate the loss J for some set of models $\mathcal{M}$, some auxiliary models $\mathcal{M}_a$ and some selected parameter $\theta \in \Theta$, using an auxiliary optimizer $\mathcal{F}_a$. In particular, it encapsulates the generation of synthetic data and subsequent fitting of an auxiliary model into a subroutine which we shall refer to later for the sake of convenience.

Algorithm 2 Calculating loss J

```
Parameters: Parameter $\theta \in \Theta$, set of models $\mathcal{M}$, set of
    auxiliary models $\mathcal{M}_a$, auxiliary fit optimizer $\mathcal{F}_a$, data $D_N$
Output: Loss J
Fit $\beta^* \leftarrow \mathcal{F}_a(D_N, \mathcal{M}_a)$
Draw simulated data $\tilde{D}_N \sim M(\theta) \in \mathcal{M}$
Fit $\beta(\theta) \leftarrow \mathcal{F}_a(\tilde{D}_N, \mathcal{M}_a)$
$J \leftarrow (\beta(\theta) - \beta^*)^T (\beta(\theta) - \beta^*)$
Return $J$
```

B. Indirect inference as a stochastic bandit

A generic implementation of indirect inference using algorithm 2 to compute the loss J may look like algorithm 3.

Algorithm 3 Indirect inference

```
Parameters: Set of models $\mathcal{M}$ w. parameter set $\Theta$, set of
    auxiliary models $\mathcal{M}_a$ w. aux. parameter set $\mathcal{B}$, optimizer
    $\mathcal{O}$, auxiliary fit optimizer $\mathcal{F}_a$, data $D_N$
$\tilde{J} \leftarrow J(\cdot, \mathcal{M}, \mathcal{M}_a, \mathcal{F}_a, D_N)$
According to optimizer $\mathcal{O}$: $\theta^* \leftarrow \arg\min_{\theta \in \Theta} \tilde{J}(\theta)$
Return $\theta^*$
```

To tackle the problem of stochastic score function we formulate the general problem of indirect inference as a stochastic bandit problem. Remember the assumption that $\Theta = \{\theta^1, \theta^2, \ldots, \theta^K\}$ and that a general agent A chooses arms in the index set $\{1, 2, \ldots, K\}$.

Algorithm 4 Indirect inference as a stochastic bandit

```
Parameters: Set of models $\mathcal{M}$ w. parameter set $\Theta$, set of
    auxiliary models $\mathcal{M}_a$ w. aux. parameter set $\mathcal{B}$, agent $A$,
    auxiliary fit optimizer $\mathcal{F}_a$, data $D_N$, Time horizon $T$
$\tilde{J} \leftarrow -J(\cdot, \mathcal{M}, \mathcal{M}_a, \mathcal{F}_a, D_N)$
for $t = 1, \ldots T$ do
    Arm chosen $a_t$ according to $A$'s decision rule
    Sample reward $r_t \sim \tilde{J}(\theta_{a_t})$
    Agent A observes $r_t$ and updates decision rule according to reward history $(r_n)_{n=1}^t$ and action history $(a_n)_{n=1}^t$.
end for
$a^* \leftarrow$ Best arm determined by A
Return $\theta_{a^*}$
```

Notice that $\tilde{J}$ is set to $-J(\cdot, \mathcal{M}, \mathcal{M}_a, \mathcal{F}_a, D_N)$. This is due to the convention that the objective in a stochastic bandit problem is to maximize the accumulated reward. Maximizing the negative of a function is equivalent to minimizing the function.

V. Experimental validation

We evaluate the performance of indirect inference with bandit optimization on two sample problems. The selected problems demonstrate different aspects of the estimation method: Where the first problem shows the method producing reasonable estimates on a simple, well-understood estimation task, the second indicates the generalizability of the method to a non-trivial estimation task.

A. Gaussian model

The performances of indirect inference with ϵ-greedy and UCB1 optimizers (see Appendix A) are compared on the simple task of parameter estimation for a Gaussian model. Consider a model $\mathcal{M} = \{M(\theta) \mid \theta \in \Theta\}$ of Gaussian distributions $M(\theta) = \mathcal{N}(\mu, \sigma^2)$ for some $\theta = (\theta_1, \theta_2)^T = (\mu, \sigma)^T \in \Theta$. The parameter space $\Theta = \{\theta^1, \theta^2, \dots, \theta^K\} \subset \mathbb{R}^2$ is a finite set of two-dimensional real vectors $\theta^i = (\theta_1^i, \theta_2^i)^T$ conceptually corresponding to a discretization of the product of some intervals encoding prior belief or knowledge about the mean μ and standard deviation σ of the true distribution $\mathcal{N}(\mu, \sigma^2) \in \mathcal{M}$ according to which a sequence of N data $D_N = (y_i)_{i=1}^N$ is generated as independent samples from identical distributions. D_N is considered fixed true data: In a real application it would come from some physical data generation process. The set of auxiliary models $\mathcal{M}_a$ is taken to be Gaussian also, with β^* and $\beta(\theta)$ being two-dimensional vectors of sample mean and sample standard deviation of D_N and $\tilde{D}_N$, respectively.

