DEPARTMENT OF
STATISTICS

Modern Bayesian Experimental Design

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2021

Adam Evan Foster
University College

Department of Statistics
University of Oxford

Abstract

Bayesian experimental design (BED) is a powerful mathematical framework with diverse applications that include asking the most pertinent question in an online survey, designing a laboratory experiment, choosing sensor locations, obtaining labels in active learning, searching for the maximum of an unknown function, and exploring an unknown environment. Automated design of optimal experiments will allow scientists to undertake experiments more efficiently, reach statistically valid conclusions more quickly, and unlock kinds of experiments that have been hitherto considered impractical. Unfortunately, widespread utilisation of BED for designing optimal experiments is not yet a reality. Adoption of BED is hampered by computational challenges that are inherent in finding experimental designs that maximise the expected information that will be gained about the underlying process by running the experiment. Broadly, the computational challenges can be broken down into three parts of increasing complexity: 1) estimating the expected information gain, 2) optimising the expected information gain over the space of possible designs, and 3) choosing a sequence of optimal designs whilst incorporating feedback from the experiment.

The goal of this thesis is to present methods to tackle these computational challenges by taking inspiration from the rapid development of modern probabilistic machine learning in areas such as variational inference, amortised inference, stochastic gradient optimisation, probabilistic deep learning, Monte Carlo methods, likelihood-free inference, constrastive learning and reinforcement learning. We show how concepts from these areas can be brought together to create a modern approach to BED that begins to overcome the computational restrictions on the Bayesian design of experiments.

Specifically, we begin by presenting advances in the estimation of expected information gain, incorporating ideas from variational and amortised inference. We make several contributions to the optimisation of experimental designs using stochastic gradient methods. Finally, we turn to the sequential design problem, and demonstrate how efficient, adaptive design may be achieved through the use of a policy. In concert, these methods allow us to expand the range of circumstances in which BED can be used to design optimal experiments, with implications for both machine learning and the sciences.

Acknowledgements

I would like to begin by thanking my supervisors Yee Whye Teh and Tom Rainforth for their kindness, patience and guidance throughout my DPhil. To Yee Whye, I particularly owe gratitude for his thoughtfulness and calmness, his ability to always give some insight on the mathematical problem that I happened to be considering and for his unwavering commitment to allow me to work on whichever project I find most interesting. To Tom, I would like to express my thanks for the enthusiasm and interest he has for our shared work—I would have given up many years ago without his input and his drive to continually improve our understanding—and for teaching me many of the skills that I needed to succeed in my DPhil, including writing in something other than the 'Russian mathematician' style.

In the Department of Statistics at Oxford, I would like to thank Emile Mathieu for his friendship and support during our four year DPhil adventure, as well as Benjamin Bloem-Reddy, Chris Maddison, Dominic Richards, Jean-François Ton, Adam Kosiorek, Emilien Dupont, Anthony Caterini, Adam Goliński, Bobby He, Hyunjik Kim, Qinyi Zhang, Yuan Zhou, Michael Hutchinson and Desi Ivanova for their friendship and for making the department a good place to be.

I also owe a huge debt of gratitude to Noah Goodman for his inspiration and guidance during my time at Uber AI Labs, and for helping to set me on the exciting research path that this thesis is a partial culmination of. I would also like to thank Martin Jankowiak for his help and support with our joint research, as well as Eli Bingham, Paul Horsfall and the whole of the Pyro team for good times both in the office and out of it. To Takashi Goda, Tomohiko Hironaka and Wataru Kitade of the University of Tokyo I express my gratitude for their willingness to collaborate and share their expertise on MLMC. To Rattana Pukdee, Ilyas Malik and Matthew O'Meara, thank you for being wonderful co-authors. I am grateful to Árpi Vezér, Craig Glastonbury, Páidí Creed, Sam Abujudeh and Aaron Sim for their kindness and help during my time as an intern at BenevolentAI. This research would not have been possible without the generous financial suppport that I received from an EPSRC Excellence Award.

Finally, I would like to thank the friends and family who have been with me during the highs and lows of my studies. I am humbled by the kindness that my parents, siblings, and family have always shown me. To Diallo, thank you for the love and support that has sustained me on the last leg of my DPhil journey. To my friends from University College—James, Seb, Miles, Tales, Nonie, Colm, Mitch & Jess, Max and Rieke—thank you for so many happy memories of Oxford. And to Alexander Temple McCune for being the man you were and my friend, thank you.

Contents

Chapter 1

Introduction and Literature Review

In this chapter, we present a self-contained literature review of the field of Bayesian Experimental Design. In Chapter 2, we present work on the variational estimation of the expected information gain. Chapter 3 discusses stochastic-gradient optimisation of Bayesian experimental designs. In Chapter 4, we extend this by deriving an unbiased gradient estimator of the expected information gain. In Chapter 5, we consider the problem of sequential experimental design, and present a policy-based algorithm for real-time adaptive experimental design. Chapter 6 is an extension of Chapter 5 to implicit models. We conclude with the discussion, in Chapter 7.

Bayesian Experimental Design Literature Review

Connections with Bayesian Active Learning, Bayesian Optimisation and Bayesian Reinforcement Learning

Adam Foster
Department of Statistics, University of Oxford

July 2021

1 Introduction

If true knowledge arises from empirical observations, it is natural to ask which kinds of observations we should actively seek out to further our understanding of nature. In its broadest sense, this is the question that the design of experiments seeks to answer. An *experimental design* is an allocation of resources—e.g. time, human attention, chemical reagents, physical space—that will be used to obtain empirical observations. The *design space* is the set of designs that we could feasibly choose for the experiment; the problem of experimental design is to pick a design to use for the real experiment. The choice of design is an important one: we could easily waste resources on poorly designed experiments that do not further our understanding. By carefully designing experiments, we can efficiently gather empirical observations that lead to new ideas, hypotheses, conclusions and models.

It is therefore unsurprising that we find experimental design to be a key concern in scientific disciplines as diverse as psychology (Myung et al., 2013), bioinformatics (Vanlier et al., 2012), pharmacology (Lyu et al., 2019), physics (Dushenko et al., 2020), neuroscience (Shababo et al., 2013), astronomy (Loredo, 2004) and engineering (Papadimitriou, 2004). It is also is a natural abstraction for several central problems in machine learning, including active learning (Houlsby et al., 2011; Gal et al., 2017), Bayesian optimisation (Hernández-Lobato et al., 2014; Shahriari et al., 2015) and exploration (Sun et al., 2011; Shyam et al., 2019).

In many practical cases, experimental design is not used just once. Indeed, many experiments are naturally *adaptive*: they are an iterative process in which we can select the designs for later iterations on the basis of data already gathered. This allows feedback from the outcome of one experiment iteration to be used to guide the design of the next iteration. This setting can be particularly powerful because, as we gain some information about the system, it may become clearer how we should proceed to design our experiments to investigate further, thereby honing in quickly on the truth.

To choose between different possible experimental designs requires an objective function. In general, the objective depends not only on known quantities (such as the cost of the experiment), but also on the not-yet-observed outcome of the experiment and potentially on other unobserved quantities. For example, the objective function for a chemical experiment might reward correctly synthesising a product, something that will only be observed once the experiment is completed. To reason about objective functions that depend on unknowns in this way requires the incorporation of some *a priori* knowledge. This *a priori* knowledge is then used to select the design before commencing the experiment. In this work, we focus on the Bayesian approach to this problem (Lindley, 1956, 1972; Chaloner and Verdinelli, 1995; Ryan et al., 2016; Foster et al., 2019) in which *a priori* knowledge is encoded in two ways—first, the specification of a model for the experiment, and second in the prior distribution for the unknown parameters of that model. Typically, the model itself is assumed to be correct. The prior distribution explicitly represents initial beliefs about unknown parameters of the model. Furthermore, uncertainty in the prior is exactly the epistemic uncertainty that can be reduced by running experiments and collecting data, resulting in more precise *a posteriori* knowledge.

In this literature review, we begin with a brief survey of foundational concepts in Bayesian data analysis

(Sec. 2). We then turn to the core theory of Bayesian experimental design (Sec. 3), discussing criteria that have been used within the statistics community, with an emphasis on expected information gain. In Sec. 4, we discuss computational methods for Bayesian experimental that have been used within statistics, and in Sec. 5 we discuss active learning. We then discuss models in which the target of experimental design is embedded in a larger model (Sec. 6); Bayesian optimisation (Sec. 7) is a specific instance of this. Finally, we delve into the theory of the sequential experimental design problem (Sec. 8), and highlight connections with exploration and Bayesian reinforcement learning (Sec. 9).

2 Background on Bayesian statistics

We first introduce necessary notation and key concepts in Bayesian data analysis[1]. The first ingredient of any Bayesian analysis is a full probability model that places a joint distribution over all observable and unobservable quantities. We denote the parameters of interest, also called the latent variable of interest, by $\theta \in \Theta$. This may be a scalar, vector, or a function depending on the model. We denote the observed data, or outcome, as $y \in \mathcal{Y}$. The full probability model is simply a probability distribution $p(\theta, y)$ on $\Theta \times \mathcal{Y}$. Typically, the full probability model can be factorised as

$$p(\theta, y) = p(\theta)p(y|\theta) \tag{1}$$

where $p(\theta)$ denotes the *prior* on θ, and $p(y|\theta)$ is the *likelihood* function[2], or sampling distribution.

Since we are interested in experimental design, we also introduce the *design* or *covariate* $\xi \in \Xi$. This is not typically treated as a random variable, because it is assumed to be directly under the experimenter's control. Instead, for each possible design ξ, we have a different probability model $p(\theta, y|\xi)$. Different choices of ξ should not alter our prior $p(\theta)$, thus the change in the probability model is only felt through the likelihood, so we can write $p(\theta, y|\xi) = p(\theta)p(y|\theta, \xi)$. Intuitively, this says that the design of the experiment ξ does not change the natural environment, but it can change the outcome of an experiment that we choose to run.

Once we have chosen ξ and run our experiment to obtain y, we can make probability statements about θ by applying Bayes' Rule to calculate the posterior

$$p(\theta|\xi, y) = \frac{p(\theta)p(y|\theta, \xi)}{\int_\Theta p(\theta')p(y|\theta', \xi)d\theta'} = \frac{p(\theta)p(y|\theta, \xi)}{p(y|\xi)}. \tag{2}$$

In general, actually performing Bayesian inference to calculate $p(\theta|\xi, y)$ can be computationally challenging.

2.1 Explicit and implicit models

If the likelihood $p(y|\theta, \xi)$ is known in closed form, then the probability model is called an *explicit likelihood* model. Most Bayesian statistics assumes an explicit likelihood. If no closed form likelihood is available, the model is an *implicit likelihood* model (Sisson et al., 2018). Implicit models often arise when θ, y and ξ are related by a simulator (Alsing et al., 2019; Brehmer et al., 2018; Gonçalves et al., 2020) that can produce samples of $p(y|\theta, \xi)$, but does not have a closed form probability density.

Similarly, if $p(\theta)$ is known in closed form, then the model is said to have an *explicit prior*, otherwise, the prior is said to be *implicit*.

2.2 Sequential data collection

So far, we have considered choosing ξ, collecting y, and analysing the data by computing $p(\theta|\xi, y)$. A more realistic setting is to consider a sequence $\xi_1 \dots \xi_T$ of designs with corresponding outcomes $y_1, \dots, y_T$. This

[1] More details on Bayesian data analysis can be found in modern textbooks on the topic, such as Gelman et al. (2013) and Kruschke (2014).

[2] Strictly, the likelihood describes the sampling distribution $p(y|\theta)$ as a function of θ for a fixed y; we use likelihood in a slightly looser sense to refer to $p(y|\theta)$ in general.

means that we run T different experiments with T different designs, each with its own corresponding outcome. The value of θ, although unknown, is assumed to be the same across all the T experiments—that means that we are conducting multiple experiments in the same natural environment to gather further information about it, instead of starting afresh in a new environment for each new experiment.

In an *exchangeable* model (Bloem-Reddy and Teh, 2019), the order of the experiments does not matter. This is equivalent (Øksendal, 2003) to the following factorisation of the full probability model

$$p(\theta, y_{1:T}|\xi_{1:T}) = p(\theta)\prod_{t=1}^{T} p(y_t|\theta, \xi_t). \tag{3}$$

for some random variable θ. The question is whether we can identify this θ with the model parameters of interest θ. In general, this is valid when there are no other model parameters besides θ. Indeed, in a full statistical model with parameters θ (Cox, 2006), it is common to assume that the outcomes of different experiments are independent given θ, which is equivalent to the factorisation in equation (3). We discuss the case in which there are other model parameters aside from θ in Sec. 6.

In non-exchangeable models, there is no assumption of conditional independence between experiments. Such models are uncommon, but can arise in settings such as time series (Pole et al., 2018). In a non-exchangeable model, the distribution of y_t can, for example, be influenced by y_{t-1} as well as by θ and ξ_t. Without loss of generality, the probability model for a non-exchangeable model can be written

$$p(\theta, y_{1:T}|\xi_{1:T}) = p(\theta)\prod_{t=1}^{T} p(y_t|\theta, \xi_{1:t}, y_{1:t-1}). \tag{4}$$

which encodes only the assumption that future experiments cannot affect the outcome of earlier experiments.

Static and adaptive experiments An orthogonal distinction in sequential experiments is how the designs are generated. In a *static* experiment, also called fixed, batch, or open loop (DiStefano III et al., 2014), the designs $\xi_1, \ldots, \xi_T$ are chosen before the beginning of the experiment. In an *adaptive* experiment (Myung et al., 2013), each ξ_t is chosen depending on data already seen $\xi_1, \ldots, \xi_{t-1}, y_1, \ldots, y_{t-1}$. A simple consequence of the likelihood principle (Barnard et al., 1962; Birnbaum, 1962) is that the mode in which the ξ_t are generated does not affect the posterior distribution on θ calculated from the data. Indeed, suppose each new design is chosen adaptively from a density $p(\xi_t|\xi_{1:t-1}, y_{1:t-1})$. Then the resulting posterior distribution is

$$p(\theta|\xi_{1:T}, y_{1:T}) = \frac{p(\theta)\prod_{t=1}^{T} p(\xi_t|\xi_{1:t-1}, y_{1:t-1})p(y_t|\theta, \xi_{1:t}, y_{1:t-1})}{\int_\Theta p(\theta')\prod_{t=1}^{T} p(\xi_t|\xi_{1:t-1}, y_{1:t-1})p(y_t|\theta', \xi_{1:t}, y_{1:t-1})d\theta'} \tag{5}$$

$$= \frac{\prod_{t=1}^{T} p(\xi_t|\xi_{1:t-1}, y_{1:t-1})\ p(\theta)\prod_{t=1}^{T} p(y_t|\theta, \xi_{1:t}, y_{1:t-1})}{\prod_{t=1}^{T} p(\xi_t|\xi_{1:t-1}, y_{1:t-1})\ \int_\Theta p(\theta')\prod_{t=1}^{T} p(y_t|\theta', \xi_{1:t}, y_{1:t-1})d\theta'} \tag{6}$$

$$= \frac{p(\theta)\prod_{t=1}^{T} p(y_t|\theta, \xi_{1:t}, y_{1:t-1})}{\int_\Theta p(\theta')\prod_{t=1}^{T} p(y_t|\theta', \xi_{1:t}, y_{1:t-1})d\theta'}, \tag{7}$$

which is independent of the mechanism of choosing designs.

2.3 Bayesian decision making

After collecting data $\xi_{1:T}, y_{1:T}$, suppose that we must choose some decision δ, for example whether to prescribe a medication or not. The Bayesian approach to selecting the optimal decision (Lindley, 1972; Robert, 2007) is to specify a utility function $U(\delta, \theta)$ which should assign a value to the decision δ *in the case that* θ is the true value of the unobserved parameter. The optimal decision is then found by maximising expected utility under the current posterior

$$\delta^* = \arg\max_{\delta\in\Delta} \quad \mathbb{E}_{p(\theta|\xi_{1:T}, y_{1:T})}[U(\delta, \theta)] \tag{8}$$

For a more extensive discussion of Bayesian decision theory, see Berger (2013).

3 Bayesian Experimental Design

Experimental design with a Bayesian data analysis model means choosing the design using the likelihood model and the prior $p(\theta)$ as *a priori* information. What criterion should be used to select the design? Following from Bayesian decision theory, Lindley (1972) proposed a decision-theoretic approach to Bayesian experimental design that focuses on maximising a utility. Chaloner and Verdinelli (1995) provides a more recent summary of Lindley's approach.

First, let us restrict ourselves to a single design ξ with outcome y, leaving the sequential design problem to Sec. 8. In the spirit of Sec. 2.3, we consider a utility function $U(\theta, \xi, y)$ that may reflect the value of obtaining the data (ξ, y) when θ is the true value of the parameters, and may also incorporate costs of the experimental design and outcome. Whilst our discussion in Sec. 2.3 assumed that the data ξ, y had already been gathered, we now need to consider the choice of the design ξ. The order of operation for the experimenter is as follows:

1. choose design ξ;
2. perform experiment with design ξ, obtaining experimental outcome y;
3. compute the posterior $p(\theta|\xi, y)$;
4. the expected utility obtained is then $\mathbb{E}_{p(\theta|\xi,y)}[U(\theta, \xi, y)]$.

In order to choose ξ optimally, we should therefore consider the different possible observations y that could arise. Specifically, we will choose ξ to maximise the expected utility, taking an outer expectation over the observation y using the Bayesian marginal (also called prior predictive) distribution $p(y|\xi) = \mathbb{E}_{p(\theta)}[p(y|\theta, \xi)]$. This leads to the following method of choosing the optimal design

$$\xi^* = \arg\max_{\xi \in \Xi} \; \mathbb{E}_{p(y|\xi)} \left[\mathbb{E}_{p(\theta|\xi,y)}[U(\theta, \xi, y)] \right]. \tag{9}$$

Proposition 1 (Lindley (1972))**.** *It is not necessary to introduce randomness into the selection of* ξ*.*

Proof. Suppose we consider a randomised way of selecting ξ with distribution $p(\xi)$. The expected reward of this approach is

$$\mathbb{E}_{p(\xi)p(y|\xi)} \left[\mathbb{E}_{p(\theta|\xi,y)}[U(\theta, \xi, y)] \right] \leq \sup_{\xi \in \Xi} \mathbb{E}_{p(y|\xi)} \left[\mathbb{E}_{p(\theta|\xi,y)}[U(\theta, \xi, y)] \right] \tag{10}$$

where the righthand side is the expected utility using the non-random ξ^*. So a randomised design is not required. □

The remaining piece of the puzzle is to select a utility function. Some applications feature a highly problem-specific utility. In other cases, we can rely on general purpose utilities.

3.1 Expected Information Gain

Perhaps the most well-studied of all criteria for Bayesian experimental design is expected information gain (EIG). Within Bayesian experimental design, EIG appears to be dominant in a number of fields. EIG was proposed by Lindley (1956). Important statistical review papers (Chaloner and Verdinelli, 1995; Ryan et al., 2016) give EIG pride of place within Bayesian experimental design. In psychology, Myung et al. (2013) promote the use of EIG to run adaptive trials. Several toolboxes (Watson, 2017; Vincent and Rainforth, 2017) have been designed specifically for the problem of performing adaptive psychology trials using EIG as the criterion for selecting designs. Heck and Erdfelder (2019) suggest EIG for experimental design for cognitive models and Cavagnaro et al. (2010) consider its application in the context of model discrimination in cognitive science. Shababo et al. (2013) applied EIG maximisation within a Bayesian model of neural microcircuits to choose the right subset of neurons to stimulate in an experiment. Dushenko et al. (2020) proposed EIG as a criterion for designing measurement settings in magnetometry. In biochemistry, Busetto

et al. (2009) compared EIG with several other criteria for the design of experiments for biochemical dynamical systems, finding EIG to perform best. In pharmacology, Lyu et al. (2019); Foster et al. (2020) applied EIG maximisation to design experiments to calibrate a docking model. Loredo (2004) used EIG for active exploration, specifically investigating the scheduling of observations of a star to characterise the orbit of a planet. EIG has also been used in active learning, Bayesian optimisation and reinforcement learning. We discuss these fields separately in Sections 5, 7 and 9.

There are several reasons for the dominance of the EIG. First, it has mathematical properties that make it very natural for describing information gained from experimentation. We discuss some key properties of the EIG in this section, and we discuss EIG in sequential settings in Sec. 8. More practically, EIG applies to a range of linear and nonlinear models (unlike some criteria which are more restricted in their applicability) and handles both continuous and discrete θ.

What does EIG measure? EIG quantifies the amount of information that the experiment with design ξ is expected to produce about the unknown parameter of interest θ. A higher EIG indicates that doing the experiment with design ξ is likely to produce data that will be helpful in reducing uncertainty about the true value of θ.

To precisely define EIG, we utilise the rigorous probabilistic definition of information that was first given by Shannon (1948). Lindley (1956) used this work to quantify the information provided by an experiment. Lindley began by considering the Shannon *entropy* of a random variable θ

$$H[p(\theta)] = -\mathbb{E}_{p(\theta)}[\log p(\theta)]. \tag{11}$$

One interpretation of entropy is uncertainty in what the true value of θ is. In the experimental design context, we measure the amount of information that is gained about θ by performing the experiment with design ξ and obtaining outcome y using the reduction in entropy from the prior to the posterior. This is referred to as the information gain (IG)

$$U_{\mathcal{I}}(\xi, y) = \mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)]. \tag{12}$$

To obtain an objective function for ξ, we can use this utility within the decision-theoretic framework laid out in the preceding section. We substitute $U_{\mathcal{I}}$ into equation (9). This gives the overall objective function to select ξ: the *expected information gain* (EIG), formed by taking the expectation of $U_{\mathcal{I}}$ over $p(y|\xi)$, giving

$$\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)}\left[\mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)]\right]. \tag{13}$$

Proposition 2 (Lindley (1956))**.** *The EIG at design ξ, $\mathcal{I}(\xi)$, is the mutual information between y and θ under design ξ.*

Proof. By repeatedly using Bayes Theorem, we have

$$\begin{aligned}
\mathcal{I}(\xi) &= \mathbb{E}_{p(y|\xi)}\left[\mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)]\right] && (14)\\
&= \mathbb{E}_{p(y|\xi)p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)] && (15)\\
&= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[\log p(\theta|\xi, y) - \log p(\theta)] && (16)\\
&= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(\theta|\xi, y)}{p(\theta)}\right] && (17)\\
&= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(\theta)p(y|\theta, \xi)}{p(\theta)p(y|\xi)}\right]. && (18)
\end{aligned}$$

□

Proposition 3. *EIG is unchanged under invertible reparametrisations of θ and y.*

Proof. This follows from the well-known property of mutual information (Cover, 1999). □

Proposition 4 (Bernardo (1979))**.** *EIG can equivalently be derived from the KL-divergence utility*

$$U_{KL}(\xi, y) = \mathrm{KL}\left(p(\theta|\xi, y)\|p(\theta)\right). \tag{19}$$

Proof. Substituting this utility into equation (9) gives us

$$I_{KL}(\xi) = \mathbb{E}_{p(y|\xi)}\left[\mathrm{KL}\left(p(\theta|\xi,y)\|p(\theta)\right)\right] \tag{20}$$

$$= \mathbb{E}_{p(y|\xi)p(\theta|y,\xi)}\left[\log \frac{p(\theta|y,\xi)}{p(\theta)}\right] \tag{21}$$

$$= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(\theta|y,\xi)}{p(\theta)}\right] = \mathcal{I}(\xi) \text{ by equation (17).} \tag{22}$$

□

Proposition 5 (Theorem 6 of Lindley (1956))**.** *EIG is convex in the likelihood.*

Proof. Let $\lambda \in [0,1]$ and ξ_0, ξ_1 be two designs. Suppose there exists a design ξ_λ with the following likelihood

$$p(y|\theta,\xi_\lambda) = \lambda p(y|\theta,\xi_0) + (1-\lambda)p(y|\theta,\xi_1). \tag{23}$$

We can interpret an experiment with likelihood $p(y|\theta,\xi_\lambda)$ as follows. With probability λ, the outcome y is sampled from $p(y|\theta,\xi_0)$, and with probability $1-\lambda$ it is sampled from $p(y|\theta,\xi_1)$, but it is unknown which of the two likelihoods was chosen. We could also consider a different experiment in which we observe y *and* the binary random variable u which indicates which likelihood was used. Intuitively, the latter experiment must contain at least as much information as the first. We can demonstrate this formally using the information chain rule. The expected information gain of the experiment with outcome u, y can be expanded as

$$I_{\xi,\lambda}(\theta;(u,y)) = I_\lambda(\theta;u) + I_{\xi,\lambda}(\theta;y|u), \tag{24}$$

we note θ and u are independent, and we expand the definition of the conditional mutual information

$$= \lambda I_{\xi_0}(\theta;y) + (1-\lambda)I_{\xi_1}(\theta;y). \tag{25}$$

We can also expand the same mutual information as

$$I_{\xi,\lambda}(\theta;(u,y)) = I_{\xi,\lambda}(\theta;y) + I_{\xi,\lambda}(\theta;u|y) \tag{26}$$

$$\geq I_{\xi,\lambda}(\theta;y) \tag{27}$$

since conditional mutual information is nonnegative. Finally, Proposition 2 tells us that $I_\lambda(\theta;y) = \mathcal{I}(\lambda)$. Hence,

$$\mathcal{I}(\xi_\lambda) \leq \lambda\mathcal{I}(\xi_0) + (1-\lambda)\mathcal{I}(\xi_1). \tag{28}$$

□

Proposition 6 (Sebastiani and Wynn (2000))**.** *EIG can be written as* $\mathcal{I}(\xi) = \mathbb{E}_{p(\theta)}\left[H[p(y|\xi)] - H[p(y|\theta,\xi)]\right]$. *Furthermore, when* $H[p(y|\theta,\xi)]$ *does not depend on* ξ*, EIG maximisation is equivalent to maximum entropy design which selects* ξ *to maximise* $H[p(y|\xi)]$.

Proof. Starting from Proposition 2, we have

$$\mathcal{I}(\xi) = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(\theta)p(y|\theta,\xi)}{p(\theta)p(y|\xi)}\right] \tag{29}$$

$$= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[\log p(y|\theta,\xi) - \log p(y|\xi)] \tag{30}$$

$$= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[\log p(y|\theta,\xi)] - \mathbb{E}_{p(y|\xi)}[\log p(y|\xi)] \tag{31}$$

$$= \mathbb{E}_{p(\theta)}\left[H[p(y|\xi)] - H[p(y|\theta,\xi)]\right]. \tag{32}$$

Now, if $H[p(y|\theta,\xi)]$ is independent of ξ, then we have $\mathcal{I}(\xi) = H[p(y|\xi)] + \text{const.}$, so EIG maximisation and maximum entropy design lead to the same optimal design. □

Remark 7 (Smith and Gal (2018))**.** *The EIG at design* ξ *can be interpreted as a measure of epistemic uncertainty in the outcome of performing an experiment with design* ξ.

Proof. Equation (32) breaks the EIG into two terms. The first is the total entropy $H[p(y|\xi)]$, called the predictive entropy. The second is $-\mathbb{E}_{p(\theta)}[H[p(y|\theta,\xi)]]$, which represents the expectation of the uncertainty in y *conditional on* θ. We can view this as a measure of aleatoric uncertainty—uncertainty which cannot be eliminated by knowing θ exactly. The EIG is the difference between the total and aleatoric uncertainties, hence we can interpret it as epistemic uncertainty—the part of $H[p(y|\xi)]$ that can be reduced by learning about θ. □

This interpretation does have its limitations. First, this definition of epistemic uncertainty is a model-dependent quantity—if we choose a more powerful model, then some variation that had previously been characterised as aleatoric would now be seen as epistemic. It also rests on our model's ability to accurately capture aleatoric uncertainty. Second, the interpretation does not hold true in the case that θ is a function of a larger set of model parameters ψ, as in Sec. 6. This is because the term $-\mathbb{E}_{p(\theta)}[H[p(y|\theta,\xi)]]$ no longer represents aleatoric uncertainty, as it also includes some uncertainty that arises from not knowing the true value of ψ.

Other important features of the EIG in sequential experiments will be discussed in Section 8.

3.2 Alphabetic criteria

The EIG is only one approach to assigning value to the design of an experiment. Whilst the EIG has a number of attractive properties, it is not the only criterion to have been explored in the literature. Perhaps the more classical approach to experimental design is to use one of the 'alphabetic' criteria. Unlike the EIG, the alphabetic criteria stem from research into *non-Bayesian linear models*, as it was in this context that the alphabetic criteria were originally proposed. Authors have then sought to generalise these alphabetic criteria, first to the Bayesian linear model, and then to general Bayesian models. Unfortunately, the resulting criteria do not always map onto Lindley's general Bayesian methodology as outlined in equation (9). As the focus in this review is on the EIG, we provide only a brief introduction to the alphabetic criteria, emphasising the historical development from linear models, and the connection to the EIG.

3.2.1 Non-Bayesian linear model

The alphabetic criteria were initially proposed in the context of non-Bayesian experimental design for the linear model

$$y|\theta,\xi \sim N(\xi\theta,\sigma^2) \tag{33}$$

where ξ is the $n \times p$ design matrix and θ is a p-vector. (In a linear model, we would conventionally replace $\theta \to \beta$ and $\xi \to X$.) The least squares estimator for θ is $\hat{\theta} = (\xi^\top\xi)^{-1}\xi^\top y$. In frequentist analysis of this estimator, the covariance matrix of $\hat{\theta}$ is proportional to $(\xi^\top\xi)^{-1}$. To guide the choice of ξ, Box (1982) discussed the following notions of optimality of ξ

A-optimality minimise $\mathrm{Tr}(\xi^\top\xi)^{-1}$, or more generally, minimise $\mathrm{Tr}\, A(\xi^\top\xi)^{-1}$ for a matrix A;

D-optimality minimise $\det(\xi^\top\xi)^{-1}$;

E-optimality minimise $\max_i \lambda_i$, where $\lambda_1,\ldots,\lambda_p$ are the eigenvalues of $(\xi^\top\xi)^{-1}$;

G-optimality minimise $\sup_{c\in\mathcal{C}} c^\top(\xi^\top\xi)^{-1}c$, where $\mathcal{C}$ is some target region for prediction.

Other alphabetic criteria include

c-optimality (Elfving, 1952) minimise $c^\top(\xi^\top\xi)^{-1}c$ for some vector c.