Concretely, $\Theta = \Theta_\mu \times \Theta_\sigma$ where $\Theta_\mu = \left\{\frac{15k}{30}\right\}_{k=0}^{30}$ and $\Theta_\sigma = \left\{0.5 + \frac{\sqrt{15}k}{30}\right\}_{k=0}^{30}$ are sets of equidistantly spaced grids of 31 points covering the intervals $[0, 15]$ and $[0.5, \sqrt{15}]$, respectively.

For different values of the true parameter θ^* ranging over the entirety of Θ we plot the estimated mean against the true mean and the estimated standard deviation against the true standard deviation for both ϵ-greedy (Fig. 2) and UCB1 (Fig. 3) optimizers. The identity line is also plotted for each parameter: When the estimate coincides with the true parameter, the blue dot lies on this red dashed line. In all cases, the closer the blue dot markers lie to the red dashed line, the better the estimate is.

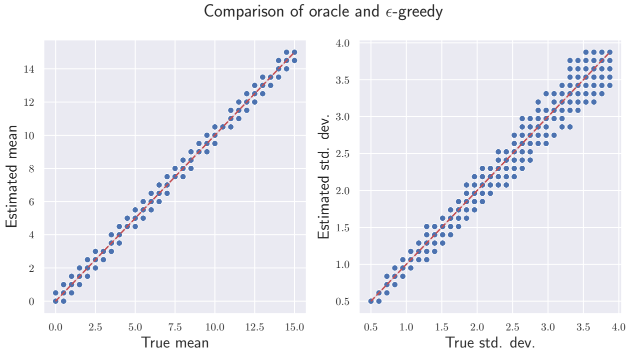

Fig. 2. Indirect inference estimates of means and standard deviations of a normal distribution using an ϵ-greedy optimizer with time horizon $T = 20000$. Left: Estimated means for a given true mean (blue dots) and the identity line corresponding to estimates of an oracle estimator (red dashed line). Right: Estimated means for a given true mean (blue dots) and the identity line corresponding to estimates of an oracle estimator (red dashed line).

For one value of the true parameter ($\theta^* = (3.10, 2.82) \in \Theta$, chosen arbitrarily somewhere in the central parts of Θ), the empirical regret of ϵ-greedy and UCB optimizers is computed and plotted over time in Fig. 4.

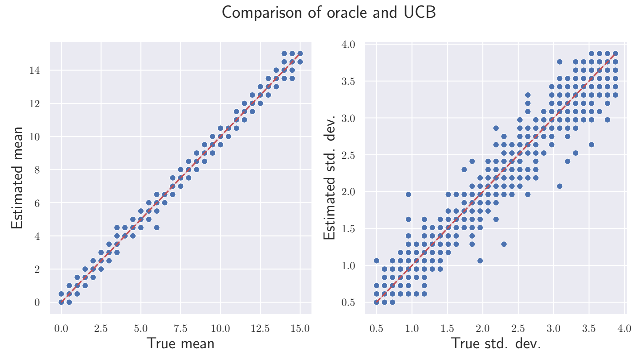

Fig. 3. Indirect inference estimates of means and standard deviations of a normal distribution using a UCB1 optimizer with time horizon $T = 20000$. Left: Estimated means for a given true mean (blue dots) and the identity line corresponding to estimates of an oracle estimator (red dashed line). Right: Estimated means for a given true mean (blue dots) and the identity line corresponding to estimates of an oracle estimator (red dashed line).

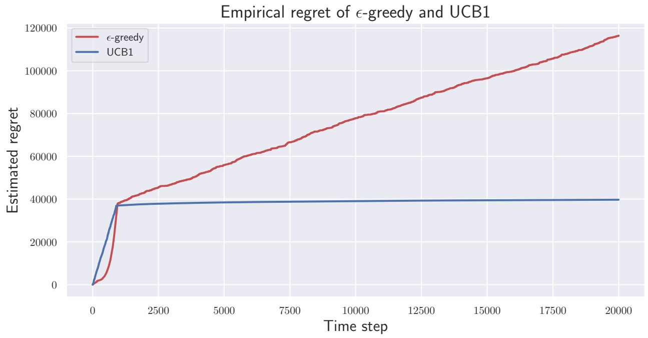

Fig. 4. Empirical regrets of ϵ-greedy (red) and UCB1 (blue) optimizers plotted over time. The true parameter is $\theta^* = (3.10, 2.82)$.

B. Mixture normal model

In applications it is common to encounter physical processes well-modeled by mixture distributions, in particular when the population may be partitioned into non-overlapping subpopulations, each of which follow the same distribution but with different parameters. Examples in the literature abound, including models of human height [16], investigations into sediment arrangements in riverbeds [17, 18], analysis of astronomical data [19], and fisheries research [20]. We consider the arguably simplest instance of such models, namely the model of mixtures of two normal distributions, that is, the model $\mathcal{M} = \{M(\theta) \mid \theta \in \Theta\}$ where $\theta \in \Theta \subseteq \mathbb{R}^5$ is a five-dimensional vector $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, p)^T$ of means μ_1, μ_2 and standard deviations σ_1, σ_2 of two independent normal distributions, and $p \in [0, 1]$ is a mixing parameter. Each indexed probability distribution $M(\theta)$ is the mixture

$$M(\theta) = p\mathcal{N}(\mu_1, \sigma_1^2) + (1 - p)\mathcal{N}(\mu_2, \sigma_2^2)$$

of two independent normal distributions.

Even a reduced parameter estimation task of estimating some components of the true parameter $\theta^* = (\mu_1^*, \sigma_1^*, \mu_2^*, \sigma_2^*, p^*)^T \in \Theta$ while keeping others fixed is difficult to solve efficiently even for this simple mixture model [8]. We consider the reduced parameter estimation task for $\mathcal{M}$ with $\Theta = \Theta_\mu \times \Theta_\sigma \times \{\theta_3^*\} \times \{\theta_4^*\} \times \Theta_p$, where $\Theta_\mu = \left\{3+\frac{9k}{36}\right\}_{k=0}^{36}$, $\Theta_\sigma = \left\{0.5+\frac{2.4k}{12}\right\}_{k=0}^{12}$ and $\Theta_p = \left\{\frac{k}{10}\right\}_{k=0}^{10}$. $\Theta_\mu, \Theta_\sigma$ and Θ_p are equidistant discretizations of the intervals $[3, 12]$, $[0.5, 2.9]$ and $[0, 1]$ using 73, 25 and 21 points, respectively. The singleton sets in the third and fourth factors of Θ correspond to the assumption that the third and fourth components of the true parameter θ^* are known. We take $\theta^* = (9, 1.7, 3.25, 0.9, 0.7)^T \in \Theta$, again chosen somewhat arbitrarily. The set of auxiliary models $\mathcal{M}_a$ is left implicit but chosen in such a way that β^* and $\beta(\theta)$ are the decile vectors of D_N and $\tilde{D}_N$, respectively.

The parameter vector $(9, 1.7, 3.25, 0.9, 0.7)^T$ is correctly identified by both ϵ-greedy-based II estimator and the UCB-based II estimator. In principle, there is no obstacle beyond the curse of dimensionality (discussed further below) against loosening the assumption that some components of the vector are known. Fig. 5 and Fig. 6 show normalized histograms of the true data set and one generated according to the correctly identified parameter estimate.

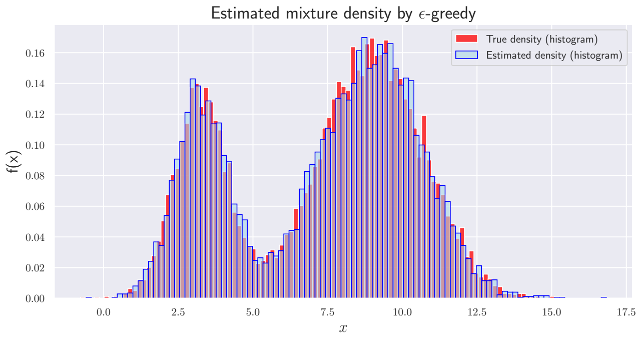

Fig. 5. Normalized histograms of the true model (corresponding to the parameter θ^*) and the estimated model (corresponding to the estimate yielded by ϵ-greedy-based II), both with time horizon $T = 10000$.