T-optimality (Atkinson and Fedorov, 1975) for model discrimination, which maximises the minimal deviation between a null model and an alternative.

Several key results relate these classical criteria, such as Kiefer and Wolfowitz (1959).

3.2.2 Bayesian linear model

The alphabetic criteria can be extended to Bayesian linear models (Chaloner and Verdinelli, 1995), using the observation that the posterior covariance matrix for θ is proportional to $(\xi^\top \xi + \Sigma_0^{-1})^{-1}$ when we augment the model in equation (33) with a Gaussian prior $\theta \sim N(0, \Sigma_0)$. This allows a direct generalisation of the alphabetic criteria with $(\xi^\top \xi + \Sigma_0^{-1})^{-1}$ playing the role of $(\xi^\top \xi)^{-1}$. For example, Bayesian A-optimality minimises $\mathrm{Tr}(\xi^\top \xi + \Sigma_0^{-1})^{-1}$, and Bayesian D-optimality minimises $\det(\xi^\top \xi + \Sigma_0^{-1})^{-1}$.

One may well ask how these alphabetic criteria relate to our preceding work on the EIG. Superficially, the alphabetic criteria are simply functionals of the Gram matrix $\xi^\top \xi$, whilst the EIG is defined in terms of posterior entropy. Fortunately, there is a point of close connection between the two for the Bayesian linear model.

Proposition 8 (Chaloner and Verdinelli (1995))**.** *For a Bayesian linear model, Bayesian D-optimality and EIG optimality are equivalent.*

Proof. In the Bayesian linear model, the posterior on θ is Gaussian with covariance matrix that is proportional to $(\xi^\top \xi + \Sigma_0^{-1})^{-1}$, and is independent of y. The entropy of this Gaussian posterior is $\frac{1}{2} \log \det(\xi^\top \xi + \Sigma_0^{-1})^{-1} +$ a constant. Substituting this into equation (13), the EIG for the Bayesian linear model is

$$\mathcal{I}(\xi) = \tfrac{1}{2} \log \det \Sigma_0 - \tfrac{1}{2} \log \det(\xi^\top \xi + \Sigma_0^{-1})^{-1} - \text{const} = -\tfrac{1}{2} \log \det(\xi^\top \xi + \Sigma_0^{-1})^{-1} + \text{const}'. \tag{34}$$

Thus, EIG optimality (maximise $\mathcal{I}(\xi)$) and Bayesian D-optimality (minimise $\det(\xi^\top \xi + \Sigma_0^{-1})^{-1}$) lead to the same optimal design. □

3.2.3 Bayesian non-linear models

The 'classical' approach (Tsutakawa, 1972; Chaloner and Verdinelli, 1995) to generalising the alphabetic criteria to non-linear Bayesian models is to consider the Fisher information matrix (FIM), which is defined as

$$M(\theta, \xi) = -\mathbb{E}_{p(y|\theta,\xi)} \left[\frac{\partial^2}{\partial \theta^2} \log p(y|\theta, \xi) \right] \tag{35}$$

where $\partial^2 / \partial \theta^2$ denotes the Hessian when θ is a vector. The FIM has two important properties that motivate its use to extend the alphabetic criteria:

1. the FIM for the linear regression model is proportional to $(\xi^\top \xi)$;
2. the inverse FIM is related to the asymptotic covariance matrix of the Bayesian posterior by the Bernstein–von Mises Theorem (Van der Vaart, 2000).

For non-linear models, the FIM generally depends on θ as well as ξ, so forming a criterion for ξ involves an integral over $p(\theta)$. For instance, Chaloner and Verdinelli (1995) gives a Bayesian non-linear version of D-optimality as

$$U_{\text{Bayesian-}D}(\theta, \xi) = \log \det M(\theta, \xi)^{-1}; \tag{36}$$

substituting this utility in equation (9), leads to the optimality condition

$$\xi^* = \arg\max_{\xi} \mathbb{E}_{p(\theta)}[\log \det M(\theta, \xi)^{-1}]. \tag{37}$$

Using the FIM is not the only way to generalise the alphabetic criteria to non-linear models. Indeed, Ryan et al. (2016) takes issue with the classical FIM approach, suggesting that "to qualify as a 'fully Bayesian design', one must obtain the design by using a design criterion that is a functional of the posterior distribution". Whilst the EIG satisfies this requirement, the FIM extensions of the alphabetic criteria do not.

An approach to generalising the alphabetic criteria that is consistent with Ryan's definition of 'fully Bayesian' is to look at the covariance matrix of the Bayesian posterior $\mathrm{Cov}_{p(\theta|y,\xi)}[\theta]$, which depends on ξ and y and is

a functional of the posterior. For example, Ryan et al. (2016) mention two scalar objectives that can arise from this covariance matrix. One is termed the Bayesian D-posterior precision

$$U_{D\text{-precision}}(\xi, y) = \frac{1}{\det \operatorname{Cov}_{p(\theta'|y,\xi)}[\theta']} \tag{38}$$

the other is quadratic loss

$$U_Q(\xi, y, \theta) = (\theta - \hat{\theta}(y, \xi))^\top A(\theta - \hat{\theta}(y, \xi)) \tag{39}$$

for some matrix A and for some posterior functional estimate $\hat{\theta}(y, \xi)$ of θ, such as the posterior mean. Both can be applied in the general framework of equation (9).

4 Computational methods for one-step design

Choosing Bayesian-optimal experimental designs brings tremendous promise for obtaining information more efficiently. The utilisation of this method, however, is practically limited by the difficulty of quickly obtaining accurate estimates of the design criterion. This is particularly true of the EIG, and we focus on computational methods for the EIG in this section. The mathematical problem that must be solved to find the optimal design is the EIG maximisation problem

$$\begin{aligned} \xi^* &= \arg\max_{\xi \in \Xi} \mathcal{I}(\xi) \\ &= \arg\max_{\xi \in \Xi} \mathbb{E}_{p(y|\xi)} \left[\mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)] \right]. \end{aligned} \tag{40}$$

Note that here we are restricting ourselves to one-step experimental design, with sequential and adaptive design being left to Sec. 8.

Computational methods for solving the EIG maximisation problem defined in equation (40) can generally be further broken down into two parts. First, they often make point estimates of the EIG criterion at various candidate designs ξ. The difficulty of this estimation procedure can immediately be seen from the definition of the EIG. It entails the calculation of the posterior entropy $-\mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)]$. For large-scale Bayesian models, density estimation of the posterior constitutes a complex computational task in of itself. However, for the EIG, the problem is more challenging because the posterior entropy occurs inside an expectation $\mathbb{E}_{p(y|\xi)}$ over the observation y. Thus, a direct approach to estimating the EIG amounts to *nested* estimation of potentially intractable posterior distributions. It is for this reason that EIG estimation is sometimes referred to as a 'double intractability' (Foster et al., 2019). In Sec. 4.1, we review methods that have been proposed for EIG estimation.

Second, there are still further difficulties in the problem of *optimising* the EIG objective function over the space Ξ of possible designs. In the most naive methods, the optimisation simply adds an additional layer of nesting onto the EIG estimation computations, with an outer optimiser searching over candidate designs and feeding them into the EIG estimation. Such optimisation procedures are generally best suited for smaller problems; for larger ones, more sophisticated approaches to the optimisation of the design have been studied. In Sec. 4.2 we review a range of techniques that have been proposed for this optimisation.

Computational advances, both in the estimation and the optimisation of EIG, have significantly broadened the range of Bayesian models and design spaces for which Bayesian experimental design is a realistic possibility for practitioners.

4.1 Point estimates of EIG

The EIG, $\mathcal{I}(\xi)$, represents the expected reduction in Shannon entropy between the prior and posterior (see Sec. 3.1). The first step in utilising EIG for experimental design is to compute an estimate of the EIG for a single design ξ. Since $\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)}\left[H[p(\theta)] - H[p(\theta|\xi, y)]\right]$ involves an expectation over $y \sim p(y|\xi)$ of the posterior entropy $H[p(\theta|y, \xi)]$, a direct approach to its estimation requires repeated computations of the

posterior $p(\theta|y,\xi)$ with different simulated observations y. Given that calculating just one posterior can be intractable, it can readily be observed that EIG estimation is a computationally challenging problem.

A critical distinction when computing the EIG is whether the model has an explicit or an implicit likelihood (see Sec. 2.1 for a definition). In general, the explicit likelihood case contains strictly more information about the model, and so results in an easier, yet still doubly intractable, computational problem for the EIG. The implicit likelihood case is more challenging still, as the unknown likelihood typically has to be estimated in some way. We review computational methods for EIG estimation in both cases.

4.1.1 Explicit likelihood models

The existence of an explicit likelihood allows conventional approaches to posterior estimation, including MCMC, importance sampling, and Laplace approximation, to be used. Each leads to a family of well-studied approaches to EIG estimation. However, perhaps the most important computational methods are those which side-step direct estimation of the posterior, focusing on estimates of the marginal likelihoods $p(y|\xi)$ only. The Nested Monte Carlo estimator (Ryan, 2003) is the canonical method in this class, and been widely applied with a number of explicit likelihood models.

MCMC A natural approach that is mostly suited to low θ dimension problems, is to estimate the posterior using Markov Chain Monte Carlo (MCMC) (Andrieu et al., 2003). Unfortunately, MCMC only produces samples of the target density. This is problematic for EIG estimation, which also requires access to the posterior density $p(\theta|\xi,y)$. To overcome this, Heinrich et al. (2020) used MCMC to sample the posterior, and a Gaussian Mixture Model (Hastie et al., 2009, Sec. 6.8) to perform density estimation of the posterior. MCMC has also been applied to estimate non-EIG criteria for Bayesian experimental design (Wakefield, 1994; Han and Chaloner, 2004).

Importance sampling Another family of methods for EIG estimation is based on importance sampling. These methods begin with the key observation that estimating the posterior density is not actually required for EIG estimation, because we can write

$$\log \frac{p(\theta|\xi,y)}{p(\theta)} = \log \frac{p(y|\theta,\xi)}{p(y|\xi)}. \tag{41}$$

The approach of Cook et al. (2008); Ryan et al. (2014) is to estimate $p(y|\xi)$ using Monte Carlo samples from the prior, leading to the estimator

$$\log p(y|\xi) \approx \log\left(\frac{1}{N}\sum_{n=1}^{N} p(y|\theta_n,\xi)\right) \text{ where } \theta_1,\ldots,\theta_N \overset{\text{i.i.d.}}{\sim} p(\theta), \tag{42}$$

the other component of the likelihood ratio $p(y|\theta,\xi)/p(y|\xi)$ is the known likelihood. Cook et al. (2008); Ryan et al. (2014) then estimate $\mathbb{E}_{p(\theta|\xi,y)}[\log p(y|\theta,\xi)]$ for some fixed y by using importance sampling. Specifically, given some fixed y and the set of samples $\theta_1,\ldots,\theta_n$ drawn independently of $p(\theta)$, they use the estimator

$$\mathbb{E}_{p(\theta|\xi,y)}[\log p(y|\theta,\xi)] \approx \frac{1}{N}\sum_{n=1}^{N} \frac{p(y|\theta_n,\xi)}{\frac{1}{N}\sum_{p=1}^{N} p(y|\theta_p,\xi)} \log p(y|\theta_n,\xi). \tag{43}$$

The final estimator of EIG is formed by combining this estimator with the estimator in equation (42) for $\log p(y|\xi)$, and then taking the Monte Carlo integral over $y \sim p(y|\xi)$, giving

$$\mathcal{I}(\xi) \approx \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{N}\sum_{n=1}^{N} \frac{p(y_m|\theta_n,\xi)}{\frac{1}{N}\sum_{p=1}^{N} p(y_m|\theta_p,\xi)} \log p(y_m|\theta_n,\xi) - \log\left(\frac{1}{N}\sum_{n=1}^{N} p(y_m|\theta_n,\xi)\right)\right] \tag{44}$$

where $y_1\ldots,y_m \overset{\text{i.i.d.}}{\sim} p(y|\xi)$ and $\theta_1,\ldots,\theta_n \overset{\text{i.i.d.}}{\sim} p(\theta)$ are independent.

Monte Carlo and Nested Monte Carlo Hamada et al. (2001); Ryan (2003) considered a closely related family of estimators. They also used equation (41) to avoid computing posterior densities. Unlike Cook et al. (2008); Ryan et al. (2014), they observed that $p(y|\xi)p(\theta|\xi,y) = p(\theta)p(y|\theta,\xi)$, allowing them to write the EIG as

$$\mathcal{I}(\xi) = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(y|\theta,\xi)}{p(y|\xi)}\right]. \tag{45}$$

The only unknown quantity in the integrand here is $p(y|\xi)$. Assuming some estimator $\hat{p}(y|\xi)$ for $p(y|\xi)$, we have the Monte Carlo estimator

$$\mathcal{I}(\xi) \approx \frac{1}{N}\sum_{n=1}^{N} \log \frac{p(y_n|\theta_n,\xi)}{\hat{p}(y_n|\xi)} \text{ where } \theta_n, y_n \overset{\text{i.i.d.}}{\sim} p(\theta)p(y|\theta,\xi). \tag{46}$$

In Hamada et al. (2001), $\hat{p}$ was computed by numerical integration, for a low dimensional θ. In Ryan (2003), two approaches for $\hat{p}$ were considered—the first was a Laplacian approximation using the posterior mode $\hat{\theta}$. The second was to use to an inner Monte Carlo estimation step to estimate $p(y|\xi)$ as in equation (42). This latter approach, also considered by Myung et al. (2013); Rainforth (2017), results in the double loop, or Nested Monte Carlo (NMC) estimator of EIG

$$\hat{\mathcal{I}}_{NMC}(\xi) = \frac{1}{N}\sum_{n=1}^{N} \log \frac{p(y_n|\theta_n,\xi)}{\frac{1}{M}\sum_{m=1}^{M} p(y_n|\theta'_m,\xi)} \text{ where } \theta_n, y_n \overset{\text{i.i.d.}}{\sim} p(\theta)p(y|\theta,\xi), \theta'_m \overset{\text{i.i.d.}}{\sim} p(\theta). \tag{47}$$

The asymptotic properties of this estimator were studied by Rainforth et al. (2018); Zheng et al. (2018); Beck et al. (2018), showing that $\hat{\mathcal{I}}_{NMC}(\xi)$ converges to $\mathcal{I}(\xi)$ with asymptotic error $\mathcal{O}(N^{-1}) + \mathcal{O}(M^{-2})$. Hence, it is optimal to set $M \propto \sqrt{N}$.

Laplace approximation Another important line of work (Lewi et al., 2009; Cavagnaro et al., 2010; Long et al., 2013) uses a Laplace approximation to the posterior to estimate the posterior entropy. The Laplace estimate uses the following Taylor expansion of a scalar function about a point $\hat{\theta}$

$$f(\theta) \approx f(\hat{\theta}) + (\theta-\hat{\theta})^\top \left.\frac{\partial f}{\partial \theta}\right|_{\hat{\theta}} + (\theta-\hat{\theta})^\top \left.\frac{\partial^2 f}{\partial \theta^2}\right|_{\hat{\theta}} (\theta-\hat{\theta}). \tag{48}$$

If we apply this approximation to the log posterior density $f(\theta) = \log p(\theta|\xi,y) = \log p(\theta) + \log p(y|\theta,\xi) + C$ at a point $\hat{\theta}$ for which the log posterior density has zero gradient, then we find the following Gaussian approximation

$$\log p(\theta|\xi,y) \approx (\theta-\hat{\theta})^\top \hat{\Sigma}^{-1}(\theta-\hat{\theta}) + C' \text{ where } \hat{\Sigma}^{-1} = \left.\frac{\partial^2 \log\left(p(\theta)p(y|\theta,\xi)\right)}{\partial \theta^2}\right|_{\hat{\theta}}. \tag{49}$$

One advantage of this approach is that the entropy of this Gaussian approximation is known in closed form. A drawback is that the Laplace approximation makes a strong structural assumption about the posterior. This was partially relaxed by Long (2021), who considered a multi-modal Laplace approximation. Another approach is to combine Laplace estimation and importance sampling (Ryan et al., 2015). Finally, Beck et al. (2018) analysed the standard Laplace estimator, and further proposed combining Laplace importance sampling with the NMC estimator.

4.1.2 Implicit likelihood models

When the likelihood is not available, EIG estimation is strictly more difficult than when the likelihood is known in closed form. A direct approach to re-use explicit likelihood EIG estimators is to approximate the likelihood, and use this surrogate approximation as if it were the true likelihood. Alternatively, authors have focused on existing methods for likelihood-free inference, which give posterior estimates for implicit likelihood models without requiring knowledge of the likelihood.

Approximating the likelihood In some models, the likelihood $p(y|\theta,\xi)$ can be computed, but it is too expensive to be used in extensive calculation. The approach of Huan and Marzouk (2013) is to approximate the likelihood using a polynomial chaos expansion. Here, it is necessary to use a small number of evaluations of the likelihood to compute the polynomial chaos coefficients, but once this is done, the surrogate polynomial chaos approximate likelihood can be used in place of the true likelihood for all other calculations. Huan and Marzouk (2013) specifically use the polynomial chaos approximation within a NMC estimator of the EIG.

Overstall and McGree (2020) also consider approximating the likelihood. They assume a parametric family for distributions over y with parameters ϕ, so that $y|\theta,\xi \sim \mathcal{H}_X(\phi_f(\theta,\xi))$. They estimate the function $\phi_f(\theta,\xi)$ using a Gaussian Process (Williams and Rasmussen, 2006), trained with data obtained by maximum likelihood estimation of ϕ_f. We note the close connections between this idea and Foster et al. (2019).

Approximate Bayesian Computation Approximate Bayesian Computation (ABC) (Csilléry et al., 2010) is a family of methods for performing inference without a tractable likelihood. In its simplest form, ABC simulates $(\tilde{\theta}_i, \tilde{y}_i)_{i=1}^N$ from the joint model $p(\theta, y|\xi)$. Given a metric ρ on $\mathcal{Y}$, a sample $\tilde{\theta}_i$ is accepted as a valid sample from $p(\theta|\xi, y)$ if

$$\rho(y, \tilde{y}_i) < \epsilon \tag{50}$$

for tolerance ϵ. Drovandi and Pettitt (2013), Hainy et al. (2016), Price et al. (2016) and Dehideniya et al. (2018) have applied ABC within the context of Bayesian experimental design.

LFIRE Another more recent approach to inference in intractable likelihood models is Likelihood-free Inference by Ratio Estimation (LFIRE) (Thomas et al., 2016). This method uses logistic regression to approximate the *likelihood ratio*

$$r(\xi,\theta,y) = \frac{p(y|\theta,\xi)}{p(y|\xi)}. \tag{51}$$

Importantly, this ratio is exactly the likelihood ratio that appears in the definition of the EIG, indeed we have $\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)p(\theta|\xi,y)}[\log r(\xi,\theta,y)] = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[\log r(\xi,\theta,y)]$. Thus, if we are able to estimate $r(\xi,\theta,y)$ accurately, then EIG estimation can be performed by simple Monte Carlo integration. On this basis, LFIRE was applied in a Bayesian experimental design context by Kleinegesse and Gutmann (2018). They sampled $(\theta_i, y_i)_{i=1}^N$ from the joint model $p(\theta)p(y|\theta,\xi)$. For each θ_i, they trained a logistic regression model to distinguish samples from $p(y|\theta_i,\xi)$ and $p(y|\xi)$. This results in an estimate $\hat{r}(\xi,\theta_i,y)$ of $r(\xi,\theta_i,y)$ across different values of y. Finally, they form the Monte Carlo estimate of the EIG

$$\mathcal{I}(\xi) \approx \frac{1}{N}\sum_{i=1}^{N} \hat{r}(\xi,\theta_i,y_i). \tag{52}$$

4.2 Optimisation of EIG

We now turn to the problem of optimising the EIG over the design space Ξ. The simpler methods to perform this optimisation use point estimates of EIG. For finite Ξ, we can estimate EIG for every design. For infinite Ξ, we can use the EIG estimates at some design points to try and infer the EIG at others. Most simply, this could take the form of fitting a regression model to predict $\mathcal{I}(\xi)$ from ξ. More advanced methods use Bayesian optimisation—this both fits a Bayesian regression model of this form and uses Bayesian uncertainty estimates to propose new designs at which to compute the EIG. Another branch of thinking folds the EIG optimisation problem back into the problem of sampling from an unnormalised density. In this approach, the unnormalised density in question places high mass where the utility $U(\theta,\xi,y)$ is high. All aforementioned approaches are zeroth order methods—they do not make use of derivatives of $U(\theta,\xi,y)$. Using gradient information, albeit approximate, leads to a class of first order methods for EIG optimisation.

4.2.1 Discrete design space

For a small, discrete design space, the simplest option is to form separate estimates of $\mathcal{I}(\xi)$ for each $\xi \in \Xi$, and choose the design with the highest estimated EIG. This approach was taken by Carlin et al. (1998); Palmer and Müller (1998) and others. Vincent and Rainforth (2017) dynamically allocated resources between different discrete designs using ideas from the theory of bandit optimisation (Neufeld et al., 2014). In essence, this approach provides more accurate EIG estimates for designs that are likely to be optimal, spending less time on designs that are not promising.

4.2.2 Continuous design space

Discretisation Perhaps the simplest approach to continuous design optimisation is to discretise the design space, for example using uniformly or log-uniformly spaced points (Ryan, 2003; van den Berg et al., 2003; Watson, 2017; Vincent and Rainforth, 2017). Alternatively, a discrete set of candidate designs can be chosen by hand by the experimenter, and each evaluated (Han and Chaloner, 2004; Terejanu et al., 2012; Lyu et al., 2019).

Curve fitting Given a finite set of randomly sampled designs ξ_i with EIG estimates $\hat{\mathcal{I}}(\xi_i)$, Müller and Parmigiani (1995) proposed a curve fitting approach that fits a regression model to this data. The optimal design is then estimated as the optimum of the fitted regression model.

Bayesian optimisation Beyond simple curve fitting, Bayesian Optimisation (BO) (Snoek et al., 2012) is a well-established method for gradient-free optimisation. Like any other curve fitting approach, BO fits a model, specifically a Gaussian Process (GP), (Williams and Rasmussen, 2006) to the observed data $(\xi_i, \hat{\mathcal{I}}(\xi_i))$. However, BO iteratively suggests new designs at which to estimate the EIG, in order to efficiently seek the optimal design. We fully discuss BO and its connection with Bayesian experimental design itself in Sec. 7. For the purposes of solving the EIG optimisation problem, equation (40), we treat BO as a black box optimisation algorithm. The application of BO to optimising EIG over the design space was explored by Kleinegesse and Gutmann (2018); Foster et al. (2019); von Kügelgen et al. (2019).

Co-ordinate exchange The classical co-ordinate exchange algorithm for optimising design was proposed by Meyer and Nachtsheim (1995). Overstall and Woods (2017) proposed Approximate Co-ordinate Exchange. This is a two phase optimisation algorithm specifically designed for Bayesian experimental design. In the first phase, designs are optimised co-ordinate-wise by fitting a one-dimensional GP to the EIG surface for each co-ordinate in turn, with other elements of the design held fixed, and selecting the optimal value for that co-ordinate. In the second phase, different co-ordinates of the design are aggregated using a point exchange algorithm (Meyer and Nachtsheim, 1995; Atkinson et al., 2007).

Optimisation by sampling Clyde et al. (1996) proposed an approach to optimising the design that uses algorithms for sampling unnormalised densities. Their approach applies to any utility $U(\theta, \xi, y) > 0$ in the framework of equation (9). The authors define an augmented probability model on $\Xi \times \Theta \times \mathcal{Y}$ by

$$h(\xi, \theta, y) \propto p(\theta)p(y|\theta, \xi)U(\theta, \xi, y). \tag{53}$$

The marginal distribution for ξ is then

$$h(\xi) \propto \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[U(\theta, \xi, y)], \tag{54}$$

this guarantees that high probability regions for ξ correspond to regions with a large utility. The core approach, then, is to sample from the joint density $h(\xi, \theta, y)$ using a technique such as MCMC—Clyde et al. (1996) used the Metropolis–Hastings algorithm (Hastings, 1970). MCMC on $h(\theta, \xi, y)$ was also used by Bielza et al. (1999); Müller (2005). Cook et al. (2008); Drovandi and Pettitt (2013) used the MCMC technique, and fitted a density estimator to the MCMC samples to improve their estimation of the optimal

design. Ryan et al. (2014) applied MCMC in combination with dimensionality reduction on the latent space to avoid problems with MCMC in higher dimensions.

An extension of this idea, inspired by simulated annealing (Van Laarhoven and Aarts, 1987), is to include the utility contributions from J independent (θ, y) pairs, to create an unnormalised density on $\Xi \times \Theta^J \times \mathcal{Y}^J$

$$h_J(\xi, \theta_{1:J}, y_{1:J}) = \prod_{j=1}^{J} p(\theta_j) p(y_j|\theta_j, \xi) U(\theta_j, \xi, y_j). \tag{55}$$

One can see that for larger J, the probability mass concentrates more strongly around the optimal ξ. The simulated annealing mechanism is applied by *increasing J during the course of optimisation.* This approach has been applied by Müller et al. (2004); Müller (2005); Stroud et al. (2001); Cook et al. (2008).

Alternatively, one can sample $h_J(\xi, \theta_{1:J}, y_{1:J})$ using Sequential Monte Carlo (SMC) (Doucet et al., 2000). This was the approach taken by Amzal et al. (2006); Kuck et al. (2006).

Evolutionary algorithms Another approach to solving equation (40) is to optimise over the design space using evolutionary algorithms (Eiben et al., 2003). Hamada et al. (2001) applied genetic algorithms to this problem, and Price et al. (2018) proposed the Induced Natural Selection Heuristic (INSH) method to optimise the design.

Gradient-based optimisation Whilst gradient-driven optimisation methods are commonplace in optimisation theory, their adoption in Bayesian experimental design has been limited. This is likely because, as with the EIG itself, estimating the gradient $\nabla_\xi \mathcal{I}$ is a computationally challenging problem, and it is rare that we can place a guarantee of accuracy on EIG gradient estimates. Standard stepwise optimisation methods are relatively data hungry, requiring the estimate of gradients at many design points. Given these challenges, approaches such as Bayesian optimisation, which emphasises optimisation with limited, expensive evaluations of the underlying objective function, predominate.

Nevertheless, Huan and Marzouk (2014) considered gradient-based methods for solving the EIG optimisation problem. They considered the Robbins–Monroe stochastic gradient descent (SGD) (Robbins and Monro, 1951) algorithm applied to the NMC estimator of the EIG, equation (47), resampling θ_n, y_n and θ'_m at each iteration. This leads to a gradient descent method with noisy and biased gradients. They also considered applying the SAA-BFGS algorithm (Fletcher, 2013) to the NMC estimator, without resampling at each iteration. Carlon et al. (2020) considered gradient optimisation of both NMC and Laplace estimators of the EIG using SGD.

Note: At the time of writing Chapter 3, we were not aware of the work of Huan and Marzouk (2014). This is an important piece of prior work for this chapter, which also deals with the stochastic gradient optimisation of the EIG. We regret the omission. The key distinctions between Chapter 3 and Huan and Marzouk (2014) are 1) the use of EIG lower bounds as surrogate differentiable objectives by the former as compared to the NMC surrogate used by the latter, 2) the simultaneous optimisation of a variational parameter to produce more accurate estimates of $\nabla_\xi \mathcal{I}$ of the former.

Other methods Huan and Marzouk (2013) proposed the Nelder–Mead simplex method (Nelder and Mead, 1965), a gradient-free optimisation algorithm, and simultaneous perturbation stochastic approximation (Spall, 1998) as two alternative optimisation algorithms for Bayesian experimental designs.

5 Bayesian Active Learning

Active learning allows a learning algorithm to "choose the data from which it learns" (Settles, 2009). In the Bayesian setting, the learning algorithm is a Bayesian model. In its most abstract form, then, Bayesian Active Learning is identical to Bayesian Experimental Design, but with different vocabulary: designs ξ are referred to as queries, observations y are referred to as labels and are often provided by a human labeller, the

Algorithm 1 Pool-based Bayesian active learning with greedy acquisition

Require: Acquisition function α, prior $p(\theta)$ on model weights, pool Ξ, initial dataset $\mathcal{D}_0$ may be empty.
for step $t = 1, \ldots, T$ **do**
  Find $\xi_t = \arg\max_{\xi \in \Xi} \alpha(\xi; \mathcal{D}_{t-1})$ by scoring each unlabelled element of the pool
  Obtain label y_t for query ξ_t
  Set $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(\xi_t, y_t)\}$ and retrain model to compute $p(\theta|\mathcal{D}_t)$
end for

design criterion is referred to as the acquisition function. Queries are selected to maximise the acquisition function, typically in an iterative process.

Pool-based active learning However, this abstract similarity disguises the common differences in applications of active learning and experimental design. One hugely important sub-field of active learning, including Bayesian active learning, is *pool-based active learning* (Lewis and Gale, 1994). Here, the design space Ξ consists of unlabelled examples (such as images or sentences), the observation y is a human-provided label that corresponds to the unlabelled instance ξ, and the model is a classifier with parameters θ that predicts y from ξ. Pool-based active learning also applies less commonly to regression problems, for which y is a continuous label.

Sequential active learning with greedy acquisition In Sections 3 and 4, we focused on one-step design in which we begin with a prior $p(\theta)$, select a design ξ, obtain outcome y, and the experiment terminates. In active learning, we rarely want to acquire just one label or one batch of labels—the true power of the framework is apparent in a sequential setting (Lewis and Gale, 1994). This means that we pick design ξ_1 obtaining label y_1, then choose ξ_2 and receive label y_2, and so on. The dataset that we have after t experiments is $\mathcal{D}_t = \{(\xi_1, y_1), \ldots, (\xi_t, y_t)\}$. A simple approach to the sequential problem that is adopted in almost all of active learning (Gal et al., 2017) is *greedy acquisition*. In short, this strategy picks the next design to maximise the utility of the next label, without any consideration of how this will affect future queries.