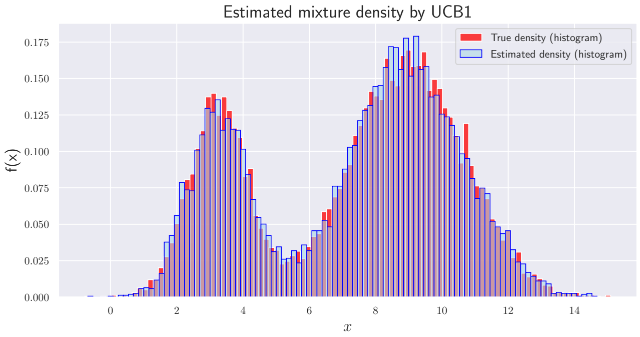

Fig. 6. Normalized histograms of the true model (corresponding to the parameter θ^*) and the estimated model (corresponding to the estimate yielded by UCB-based II), both with time horizon $T = 10000$.

VI. Discussion

A. Gaussian model

Based on the plots presented in Fig. 2 and Fig. 3, both the ϵ-greedy algorithm and the UCB1 algorithm seem to accurately estimate the parameters of a Gaussian distribution. The scatter plots of the estimated means displayed in the left panel of both figures suggest that both algorithms effectively estimate the mean, as all markers lie close to the identity line. One notices that the estimates of the standard deviation yielded by the ϵ-greedy algorithm vary more around the true standard deviation for higher values of the true standard deviation. The standard deviation estimates yielded by UCB1 are greater across the board, but do not show this increase for greater true values. This is slightly unexpected, but likely to disappear for simulations on a greater time horizion.

Turning to Fig. 4 we can see that the estimated regret increase rapidly in the early time steps for both algorithms. This is to be expected since neither algorithm has had any time to explore the different arms. For the ϵ-greedy algorithm, after the initial fast increase, the regret trends upwards linearly. The UCB1 algorithm, on the other hand, displays a slight concavity in the empirical regret curve after the initial phase, which is in line with theoretical results regarding UCB.

This toy problem showcases one limitation of the discretization approach in that it assumes that the true parameter belongs to the discrete parameter space Θ. When applying such a discretization approach to real data, it is improbable that the true parameter belongs to any discretization of the continuous real parameter space. As in any discretization method, a fine enough discretization is required, and what is fine enough needs to be determined on a case-by-case basis. To avoid the estimation becoming prohibitively computationally expensive, adaptive grid schemes may be used, wherein the discretization is successively refined around consecutively more precise estimates. Another approach is to forego the discretization completely, instead turning to the existing theory of continuous bandits. This is beyond the scope of this thesis, but is discussed in the section on *Further research* below.

B. Mixture normal model

The experimental evaluation on a mixture normal model displays the generalizability of the bandit approach to indirect inference. Unlike for other common distributions such as the regular normal distribution, commonly used sample statistics like the sample mean or the sample variance are not unbiased estimators of any quantity of interest. Approaches based on maximum likelihood estimation of the means and variances of the component distributions under separate maximization over the mixing parameter p might show promise for the simple case of a mixture of two distributions, but become intractable quickly for mixtures of a greater number of component distributions.

Taking the auxiliary parameters in $\mathcal{B}$ to be deciles (or, more generally, k-quantiles for some sufficiently large $k \in \mathbb{N}$) of the true and synthetic data sequences shows great potential due to the combination of ease of use and general applicability. Intuitively, it is desirable that the auxiliary parameters in $\mathcal{B}$ computed from the different data sequences are similar when the models generating the data are similar. Models indexed by similar parameters in Θ do in fact generate similar data sequences, so it is plausible that they yield similar quantiles. Furthermore, it is easy to refine this approach if need be (i.e. if the used quantiles do not provide a sufficiently truthful summary of the true data sequences) by increasing the number of quantiles used, or by increasing the number of samples used in the construction of the quantile vectors to reduce variance.

C. Further research

Our formulation of II as a stochastic bandit problem assumes a finite parameter space for ease of an initial analysis. While the general approach of bandit optimization to II remains sound, the discretization may become prohibitively expensive in order to approximate continuous true parameter spaces with very high numerical accuracy, or if the parameter space is of very high dimension. This is a natural consequence of the fact that the complexity of bandit algorithms for a finite number of arms generally scales with the number of arms [11], and the discretization approach poses a bijection between the finite parameter space and the set of arms.