However, it is still essential to incorporate all existing data $\mathcal{D}_t$ into the model before making this choice. To do this, we use the posterior[3] given existing data $p(\theta|\mathcal{D}_t)$ in place of the original prior $p(\theta)$. For the EIG, for example, at each step we would choose the design that maximises

$$\mathcal{I}(\xi; \mathcal{D}_t) = \mathbb{E}_{p(y|\xi, \mathcal{D}_t)} \left[\mathbb{E}_{p(\theta|\xi, y, \mathcal{D}_t)}[\log p(\theta|\xi, y, \mathcal{D}_t)] - \mathbb{E}_{p(\theta|\mathcal{D}_t)}[\log p(\theta|\mathcal{D}_t)] \right] \tag{56}$$

where $p(y|\xi, \mathcal{D}_t) = \mathbb{E}_{p(\theta|\mathcal{D}_t)}[p(y|\theta, \xi)]$. The high-level framework of greedy, sequential pool-based Bayesian active learning with a general acquisition function α is summarised in Algorithm 1. We discuss the theory of sequential experimentation in more detail in Sec. 8.

5.1 Acquisition functions

5.1.1 Bayesian Active Learning by Disagreement

A key point of intersection between Bayesian active learning and Bayesian experimental design is the Bayesian Active Learning by Disagreement (BALD) score (Houlsby et al., 2011), a widely adopted acquisition function within Bayesian Active Learning.

Proposition 9 (Houlsby et al. (2011))**.** *The BALD score is equivalent to the EIG.*

Proof. The BALD score is the mutual information between θ and y, but typically rearranged as

$$\alpha_{\text{BALD}}(\xi; \mathcal{D}_t) = \mathbb{E}_{p(\theta|\mathcal{D}_t)} \left[H[p(y|\xi, \mathcal{D}_t)] - H[p(y|\xi, \theta, \mathcal{D}_t)] \right]. \tag{57}$$

[3]In active learning, we make the assumption that θ represents the full set of model parameters (see Sec. 6).

We have

$$= \mathbb{E}_{p(\theta|\mathcal{D}_t)} \left[-\mathbb{E}_{p(y|\xi,\mathcal{D}_t)}[\log p(y|\xi,\mathcal{D}_t)] + \mathbb{E}_{p(y|\xi,\theta,\mathcal{D}_t)}[\log p(y|\xi,\theta,\mathcal{D}_t)] \right] \tag{58}$$

$$= \mathbb{E}_{p(\theta|\mathcal{D}_t)p(y|\xi,\theta,\mathcal{D}_t)} \left[-\log p(y|\xi,\mathcal{D}_t) + \log p(y|\xi,\theta,\mathcal{D}_t) \right] \tag{59}$$

$$= \mathbb{E}_{p(\theta|\mathcal{D}_t)p(y|\xi,\theta,\mathcal{D}_t)} \left[\log \frac{p(y|\xi,\theta,\mathcal{D}_t)}{p(y|\xi,\mathcal{D}_t)} \right] \tag{60}$$

applying Bayes Theorem gives

$$= \mathbb{E}_{p(\theta|\mathcal{D}_t)p(y|\xi,\theta,\mathcal{D}_t)} \left[\log \frac{p(\theta|\xi,y,\mathcal{D}_t)}{p(\theta|\mathcal{D}_t)} \right] \tag{61}$$

$$= \mathbb{E}_{p(y|\xi,\mathcal{D}_t)} \left[\mathbb{E}_{p(\theta|\xi,y,\mathcal{D}_t)}[\log p(\theta|\xi,y,\mathcal{D}_t)] - \mathbb{E}_{p(\theta|\mathcal{D}_t)}[\log p(\theta|\mathcal{D}_t)] \right] = \mathcal{I}(\xi;\mathcal{D}_t). \tag{62}$$

Note this is essentially the same proof as Proposition 6. □

The BALD score can be utilised directly in Algorithm 1. One important feature of writing EIG in BALD form is that it only depends on the actual experimental observation y, and does not require a probability density on θ. This can be important if we do not have a closed form density for θ either in the prior $p(\theta)$ or in the posterior $p(\theta|\mathcal{D}_t)$. This is particularly useful in active learning, where we may consider particularly complex models with high-dimensional θ.

In Deep Bayesian Active Learning (Gal et al., 2017), for instance, the model that predicts y from ξ is a neural network with parameters θ. In order to treat this model in a Bayesian manner, methods for Bayesian deep learning must be utilised. Gal et al. (2017) specifically used Dropout as a way of estimating prior and posterior distributions on θ (Gal and Ghahramani, 2016). Here, fitting $p(\theta|\mathcal{D}_t)$ amounts to retraining the network with Dropout. Beluch et al. (2018) and Pop and Fulop (2018) used a simple ensemble of models, treating different members of the ensemble as posterior samples of θ. To fit $p(\theta|\mathcal{D}_t)$, each deterministic model in the ensemble is retrained separately.

A key computational insight when estimating $\mathcal{I}(\xi)$ for a classification model in which the observation space $\mathcal{Y}$ is finite was made by Houlsby et al. (2011); Gal et al. (2017). We have

$$\mathcal{I}(\xi) = \sum_{y \in \mathcal{Y}} \mathbb{E}_{p(\theta)} \left[p(y|\theta,\xi) \log \frac{p(y|\theta,\xi)}{\mathbb{E}_{p(\theta)}[p(y|\theta,\xi)]} \right] \tag{63}$$

which can simply be estimated with Monte Carlo using samples $\theta_1, \ldots, \theta_N \sim p(\theta)$. The same idea applied when we have $p(\theta|\mathcal{D}_t)$ in place of $p(\theta)$. This estimator was also used by Vincent and Rainforth (2017), who observed that, unlike the NMC estimator of equation (47), this estimator converges at the standard Monte Carlo rate with error $\mathcal{O}(N^{-1/2})$. This speed-up is a consequence of being able to sum over $\mathcal{Y}$.

BatchBALD In the pool-based active learning setting with a discrete pool of size p, each acquisition involves computing the BALD score for every element of the pool and choosing the best one (Algorithm 1), which is an $\mathcal{O}(p)$ operation. Kirsch et al. (2019) considered the problem of batch active learning, in which designs are k-subsets of the pool. This means that, at each iteration of active learning, k different unlabelled examples will be selected and labelled. Naively scoring each k-subset of the unlabelled pool costs $\binom{p}{k}$, which rapidly becomes prohibitive. BatchBALD instead creates the design by greedily adding elements from the pool one at a time, giving a more efficiently scalable algorithm. This approach can be justified theoretically using the notion of submodularity—see Sec. 8.1.1.

5.1.2 Other acquisition functions

Within the Bayesian active learning framework, a range of other acquisition functions and computational methods have been proposed. It is possible to extend most common *non-Bayesian* acquisition functions

for use with Bayesian models. These non-Bayesian acquisition rules are generally a function of the predictive distribution $p(y|\xi, \mathcal{D}_t)$. When using a Bayesian model we can use the Bayesian marginal (posterior predictive) $p(y|\xi, \mathcal{D}_t) = \mathbb{E}_{p(\theta|\mathcal{D}_t)}[p(y|\theta, \xi)]$ in place of the deterministic predictive distribution that arises in non-Bayesian models. Standard acquisition functions such as uncertainty sampling (Lewis and Gale, 1994), margin sampling (Scheffer et al., 2001), and variation ratios (Freeman, 1965) can be therefore be employed in this context. Of particular note is the maximum entropy sampling method (Shannon, 1948; Settles and Craven, 2008), which uses

$$\alpha_{\text{Entropy}}(\xi; \mathcal{D}_t) = H[p(y|\xi, \mathcal{D}_t)]. \tag{64}$$

As shown in Proposition 6, this approach is equivalent to EIG maximisation when the entropy $H[p(y|\theta, \xi)]$ does not depend on ξ. This can be interpreted as saying that, given the correct model, the level of noise is uniform across all examples in the pool Ξ. For instance, we could assume that every example has a true label that a human will assign with 100% accuracy. However, maximum entropy sampling (and, in general, rules based on uncertainty in the predictive distribution $p(y|\xi, \mathcal{D}_t)$) break down when there are designs ξ which are very ambiguous, e.g. the correct label is missing from the taxonomy. Maximum entropy and related acquisition rules can become fixated on ambiguous queries.

Active learning has also considered Bayesian-specific acquisition functions. Kendall et al. (2015) proposed the mean standard deviation (Mean STD) acquisition rule for classification models. Define $\sigma_y(\xi; \mathcal{D}_t)$ as the standard deviation over $\theta|\mathcal{D}_t$ of the probability of example ξ being assigned to class y, i.e.

$$\sigma_y(\xi; \mathcal{D}_t) = \sqrt{\text{Var}_{p(\theta|\mathcal{D}_t)}[p(y|\theta, \xi)]}, \tag{65}$$

then the MeanSTD acquisition function is,

$$\alpha_{\text{MeanSTD}}(\xi; \mathcal{D}_t) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} \sigma_y(\xi; \mathcal{D}_t). \tag{66}$$

Following Bayesian decision theory, Roy and McCallum (2001) considered minimising the Bayes posterior risk, focusing on log loss and 0/1 loss. Kapoor et al. (2007) considered a range of Bayesian acquisition functions for binary classification, focusing on a score which combines the mean and variance of the prediction. Yang et al. (2012) applied Bayesian active learning to metric learning, and used an acquisition function based on maximum entropy.

6 Embedded models

So far, we have assumed that θ, the parameters of interest, and θ, the full set of model parameters, are one and the same. In this section, we explore the case in which the parameters of interest and the full set of model parameters are different. Model selection (Vanlier et al., 2014; Drovandi et al., 2014), in which we are only interesting in deciding which model is correct and not interested in learning the exact model parameters, is one important example of this setting. Bayesian optimisation (Sec. 7) is also an example in which we have a probability model for an unknown function, but we are only interest in learning the location of the maximum of that function.

For this more general case, we assume that the model is fully specified by a set of parameters ψ, and that our parameters of interest θ are a function of the full parameter set $\theta = f_\theta(\psi)$. In this case, the full joint distribution of the model is $p(\psi, y|\xi)$, and we obtain a joint over $\Theta \times \mathcal{Y}$ by integrating

$$p(\theta, y|\xi) = \int_{f_\theta^{-1}(\{\theta\})} p(\psi, y|\xi)\, d\psi. \tag{67}$$

Semi-implicit likelihood The embedded model setting allows us to extend our discussion of explicit and implicit models (Sec. 2.1). It could be the case that we do have an explicit prior for ψ and an explicit

likelihood $p(y|\psi,\xi)$ for the observation y given the full set of parameters ψ. Then the likelihood $p(y|\theta,\xi)$ is given by

$$p(y|\theta,\xi) = \int_{f_\theta^{-1}(\{\theta\})} p(y|\psi,\xi)\, d\psi, \tag{68}$$

and the prior is given by

$$p(\theta) = \int_{f_\theta^{-1}(\{\theta\})} p(\psi)\, d\psi. \tag{69}$$

First, note that $p(\theta, y|\xi) \neq p(\theta)p(y|\theta,\xi)$ in an embedded model. Second, computing one or both of these integrals may be an intractable computation. We use the term *semi-implicit* for this case in which the likelihood or prior for ψ is explicit, but the likelihood or prior for θ involves an intractable integral. So a semi-implicit likelihood is one which is formed as an integral of an explicit likelihood $p(y|\psi,\xi)$ and a semi-implicit prior is one which is an integral of explicit prior $p(\psi)$.

Exchangeability In an exchangeable embedded model, it is no longer true that different experiments are independent *conditional on* θ. Intuitively, the reason for this is that one experiment gives us information about *all of* ψ. Without extra assumptions, information from the first experiment tells us something about ψ even when we condition on θ, and this influences the predictive distribution for the second experiment. More formally, the factorisation equation (3) must be replaced by a factorisation conditional on ψ, and the natural assumption to make is that experiments are independent conditional on ψ

$$p(\psi, y_{1:T}|\xi_{1:T}) = p(\psi)\prod_{t=1}^{T} p(y_t|\psi,\xi_t). \tag{70}$$

Sequential learning with greedy acquisition One of the consequences of equation (70) is that Algorithm 1 is not quite correct for an embedded model. Specifically, between iterations, it is necessary to update the full model on ψ by fitting $p(\psi|\mathcal{D}_t)$, it is not enough to update beliefs about θ.

6.1 Expected Information Gain for embedded models

The EIG can naturally extend to the case of embedded models. The definition of information gain on the parameter of interest θ remains the same: $U_{\mathcal{I}}(\xi, y) = \mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|y,\xi)] - \mathbb{E}_{p(\theta)}[\log p(\theta)]$. When we take the expectation over y, however, we use the Bayesian marginal that integrates over all of ψ, i.e. $p(y|\xi) = \mathbb{E}_{p(\psi)}[p(y|\psi,\xi)]$, to give

$$I(\xi) = \mathbb{E}_{p(\psi)p(\theta|\psi)p(y|\psi,\xi)}\left[\log \frac{p(\theta|y,\xi)}{p(\theta)}\right]. \tag{71}$$

where $p(\theta|\psi)$ is a delta function on $f_\theta(\psi)$. This is different to the definition in equation (17) because the expectation is taken over $p(\theta, y|\xi) \neq p(\theta)p(y|\theta,\xi)$ for an embedded model. The EIG in an embedded model can also be expressed in BALD form (Proposition 9) as

$$I(\xi) = H[p(y|\xi)] - \mathbb{E}_{p(\theta)}[H[p(y|\theta,\xi)]] \tag{72}$$

$$= H\left[\mathbb{E}_{p(\psi)}[p(y|\psi,\xi)]\right] - \mathbb{E}_{p(\psi)p(\theta|\psi)}[H[p(y|\theta,\xi)]], \tag{73}$$

where the second line emphasises the difference with the standard case.

6.2 Computational methods for semi-implicit likelihood models

The NMC estimator of the EIG (Ryan, 2003) can be extended to the semi-implicit case. The central idea is that we form a Monte Carlo estimator of both $p(y|\theta,\xi)$ and $p(y|\xi)$ using appropriate Monte Carlo integrals over ψ. As in the standard NMC estimator, we have

$$p(y|\xi) = \mathbb{E}_{p(\psi)}[p(y|\psi,\xi)] \approx \frac{1}{M}\sum_{m=1}^{M} p(y|\psi_m,\xi) \text{ where } \psi_1,\ldots,\psi_M \overset{\text{i.i.d.}}{\sim} p(\psi). \tag{74}$$

For $p(y|\theta,\xi)$, we need access to samples from the distribution $p(\psi|\theta)$. Then,

$$p(y|\theta,\xi) = \mathbb{E}_{p(\psi|\theta)}[p(y|\psi,\xi)] \approx \frac{1}{M}\sum_{m=1}^{M} p(y|\psi_m,\xi) \text{ where } \psi_1,\ldots,\psi_M \overset{\text{i.i.d.}}{\sim} p(\psi|\theta). \tag{75}$$

Combining, we have the semi-implicit NMC estimator of EIG

$$\hat{I}_{\text{SI-NMC}}(\xi) = \frac{1}{N}\sum_{n=1}^{N}\left[\log\left(\frac{1}{M}\sum_{m=1}^{M} p(y_n|\psi_{nm},\xi)\right) - \log\left(\frac{1}{M}\sum_{m=1}^{M} p(y_n|\psi_m,\xi)\right)\right] \tag{76}$$

where $\theta_n, y_n \overset{\text{i.i.d.}}{\sim} p(\theta,y|\xi)$, $\psi_m \overset{\text{i.i.d.}}{\sim} p(\psi)$ and $\psi_{nm} \overset{\text{i.i.d.}}{\sim} p(\psi|\theta_n)$.

Ma et al. (2018) considered information acquisition for imputation in a semi-implicit setting. They used a Partial VAE which facilitated estimation of the EIG using the conditional independence assumptions of the model. Extending this, Gong et al. (2019) considered a similar active imputation scenario. They used $\hat{I}_{\text{SI-NMC}}(\xi)$ to estimate an information criterion for experimental design. In their probabilistic model, they had $\psi = (\theta, z)$ with $p(\psi) = p(\theta)p(z)$. When conditioning on data $\mathcal{D}_t$, they used approximate inference in which the independence of θ and z was maintained. Under these conditions, sampling $p(\psi|\theta)$ amounted to fixing θ and taking new, independent samples of z. A simplified form of their estimator is

$$\hat{I}_{\text{Icebreaker}}(\xi) = \frac{1}{N}\sum_{n=1}^{N}\left[\log\left(\frac{1}{M}\sum_{m=1}^{M} p(y_n|\theta_n,z_m,\xi)\right) - \log\left(\frac{1}{ML}\sum_{m=1}^{M}\sum_{\ell=1}^{L} p(y_n|\theta_\ell,z_m,\xi)\right)\right] \tag{77}$$

where $\theta_n, y_n \overset{\text{i.i.d.}}{\sim} p(\theta,y|\xi)$, $z_m \overset{\text{i.i.d.}}{\sim} p(z)$ and $\theta'_\ell \overset{\text{i.i.d.}}{\sim} p(\theta)$. Overstall and Woods (2017) considered almost the same setting when estimating the EIG utility. Specifically, they considered a semi-implicit case in which ψ can be partitioned into parameters of interest and *independent* nuisance parameters, and used this semi-implicit NMC estimator for the EIG.

7 Bayesian Optimisation

Bayesian optimisation (BO) (Snoek et al., 2012; Shahriari et al., 2015) considers the problem of finding the maximiser of an unknown objective function

$$\xi^* = \arg\max_{\xi\in\Xi} f(\xi). \tag{78}$$

To deal with the unknown function in a Bayesian manner, we consider a statistical model for f with prior $p(f)$. We assume that we can obtain relatively expensive measurements from the true function f at design points ξ. These measurements may be corrupted by noise, meaning that we obtain observations

$$y|\xi, f \sim p(y|f(\xi)), \tag{79}$$

for example, $y = f(\xi) + \varepsilon$ for $\varepsilon \sim N(0,\sigma^2)$.

BO can naturally be cast within the framework of Bayesian experimental design. We have designs ξ and observations y connected by the Bayesian model on f and the noise model. The missing piece is to specify the parameter of interest θ. The parameter of interest is not f, because Bayesian optimisation is explicitly concerned with *maximising* f, meaning that any information about f in regions where it is well below its maximum is not useful. The most common formulation is to take to be the location of the maximiser of f (Hernández-Lobato et al., 2014), i.e. $\theta = \arg\max_{\xi\in\Xi} f(\xi)$. The fact that θ is not all of f means that BO is not an explicit likelihood (Sec. 2.1) experimental design problem, nor does it fit into the framework of Bayesian active learning (Sec. 5). BO is experimental design for an embedded model (Sec. 6), with the function f playing the role of the richer parameter set ψ. We will see that BO has its own character with a wide range of algorithms that apply specifically to the optimisation problem.

Algorithm 2 Bayesian Optimisation (Shahriari et al., 2015)

Require: Acquisition function α, prior $p(f)$ on function, design space Ξ, initial dataset $\mathcal{D}_0$ may be empty.
  for step $t = 1, \ldots, T$ **do**
    Find $\xi_t = \arg\max_{\xi \in \Xi} \alpha(\xi; \mathcal{D}_{t-1})$
    Obtain noisy measurement $y_t \sim p(y|f(\xi_t))$ at design ξ_t
    Set $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(\xi_t, y_t)\}$ and retrain the model to compute $p(f|\mathcal{D}_t)$
  end for
  Use $p(f|\mathcal{D}_T)$ to estimate the maximiser of f.

To set up a BO system, we begin by specifying a Bayesian model for f with prior $p(f)$, and a measurement noise model $p(y|f(\xi))$. We then specify an acquisition function that guides our choice of designs at which we should take measurements. The acquisition function in BO plays the same role as the design criterion in Bayesian experimental design and the acquisition function in active learning—we select the design that maximises the acquisition function to obtain new measurements of f. As in active learning, BO typically adopts the *greedy acquisition* approach that was outlined in Sec. 5. The entire approach is summarised in Algorithm 2.

We begin by discussing common choices for the Bayesian model and acquisition function in BO. We focus specifically on the Entropy Search family of acquisition rules, highlighting the connection to experimental design with EIG.

7.1 Bayesian models for optimisation

7.1.1 Parametric models

When the design space Ξ is discrete, the function f can be characterised by a finite number of latent variables. This case is closely connected to theory of multi-armed bandits (Lai and Robbins, 1985): we can view each $\xi \in \Xi$ as an 'arm' of a bandit in a casino. Each arm has an unknown payout, and the aim is to identify the best arm. Our mathematical set-up specifically relates the pure exploration scenario (Bubeck et al., 2009), in which final knowledge of the location of the best arm is important, but function evaluations during the course of Algorithm 2 are not. Finite-dimensional models such as the Beta-Bernoulli (Shahriari et al., 2015) and the Gaussian (Hoffman et al., 2014) have been applied in the bandit context.

Both Bayesian linear and generalised linear models have been utilised within the Bayesian optimisation context (Russo and Van Roy, 2014; Shahriari et al., 2015). In the bandit context, these models are applied by associating each bandit arm with a feature vector $\mathbf{x}_\xi$, and assuming that the arm payout depends on this feature vector. For a linear model, for example, we would assume $f(\xi) = \langle \mathbf{x}_\xi, \mathbf{w} \rangle$. These models can also be applied to optimisation over continuous design spaces. Snoek et al. (2015) considered Bayesian optimisation using a Bayesian neural network as the model for f; they specifically took an 'adaptive basis regression' approach that is only Bayesian on the last layer of the network.

7.1.2 Nonparametric models

For continuous Bayesian optimisation, the Gaussian Process (GP) (Williams and Rasmussen, 2006) has proved an extremely popular Bayesian nonparametric model for the unknown function f (Osborne et al., 2009). The Gaussian process with a positive definite kernel k and mean function μ assumes the following multivariate Gaussian distribution for the finite-dimensional marginal distributions (Øksendal, 2003) of f

$$\begin{pmatrix} f(\xi_1) \\ \vdots \\ f(\xi_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\xi_1) \\ \vdots \\ \mu(\xi_n) \end{pmatrix}, \begin{pmatrix} k(\xi_1, \xi_1) & \ldots & k(\xi_1, \xi_n) \\ \vdots & & \vdots \\ k(\xi_n, \xi_1) & \ldots & k(\xi_n, \xi_n) \end{pmatrix} \right). \tag{80}$$

Given a dataset of observations $\mathcal{D}_t = \{(\xi_i, y_i)\}_{i=1}^t$, the resulting posterior on f is also a Gaussian process. The mean and covariance structure of the posterior can be derived by computing the conditional form of

equation (80), however, the necessary matrix computations come at cubic cost $\mathcal{O}(t^3)$; we refer to Williams and Rasmussen (2006) for full details. As a mark of its popularity, BO with a GP model for f has been implemented in several software frameworks, such as BoTorch (Balandat et al., 2020).

Within Bayesian optimisation, several extensions of the GP have also been considered as models for f. Different variants of *sparse* GPs have been proposed (Quinonero-Candela and Rasmussen, 2005; Snelson and Ghahramani, 2006; Lázaro-Gredilla et al., 2010), aiming to reduce the computational burden of using the standard GP conditioning formula. Calandra et al. (2016) combined GPs with feature learning to propose the Manifold GP.

Beyond the GP family, Hutter et al. (2013) also considered a random forest model for f, but found GPs to be preferable. Focusing on the application of hyperparameter optimisation, Bergstra et al. (2011) proposed the Tree-structured Parzen Estimator model for f that combines a tree-structured hierarchy with mixture modelling. Finally, Neiswanger et al. (2019) considered Bayesian optimisation in which an arbitrary probabilistic program is used as the model for f.

7.2 Acquisition functions

7.2.1 The Entropy Search family

To define an information-theoretic acquisition function for Bayesian optimisation, we want to gain information about the random variable $\theta = \arg\max_{\xi \in \Xi} f(\xi)$. To this end, Villemonteix et al. (2009) proposed Stepwise Uncertainty Reduction (SUR). This method aims to reduce posterior entropy in θ using the acquisition rule

$$\alpha_{\mathrm{SUR}}(\xi; \mathcal{D}_t) := -\mathbb{E}_{p(y|\xi,\mathcal{D}_t)}[H[p(\theta|\mathcal{D}_t \cup \{(\xi, y)\})]]. \tag{81}$$

where $p(y|\xi, \mathcal{D}_t) = \mathbb{E}_{p(f|\mathcal{D}_t)}[p(y|f(\xi))]$. In practice, Villemonteix et al. (2009) estimated the acquisition function by discretising θ and using a GP model for f. Hennig and Schuler (2012) considered a closely related acquisition function called Entropy Search (ES) that maximises the KL-divergence between the posterior on θ and a base measure $b(\theta)$. This gives the acquisition function

$$\alpha_{\mathrm{ES}}(\xi; \mathcal{D}_t) := \mathbb{E}_{p(y|\xi,\mathcal{D}_t)}[\mathrm{KL}[p(\theta|\mathcal{D}_t \cup \{(\xi, y)\}) \| b(\theta)]]. \tag{82}$$

The following Proposition, due to MacKay (1992), shows that these in information measures are equivalent, and are equivalent to the EIG.

Proposition 10 (MacKay (1992))**.** *Consider the general experimental design set-up of Sec. 3. The following acquisition functions all give the same optimal design*

$$\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)}\left[\mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)]\right], \tag{83}$$

$$\mathcal{I}_2(\xi) = -\mathbb{E}_{p(y|\xi)}[H[p(\theta|\xi, y)]], \tag{84}$$

$$\mathcal{I}_3(\xi) = \mathbb{E}_{p(y|\xi)}[\mathrm{KL}[p(\theta|\xi, y) \| b(\theta)]] \tag{85}$$

where $p(y|\xi) = \mathbb{E}_{p(f)}[p(y|f(\xi))]$.

Proof. We have

$$\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)}\left[\mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)]\right] \tag{86}$$

$$= \mathbb{E}_{p(y|\xi)}[-H[p(\theta|\xi, y)] + H[p(\theta)]] \tag{87}$$

$$= \mathcal{I}_2(\xi) + H[p(\theta)]. \tag{88}$$

and

$$\mathcal{I}(\xi) = \mathbb{E}_{p(y|\xi)}\left[\mathbb{E}_{p(\theta|\xi,y)}[\log p(\theta|\xi, y)] - \mathbb{E}_{p(\theta)}[\log p(\theta)]\right] \tag{89}$$

$$= \mathbb{E}_{p(y|\xi)}\left[\mathbb{E}_{p(\theta|\xi,y)}\left[\log \frac{p(\theta|\xi, y)}{b(\theta)}\right] - \mathbb{E}_{p(\theta)}\left[\log \frac{p(\theta)}{b(\theta)}\right]\right] \tag{90}$$

$$= \mathcal{I}_3(\xi) - \mathrm{KL}[p(\theta) \| b(\theta)]. \tag{91}$$

Since $H[p(\theta)]$ and $\mathrm{KL}[p(\theta)\|b(\theta)]$ do not depend on ξ, choosing ξ to maximise EIG is equivalent to maximising $\mathcal{I}_2$ and $\mathcal{I}_3$. □

Corollary 11. *Stepwise Uncertainty Reduction (Villemonteix et al., 2009) and Entropy Search (Hennig and Schuler, 2012) are equivalent to EIG maximisation when* $\theta = \arg\max_{\xi\in\Xi} f(\xi)$.

Proof. Note that $\mathcal{I}_2 = \alpha_{\mathrm{SUR}}$ and $\mathcal{I}_3 = \alpha_{\mathrm{ES}}$ when we replace the prior $p(\theta)$ with the posterior $p(\theta|\mathcal{D}_t)$. Since the result holds for a general experimental design set-up, it specifically holds in the BO case when $\theta = \arg\max_{\xi\in\Xi} f(\xi)$. □

Hernández-Lobato et al. (2014) proposed Predictive Entropy Search (PES). Like previous methods, PES uses the EIG (accounting for the embedded model, as in Sec. 6.1) as their acquisition function

$$\alpha_{\mathrm{PES}}(\xi;\mathcal{D}_t) = \mathcal{I}(\xi;\mathcal{D}_t) = H[p(\theta|\mathcal{D}_t)] - \mathbb{E}_{p(y|\xi,\mathcal{D}_t)}[H[p(\theta|\mathcal{D}_t \cup \{(\xi,y)\})]]. \tag{92}$$

However, the authors utilise the same insight as Houlsby et al. (2011) to write EIG in the equivalent form (see Proposition 9)

$$\mathcal{I}(\xi;\mathcal{D}_t) = H[p(y|\xi,\mathcal{D}_t)] - \mathbb{E}_{p(\theta|\mathcal{D}_t)}[H[p(y|\xi,\mathcal{D}_t,\theta)]]. \tag{93}$$

Within a GP model, the first term can be computed analytically, whilst the second is approximated by drawing samples of $\theta|\mathcal{D}_t$ and estimating $H[p(y|\xi,\mathcal{D}_t,\theta)]$ using expectation propagation (Minka, 2001). PES can be extended to batch acquisition in which we query f at multiple locations simultaneously on each iteration (Shah and Ghahramani, 2015).

In Maximum Entropy Search (MES) (Wang and Jegelka, 2017), the authors approach the problem differently. Instead of focusing on the latent variable of interest $\theta = \arg\max_{\xi\in\Xi} f(\xi)$, they instead formulate the problem with variable of interest $\theta_m = \max_{\xi\in\Xi} f(\xi)$. Here, θ_m is a one-dimensional random variable that represents the maximum *value* of the function f, rather than its arg max. The objective function for MES is then the EIG between a new observation y at design ξ and their parameter of interest θ_m

$$\alpha_{\mathrm{MES}}(\xi;\mathcal{D}_t) = H[p(y|\xi,\mathcal{D}_t)] - \mathbb{E}_{p(\theta_m|\mathcal{D}_t)}[H[p(y|\xi,\mathcal{D}_t,\theta_m)]]. \tag{94}$$

The MES objective may be easier to compute than PES with a GP model for f because θ_m is always one dimensional.

Computationally, a distinctive feature of BO with a GP model that sets it apart from computing the EIG in standard models (Sec. 4) is that many calculations can be performed analytically for the GP. For example, $H[p(y|\xi,\mathcal{D}_t)]$ is computed analytically in the PES acquisition function—this calculation would be intractable in a general model.

7.2.2 Other acquisition functions

Probability of improvement Perhaps the simplest acquisition rule, probability of improvement (Kushner, 1964) computes the probability that $f(\xi)$ is greater than some threshold τ

$$\alpha_{\mathrm{PI}}(\xi;\mathcal{D}_t) := \mathbb{P}(f(\xi) < \tau|\mathcal{D}_t). \tag{95}$$

Typically, the threshold τ is chosen adaptively to be the best objective value seen so far: $\tau_t = \max\{y_1,\ldots,y_t\}$.