Two approaches present themselves clearly as candidates for future research regarding how to handle the ever-present curse of dimensionality. Firstly, function approximation is already widely used in other reinforcement learning methods [21] as a way to handle large or continuous action and state spaces in more general MDP algorithms, and it is plausible that function approximation should transfer well to the task of indirect inference using bandit algorithms. Secondly, there is theory regarding bandit optimization under continuous action spaces for restricted families of reward distributions. As our formulation equates parameter spaces with bandit action spaces, the discretization approach is admittedly naive and somewhat roundabout. A more sophisticated approach might use bandit algorithms for continuous action spaces to address the issue of continuous parameter spaces more directly.

VII. Conclusion

In this bachelor's thesis we present a novel approach to indirect inference using bandit optimization. Concretely, the approach converts the problem of indirect inference into an equivalent stateless Markov decision process optimization task and uses well-established algorithms for the optimization procedure. This circumvents some limitations imposed on the applicability of indirect inference in its original formulation. Our approach is demonstrated to work well in practice on the simple task of estimating the parameters of a Gaussian model, as well as on the non-trivial task of estimating the parameters of a mixture normal model. The main limiting factor of the approach in its current form is the linchpin role of discretization in converting a (generally continuous) parameter space into a finite action set: All well-known issues of large-scale discretization—chiefly, the curse of dimensionality—apply. Some avenues towards a more refined approach to general indirect inference with bandit optimization are introduced.

Appendix A
Pseudocode for Bandit algorithms

Acknowledgment

The authors would like to extend the greatest of thanks to our mentor Braghadeesh Lakshminarayanan for his valuable guidance, helpful explanations, and keen insights throughout the project. We are particularly thankful for Braghadeesh eagerly encouraging us to dare to explore our own ideas, and for his support in that process.

References

[1] L. Ljung, *System Identification: Theory for the User*. London, UK: Pearson Education, 1998.

[2] G. Casella and R. L. Berger, *Statistical Inference, 2nd Ed*. Pacific Grove, CA: Thomson Learning, 2002.

[3] A. A. Smith Jr, "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics*, vol. 8, no. S1, pp. S63–S84, 1993.

[4] C. Gourieroux, A. Monfort, and E. Renault, "Indirect inference," *Journal of applied econometrics*, vol. 8, no. S1, pp. S85–S118, 1993.

[5] C. Gourieroux and A. Monfort, *Simulation-based econometric methods*. Oxford, UK: Oxford University Press, 1996.

[6] C. Monfardini, "Estimating stochastic volatility models through indirect inference," *The Econometrics Journal*, vol. 1, no. 1, pp. 113–128, 1998.

[7] S. Chaudhuri, D. T. Frazier, and E. Renault, "Indirect inference with endogenously missing exogenous variables," *Journal of Econometrics*, vol. 205, no. 1, pp. 55–75, 2018.

[8] P. Rossi, *Bayesian non- and semi-parametric methods and applications*. Princeton, NJ: Princeton University Press, 2014.

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

[10] M. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. Hoboken, NJ: John Wiley & Sons, 2014.

[11] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge, UK: Cambridge University Press, 2020.

[12] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, pp. 235–256, 2002.

[13] C. A. Sims, "Macroeconomics and reality," *Econometrica*, vol. 48, no. 1, pp. 1–48, Jan. 1980.

[14] H. Robbins and S. Monro, "A stochastic approximation method," *The annals of mathematical statistics*, vol. 22, no. 3, pp. 400–407, Sep. 1951.

[15] T. L. Lai, "Stochastic approximation," *The annals of Statistics*, vol. 31, no. 2, pp. 391–406, 2003.

[16] M. F. Schilling, A. E. Watkins, and W. Watkins, "Is human height bimodal?" *The American Statistician*, vol. 56, no. 3, pp. 223–229, 2002.

[17] R. L. Folk and W. C. Ward, "Brazos river bar: a study in the significance of grain size parameters," *Journal of sedimentary petrology*, vol. 27, no. 1, pp. 3–26, Mar. 1957.

[18] G. H. S. Smith, A. P. Nicholas, and R. I. Ferguson, "Measuring and defining bimodal sediments: Problems and implications," *Water Resources Research*, vol. 33, no. 5, pp. 1179–1185, 1997.

[19] K. A. Ashman, C. M. Bird, and S. E. Zepf, "Detecting bimodality in astronomical datasets," *Astronomical Journal*, vol. 108, no. 6, Dec. 1994.

[20] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical analysis of finite mixture distributions*. Hoboken, NJ: John Wiley & Sons, 1985.

[21] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction, 2nd Ed*. Cambridge, MA: MIT Press, 2018.