Expected improvement A related acquisition rule is expected improvement (Mockus et al., 1978). This incorporates the amount by which the function value can be expected to increase at the location ξ, giving

$$\alpha_{\mathrm{EI}}(\xi;\mathcal{D}_t) := \mathbb{E}((f(\xi)-\tau)_+|\mathcal{D}_t) \tag{96}$$

where $x_+ = \max(0,x)$.

Upper confidence bound Starting with theoretical work on multi-armed bandits (Lai and Robbins, 1985), upper confidence bound (UCB) acquisition rules have been popular. Srinivas et al. (2009) explicitly considered the application of UCB functions with a GP model for BO. In its most general form, we let $q_p(\cdot)$ denote to the p-quantile of a univariate distribution. Then the UCB-p acquisition function is

$$\alpha_{\text{UCB-}p}(\xi; \mathcal{D}_t) := q_p(f(\xi)|\mathcal{D}_t). \tag{97}$$

In the Gaussian case, $f(\xi)|\mathcal{D}_t \sim N(\mu(\xi|\mathcal{D}_t), \sigma(\xi|\mathcal{D}_t)^2)$, an equivalent parametrisation of the UCB acquisition function is

$$\alpha_{\text{UCB-}p}(\xi; \mathcal{D}_t) = \mu(\xi|\mathcal{D}_t) + \beta_p \sigma(\xi|\mathcal{D}_t) \tag{98}$$

where β_p is the p-quantile of the standard Normal distribution.

Thompson sampling Thompson (1933) proposed a *stochastic* acquisition rule for Bayesian optimisation. Given a sample of the functional posterior $f_t \sim p(f|\mathcal{D}_t)$, Thompson Sampling chooses the maximiser of this sample as the next sampling location. This amounts to using the acquisition function

$$\alpha_{\text{TS}}(\xi; \mathcal{D}_t) = f_t(\xi) \text{ where } f_t \sim p(f|\mathcal{D}_t). \tag{99}$$

Hernández-Lobato et al. (2014) showed how the optimisation of a sample from the GP posterior can be approximately calculated.

8 Sequential Bayesian Experimental Design

We now lay out the theory of sequential experimentation more formally. Extending the basic sequential framework that we described for active learning in Sec. 5, we suppose that we have a sequence of T experiments. For each experiment, we pick ξ_t *adaptively* using the data that has already been observed[4] $\mathcal{D}_{t-1} = (\xi_1, y_1), \ldots, (\xi_{t-1}, y_{t-1})$. Given this design, we conduct an experiment using ξ_t and obtain outcome y_t. After each step of the experiment, our beliefs about θ are summarised by the posterior $p(\theta|\mathcal{D}_t)$, which is calculated as in Sec. 2.2. For an embedded model (Sec. 6), we would update our beliefs on the extended parameters ψ. For simplicity in this section, we assume we are not in an embedded model, unless otherwise stated, so the parameter θ is a full description of the model.

Policies and objectives The design ξ_t must be chosen on the basis of $\mathcal{D}_{t-1}$. A general abstraction to describe this is to introduce a *stochastic policy* $\pi(\xi|\mathcal{D}_{t-1})$ that maps from $\mathcal{D}_{t-1}$ to a distribution over designs. A special case of this is a deterministic policy, for which ξ_t is simply a function of $\mathcal{D}_{t-1}$.

In the sequential setting, it no longer makes sense to talk of the optimality of individual designs. Indeed, we cannot say whether a design ξ_2 will be optimal until we have observed the outcome y_1. Instead, we can describe optimality in terms of the *policy*—the policy which makes the best decision for ξ_2 for every possible value of y_1 would be an optimal policy.

Optimality also requires a criterion, so we must extend the utility-based approach of Sec. 3 based on Lindley (1972) to the sequential setting. Perhaps the most natural extension of Lindley's original theory, which is used implicitly by Huan and Marzouk (2016); Foster et al. (2021) is to consider a final utility, or reward, which is obtained after all data has been collected. In this *terminal reward* framework, we assume that we have a utility function $U(\theta, \mathcal{D}_T)$. Once we have collected all our data, we have expected utility $\mathbb{E}_{p(\theta|\mathcal{D}_T)}[U(\theta, \mathcal{D}_T)]$. The optimal policy, therefore, is the natural counterpart to equation (9), namely

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{p(\mathcal{D}_T|\pi)} \left[\mathbb{E}_{p(\theta|\mathcal{D}_T)}[U(\theta, \mathcal{D}_T)] \right] \tag{100}$$

[4]For an exchangeable model, the order of the data does not matter, so we could write $\mathcal{D}_{t-1} = \{(\xi_1, y_1), \ldots, (\xi_{t-1}, y_{t-1})\}$, which we implicitly assumed was the case in Sections 5 and 7. For a non-exchangeable model, we need to know the order that data was collected to conduct valid inference.

Algorithm 3 Terminal reward Sequential Bayesian Experimental Design

Require: Prior $p(\theta)$, model $p(y|\xi,\theta)$, initial data $\mathcal{D}_0$ may be empty.
  for step $t = 1, \ldots, T$ **do**
    Use policy to compute design $\xi_t \sim \pi(\xi|\mathcal{D}_{t-1})$
    Obtain experimental observation $y_t \sim p(y|\theta, \xi_t)$ with design ξ_t
    Set $\mathcal{D}_t = (\xi_1, y_1), \ldots, (\xi_t, y_t)$
  end for
  Obtain reward $U(\theta, \mathcal{D}_T)$

where $p(\mathcal{D}_T|\pi) = \mathbb{E}_{p(\theta)}\left[\prod_{t=1}^{T} \pi(\xi_t|\mathcal{D}_{t-1})p(y_t|\theta,\xi_t)\right]$ for an exchangeable model[5]. The whole sequential experiment process is described in Algorithm 3.

Sequential EIG The natural extension of EIG to the sequential setting is to let (Foster et al., 2021)

$$U_{\mathcal{I}}(\mathcal{D}_T) = H[p(\theta)] - H[p(\theta|\mathcal{D}_T)] \tag{101}$$

or equivalently (Huan and Marzouk, 2016)

$$U_{\mathrm{KL}}(\mathcal{D}_T) = \mathrm{KL}[p(\theta|\mathcal{D}_T)\|p(\theta)]. \tag{102}$$

The intuition behind this utility is to reduce our uncertainty in the value of θ from the sum total of all our experiments. It is in the sequential setting that the naturalness of using information-theoretic objectives for experimental design becomes most apparent.

Example 12 (Shannon (1948); Lindley (1956))**.** *Consider a model with $\theta = (L, R)$ where L and R are discrete random variables with independent uniform priors $\theta \sim \mathrm{Unif}(n_L) \times \mathrm{Unif}(n_R)$. Suppose we have two experimental designs at our disposal: ξ_L and ξ_R which produce noiseless outcomes giving the values of L and R respectively. Then, the utility of the sequence of experiments ξ_L, ξ_R is equal to sum of the utilities of the separate experiments ξ_L and ξ_R, i.e.*

$$U_{\mathcal{I}}((\xi_L, L), (\xi_R, R)) = U_{\mathcal{I}}((\xi_L, L)) + U_{\mathcal{I}}((\xi_R, R)). \tag{103}$$

Proof. Direct calculation using equation (101) gives

$$U_{\mathcal{I}}((\xi_L, L), (\xi_R, R)) = \log(n_L n_R) = \log n_L + \log n_R \tag{104}$$
$$U_{\mathcal{I}}((\xi_L, L)) = \log n_L \tag{105}$$
$$U_{\mathcal{I}}((\xi_R, R)) = \log n_R. \tag{106}$$

□

Shannon (1948) showed that this property (along with other technical requirements) can only be satisfied by utilities based on entropy, making the EIG arguably the most natural criterion for sequential Bayesian experimental design.

Static and batch policies One simple approximation to the optimal policy is to select all designs $\xi_1, \ldots, \xi_T$ before the start of the experiment. This is known as *static* design, also called *open-loop* design (DiStefano III et al., 2014). In effect, static design collapses the sequential design problem back into the one-step problem of Sec. 3, albeit with a larger design space

$$\Xi^T = \{(\xi_1, \ldots, \xi_T) : \xi_t \in \Xi \text{ for all } t\} \tag{107}$$

[5]For a non-exchangeable model, it would be $p(\mathcal{D}_T|\pi) = \mathbb{E}_{p(\theta)}\left[\prod_{t=1}^{T} \pi(\xi_t|\mathcal{D}_{t-1})p(y_t|\theta,\xi_t,\mathcal{D}_{t-1})\right]$

and corresponding observation space. The probabilistic model is also augmented as in Sec. 2.2. Unfortunately, static design may be arbitrarily worse than the performance of the best fully adaptive policy.

Rather than choosing all T designs upfront, we could instead choose design in batches of B. There can be practical benefits for choosing designs in batches (Lyu et al., 2019), as opposed to choosing them individually. Mathematically, this *batch* design procedure fits back into the sequential theory we have already laid out, with batches of designs being chosen from the new design space Ξ^B. Static design corresponds to the case $B = T$ in which we stop after one batch. Batch design in active learning is discussed on page 16.

8.1 Greedy design policies

Designing a policy to solve equation (100) can be challenging. One common approximate strategy is *greedy* design (also called *myopic* design). A greedy policy can be characterised as choosing each design ξ_t assuming that this is the final experiment—i.e. that once ξ_t has been chosen and y_t observed, the sequence of experiments will terminate. This means that the greedy policy will choose ξ_t to maximise

$$\xi_t^* = \arg\max_{\xi \in \Xi} \mathbb{E}_{p(y|\xi_t, \mathcal{D}_{t-1})} \left[\mathbb{E}_{p(\theta|\mathcal{D}_t)}[U(\theta, \mathcal{D}_t)] \right]. \tag{108}$$

where $p(y|\xi_t, \mathcal{D}_{t-1}) = \mathbb{E}_{p(\theta|\mathcal{D}_{t-1})}[p(y|\theta, \xi)]$. We can see that this amounts to solving the one-step design optimisation problem of equation (9) at each t, with the important distinction that we *replace the original prior $p(\theta)$ with the posterior given existing data $p(\theta|\mathcal{D}_{t-1})$*. This agrees exactly with the greedy acquisition strategy described in Section 5.

There is a subtle distinction when θ is embedded in a larger model with parameters ψ (Sec. 6)—we must update our beliefs about all the parameters to $p(\psi|\mathcal{D}_{t-1})$ and use the predictive distribution $p(y|\xi_t, \mathcal{D}_{t-1}) = \mathbb{E}_{p(\psi|\mathcal{D}_{t-1})}[p(y|\psi, \theta)]$ for y. This agrees with the greedy acquisition strategy of Section 7, where we update the full model on the unknown function f at each step.

The greedy (myopic) approach to experimental design is very widely adopted (Cavagnaro et al., 2010; Drovandi et al., 2014; McGree et al., 2012; Myung et al., 2013; Foster et al., 2019). As noted, it is also the typical sequential optimisation strategy in Bayesian active learning and Bayesian optimisation. One benefit of the greedy strategy is its simplicity—it effectively reduces the sequential experimental design problem to repeated applications of one-step design. It is typically observed that greedy optimisation for experimental design does not fail as catastrophically as greedy policies can do in general reinforcement learning tasks (Bakker et al., 2020). We explore a possible theoretical explanation for this phenomenon.

8.1.1 Submodularity

How much do we lose by using a greedy approach? A key theoretical tool for studying greedy policies is the notion of *submodularity* (Krause and Golovin, 2014). In short, if a utility function obeys submodularity (and a number of other conditions), then the theorem of Nemhauser et al. (1978) proves that a greedy strategy can achieve at least $(1 - 1/e) \approx 63\%$ of the best possible utility.

To precisely define submodularity, we must first define several other concepts. For any (finite) set V, the *power set* of V is $2^V = \{S : S \subseteq V\}$. A *set function* is any function $g : 2^V \to \mathbb{R}$. The *discrete derivative* of g is defined as

$$\Delta_g(e|S) = g(S \cup \{e\}) - g(S), \tag{109}$$

i.e. the extra value of adding element e to the set $S \subseteq V$. A set function $g : 2^V$ is *submodular* if, for every $A \subseteq B \subseteq V$ and for every $e \in V \setminus B$,

$$\Delta_g(e|B) \leq \Delta_g(e|A). \tag{110}$$

Intuitively, the value of adding e to the larger set B is smaller than the value of adding e to the smaller set A. Submodularity captures the intuitive notion of 'diminishing returns'. We also define a set function to be *monotone* if, for $A \subseteq B \subseteq V$, we have

$$g(A) \leq g(B). \tag{111}$$

The *greedy strategy* to maximise a monotone set function is to increment S by adding elements one at a time, following the rule

$$S_t = S_{t-1} \cup \{e_t\} \quad \text{where} \quad e_t = \arg\max_{e \in V} \Delta_g(e|S_{t-1}). \tag{112}$$

The following theorem of Nemhauser et al. (1978) shows that the greedy strategy performs near-optimally for submodular set functions.

Theorem 13 (Nemhauser et al. (1978))**.** *Let g be a monotone, submodular set function $g : 2^V \to \mathbb{R}$. Let $(S_t)_{t \geq 0}$ be obtained by the greedy strategy of equation (112). Then for any $t \leq |V|$ we have*

$$g(S_t) \geq (1 - 1/e) \max_{S \subseteq V, |S| = t} g(S). \tag{113}$$

This theorem proves that the greedy strategy can achieve at least $(1 - 1/e)$ of the best possible performance.

Submodularity for static experimental design There is a direct connection between the theory of submodularity and the *greedy construction of static experimental designs*. Indeed, the static experimental design problem is to choose $(\xi_1, \ldots, \xi_T)$ where each $\xi_t \in \Xi$. We assign a value to each static design following equation (100)

$$g(\xi_1, \ldots, \xi_T) = \mathbb{E}_{p(y_1, \ldots, y_T | \xi_1, \ldots, \xi_T)} \left[\mathbb{E}_{p(\theta | \mathcal{D}_T)} [U(\theta, \mathcal{D}_T)] \right]. \tag{114}$$

Suppose we could show that g is a monotone, submodular set function. Then the result of Theorem 13 would apply, meaning that we could construct a static design greedily by adding one element at a time.

To satisfy these conditions, we first need g to be invariant to the order of $\xi_1, \ldots, \xi_T$; we therefore assume that the model is exchangeable (Sec. 2.2). For the properties of submodularity and monotonicity, we need to choose a utility U, here we focus on the EIG with $U_{\mathcal{I}} = H[p(\theta)] - H[p(\theta|\mathcal{D}_T)]$. Then the function g becomes the mutual information between θ and $y_1, \ldots, y_T$ given $\xi_1, \ldots, \xi_T$.

Proposition 14 (Krause and Guestrin (2012))**.** *Suppose that, for any k and for designs $\xi_1, \ldots, \xi_k$, the random variables $y_1|\xi_1, \ldots, y_k|\xi_k$ are independent conditional on θ. Then the mutual information*

$$g(\{\xi_1, \ldots, \xi_k\}) = \mathbb{E}_{p(y_1, \ldots, y_k | \xi_1, \ldots, \xi_k)} \left[H[p(\theta)] - H[p(\theta|\mathcal{D}_k)] \right] \tag{115}$$

is a monontone, submodular set function.

The conditional independence assumption is equivalent to assuming an exchangeable model (Sec. 2.2) in which θ is the only model parameter (Sec. 6). Proposition 14 was also proved by Kirsch et al. (2019) in the context of BatchBALD for active learning.

8.1.2 Adaptive submodularity

The limitation of submodularity as a tool for analysing experimental design is that it does not consider *adaptive* design policies where the choice of a later design could be conditional on the outcome of earlier experiments. To address this limitation, Golovin and Krause (2011) introduced the notion of *adaptive submodularity*.

To define adaptive submodularity within our framework for experimental design, we focus on a discrete design space $|\Xi| < \infty$. We can then define the *conditional expected marginal benefit* of a design ξ as

$$\Delta(\xi|\mathcal{D}_t) = \mathbb{E}_{p(\theta|\mathcal{D}_t) p(y|\theta, \xi)} [U(\theta, \mathcal{D}_t \cup \{(\xi, y)\}) - U(\theta, \mathcal{D}_t)]. \tag{116}$$

The utility U is *adaptive monotone* with respect to model $p(\theta)p(y|\theta, \xi)$ if the conditional expected marginal benefit of all designs is positive. That is, for all $t \geq 0, \mathcal{D}_t$ and $\xi \notin \mathcal{D}_t$ we have

$$\Delta(\xi|\mathcal{D}_t) \geq 0. \tag{117}$$

Furthermore, the utility U is *adaptive submodular* with respect to model $p(\theta)p(y|\theta,\xi)$ if for all $s \leq t$ and all nested datasets $\mathcal{D}_s \subseteq \mathcal{D}_t$ and for all designs $\xi \notin \mathcal{D}_t$ we have

$$\Delta(\xi|\mathcal{D}_t) \leq \Delta(\xi|\mathcal{D}_s). \tag{118}$$

This is a natural generalisation of submodularity for set functions, and again it captures the principle of 'diminishing returns'. Golovin and Krause (2011) were able to generalise the result of Nemhauser et al. (1978) to the adaptive case for noiseless experiments in which $p(y|\theta,\xi)$ is deterministic.

Theorem 15 (Golovin and Krause (2011))**.** *Let π^{greedy} be the greedy policy of equation (108). Assume U is adaptive monotone and adaptive submodular for model $p(\theta)p(y|\theta,\xi)$. Then,*

$$\mathbb{E}_{p(\mathcal{D}_T|\pi^{greedy})}\left[\mathbb{E}_{p(\theta|\mathcal{D}_T)}[U(\theta,\mathcal{D}_T)]\right] \geq (1-1/e)\sup_{\pi}\mathbb{E}_{p(\mathcal{D}_T|\pi)}\left[\mathbb{E}_{p(\theta|\mathcal{D}_T)}[U(\theta,\mathcal{D}_T)]\right]. \tag{119}$$

Golovin et al. (2010) explored the applicability of this framework to Bayesian active learning and Bayesian experimental design, focusing on the noiseless case in which $p(y|\theta,\xi)$ is deterministic. They proved that the information gain utility $U_{\mathcal{I}}$ is adaptive monotone and adaptive submodular, so the result of Theorem 15 applies in this case.

A key results of Chen et al. (2015) did away with the noiseless assumption. Instead, they assumed that different experimental outcomes are independent conditional on θ. This matches exactly with the factorisation equation (3). They also assume that θ takes finitely many values $|\Theta| < \infty$. The key bound is as follows

Theorem 16 (Theorem 2 of Chen et al. (2015))**.** *Let π^{greedy} be the adaptive greedy experimental design policy. Assume that observations y are conditionally independent given θ. Then, for any $\delta > 0$*

$$\mathbb{E}_{p(\mathcal{D}_T|\pi^{greedy})}\left[U_{\mathcal{I}}(\mathcal{D}_T)\right] \geq \left(1-\exp\left[-\frac{1}{\gamma\max\{\log|\Theta|,\log(1/\delta)\}}\right]\right)\left(\sup_{\pi}\mathbb{E}_{p(\mathcal{D}_T|\pi)}\left[U_{\mathcal{I}}(\mathcal{D}_T)\right]-\delta\right) \tag{120}$$

where γ is a constant that depends on the noise distribution (see Chen et al. (2015)), and $U_{\mathcal{I}}$ is the information gain defined in equation (101).

Chen et al. (2017) went on to consider the case of noisy and correlated experimental outcomes (violating both the noiseless and the conditionally independent assumptions).

Finally, we note that the expected information gain is *not* adaptive submodular without assumption. This is elucidated by the following example, in which outcomes are not independent conditional on θ.

Example 17 (Inspired by Theorem 9 of Golovin et al. (2010))**.** *Consider a model with prior $\theta \sim \text{Unif}(\{-1,1\})$ and with $v \sim \text{Unif}(\{-1,1\})$. We have two potentially useful designs. ξ_v reports the value of v. $\xi_{\theta v}$ reports the value of θv. We also have M 'dummy' designs $\xi_1^d,\ldots,\xi_M^d$ which report nothing. Clearly, the optimal strategy to learn θ is to conduct experiments with ξ_v and $\xi_{\theta v}$ in any order, since $v\cdot\theta v = \theta v^2 = \theta$. However, if we analyse a one-step optimal greedy strategy, we observe that every design apart from $\xi_{\theta v}$ is independent of the value of θ, and hence has EIG 0. We can also verify that, without knowing v, the posterior on θ given the outcome of design $\xi_{\theta v}$ is still* $\text{Unif}(\{-1,1\})$*, hence the EIG of this design is also 0. Thus the greedy strategy will pick a design at random. If M is very large, the greedy strategy is likely to keep picking dummy designs.*

8.1.3 Asymptotic theory

A celebrated result of asymptotic statistics is the Bernstein–von Mises Theorem (Van der Vaart, 2000). In our experimental design set-up, this says that, under certain technical conditions and with i.i.d. random designs $\xi_t \overset{\text{i.i.d.}}{\sim} p(\xi)$, the posterior distribution $p(\theta|\mathcal{D}_t)$ is asymptotically Gaussian centred on the true value θ^* of the parameters of interest and with covariance matrix $t^{-1}M(\theta^*)^{-1}$. (Here, $M(\theta)$ is the Fisher information matrix, taking the expectation over $\xi \sim p(\xi)$.)

Paninski (2005) showed that a closely related result holds when designs are not random, but are chosen by greedy maximisation of the EIG.

Theorem 18 (Theorem 1 of Paninski (2005))**.** *Under certain technical conditions, the posterior distributions with greedy EIG maximisation are asymptotically Gaussian with mean θ^* and with covariance matrix $t^{-1}\Sigma_{info}$. Furthermore, if $t^{-1}\Sigma_{iid}$ is the asymptotic covariance with i.i.d. random designs, then*

$$\det \Sigma_{info} \leq \det \Sigma_{iid}. \tag{121}$$

This result tells us that the EIG maximisation strategy is no worse than i.i.d. sampling of designs, and that it will recover the true value θ^* in the limit as $t \to \infty$, i.e. the procedure is statistically consistent.

8.2 Non-greedy design policies

Whilst greedy policies enjoy computational tractability and some theoretical guarantees, a more direct approach to the problem of sequential experimental design is to seek the optimal policy that maximises equation (108). As we discuss in Sec. 9, finding this optimal policy can be cast in the language of reinforcement learning. In this section, we focus on computational approaches that have been suggested in the literature that specifically address non-greedy experimental design. These can generally be organised under two headings.

Forward sampling The forward sampling, or lookahead, family approaches relax the greedy assumption that the next experiment will be the last one. Instead, they assume that there will be m more experiments, and take account of these m future steps when deciding on the next experimental design. As m grows larger, this approach more closely approximates the truly optimal decision. However, with a larger m, the number of future outcomes to consider may grow exponentially. Such approaches either try to limit the number of outcomes considered, or else use a smaller value of m.

Backwards induction An alternative solution is to begin at the end. Classical optimisation theory (Bellman, 1966) shows that sequential optimisation problems are often more easily solved by starting with the final decision to be made at time T. For this final decision, the greedy solution is exactly optimal. The values of designs at later steps can be propagated backwards to inform earlier decisions. (See Sec. 9 for a fuller discussion.)

Non-greedy optimisation has typically been confined to low-dimensional cases within experimental design (Ryan et al., 2016). In medicine, Whitehead and Brunier (1995) and Whitehead and Williamson (1998) used a multi-step lookahead when finding optimal treatment doses. Berry and Ho (1988) explored optimal stopping when testing a one-sided hypothesis. Lewis and Berry (1994) applied backwards induction in a Bayesian clinical trials setting. Carlin et al. (1998) used forward sampling in a closely related clinical trial design problem. Brockwell and Kadane (2003) implemented backwards induction on a grid, and applied this to clinical trial planning. Müller et al. (2006) explored forward sampling for dose-response finding in clinical trials.

The main work to tackle the more general sequential experimental design problem, using the EIG utility, was Huan and Marzouk (2016). They used approximate dynamic programming to perform backwards induction by estimating the *value function*. The value function was then used to select optimal designs at each stage. The intermediate posterior distributions were estimated on a dynamically adapted grid.

In another line of work, González et al. (2016) build a predictor of future query locations given the current data. This allows them to use a forward sampling approach that is restricted to a single future trajectory. Jiang et al. (2020) use a related approach in which the future query locations are learned by repeatedly solving the *static* design optimisation problem with $T - t$ designs, but only using one of these designs at each step.

9 Bayesian Reinforcement Learning

Reinforcement learning (RL) (Sutton, 1990; Szepesvári, 2010) has a number of important and fascinating connections to sequential Bayesian experimental design. First, the problem of sequential experimental design

is a reinforcement learning problem. Specifically, we will show how the set-up of the preceding section can be cast as a Bayes Adaptive Markov Decision Process (BAMDP) (Ross et al., 2007; Guez et al., 2012; Ghavamzadeh et al., 2016). Second, the problem of making sequential decision to learn about a model is deeply connected to *exploration* in model-based reinforcement learning (Sun et al., 2011; Shyam et al., 2019; Sekar et al., 2020).

9.1 Sequential Bayesian Experimental Design as a BAMDP

The BAMDP is a generalisation of the Markov Decision Process (Bellman, 1957; Duff, 2002) that accommodates an unknown transition model. Adopting the notation of Guez et al. (2012), a BAMDP can be described by its augmented state space S^+, action space A, augmented transition model $\mathcal{P}^+$, reward function R^+ and discount factor γ. The augmented state space consists of the *history* of all states and actions previously visited $h_t = s_1 a_1 \dots a_{t-1} s_t$. This data is used to update the transition model in a Bayesian manner, using

$$p(\mathcal{P}|h_t) \propto p(\mathcal{P})p(h_t|\mathcal{P}). \tag{122}$$

For a sampled transition model, the probability of moving from s_t to s_{t+1} when action a_t was used is

$$p(s_{t+1}|s_t, a_t, \mathcal{P}) = \mathcal{P}(s_t, a_t, s_{t+1}). \tag{123}$$

The BAMDP transition model is therefore given by the marginal (Guez et al., 2012)

$$p(s_{t+1}|a_t, h_t) = \int p(\mathcal{P}|h_t)\mathcal{P}(s_t, a_t, s_{t+1})\, d\mathcal{P}. \tag{124}$$

The reward for using action a in state s is sampled as $r \sim R(s, a)$. Planning in a BAMDP means finding the policy that maximises

$$\mathcal{J}(\pi) = \mathbb{E}_\pi \left[\sum_{t=1}^{T} \gamma^{-t} r_t \right]. \tag{125}$$

To set up sequential Bayesian experimental design in this framework, we associate the augmented history states with the data $\mathcal{D}_t$ up to time t. The actions of the BAMDP are the experimental designs ξ_t. The transition model is associated with the model parameters θ (we assume in this section that we are not considering an embedded model). The 'transitions' of a sequential experiment are given by

$$p(\mathcal{D}_{t+1}|\mathcal{D}_t, \xi_{t+1}) = \mathbb{E}_{p(\theta|\mathcal{D}_t)}[p(y_{t+1}|\theta, \xi)] = \int p(\theta|\mathcal{D}_t)p(y_{t+1}|\theta, \xi_{t+1})\, d\theta \tag{126}$$

which agrees with equation (124) if we take $\mathcal{P}_\theta(y_{t+1}, \xi_{t+1}, y_t) = p(y_{t+1}|\theta, \xi_{t+1})$. Note that we write a_t as ξ_{t+1}, and that in the exchangeable experimental design case the transition model does not depend explicitly on y_t.

The only minor distinction from the set-up of Guez et al. (2012) is that the rewards in experimental design depend on the augmented state $\mathcal{D}_t$ rather than the state s_t. We can take the reward function for experimental design to be $R(\mathcal{D}_t) = \mathbf{1}[t = T]\mathbb{E}_{p(\theta|\mathcal{D}_t)}[U(\theta, \mathcal{D}_t)]$. Setting the discount factor $\gamma = 1$, we see that the BAMDP objective equation (125) is the same as the sequential experimental design problem equation (100). This shows the close connection between these two fields. For completeness, the value function and Q-function (Szepesvári, 2010) for Bayesian experimental design are given by

$$V^\pi(\mathcal{D}_t) = \mathbb{E}_{p(\mathcal{D}_T|\mathcal{D}_t,\pi)}\left[\mathbb{E}_{p(\theta|\mathcal{D}_T)}[U(\theta, \mathcal{D}_T)]\right] \tag{127}$$

$$Q^\pi(\mathcal{D}_t, \xi_{t+1}) = \mathbb{E}_{p(\mathcal{D}_T|\mathcal{D}_t,\xi_{t+1},\pi)}\left[\mathbb{E}_{p(\theta|\mathcal{D}_T)}[U(\theta, \mathcal{D}_T)]\right] \tag{128}$$

where

$$p(\mathcal{D}_T|\mathcal{D}_t, \pi) = \mathbb{E}_{p(\theta|\mathcal{D}_t)}\left[\prod_{\tau=t+1}^{T} \pi(\xi_\tau|\mathcal{D}_{\tau-1})p(y_\tau|\theta, \xi_\tau)\right] \tag{129}$$

$$p(\mathcal{D}_T|\mathcal{D}_t, \xi_{t+1}, \pi) = \mathbb{E}_{p(\theta|\mathcal{D}_t)}\left[p(y_{t+1}|\theta, \xi_{t+1}) \prod_{\tau=t+2}^{T} \pi(\xi_\tau|\mathcal{D}_{\tau-1})p(y_\tau|\theta, \xi_\tau)\right]. \tag{130}$$

Belief states In the previous section, we followed Guez et al. (2012) and took the state space for experimental design to be the dataset $\mathcal{D}_t$. We see from equation (126) that the transition model only depends on $\mathcal{D}_t$ via the posterior $p(\theta|\mathcal{D}_t)$. Furthermore, our choice of reward function only depends on $p(\theta|\mathcal{D}_t)$ (plus an indicator that we have reached the final stage). Thus, it is sufficient to take $p(\theta|\mathcal{D}_t)$ as our augmented state. Posterior distributions treated as states are referred to as *belief states* (Igl et al., 2018). They have been utilised extensively in Bayesian RL (Igl et al., 2018; Zintgraf et al., 2019; Ghavamzadeh et al., 2016) and are beginning to be used in Bayesian experimental design (Huan and Marzouk, 2016).

9.2 Exploration

We have seen the close connection between sequential Bayesian experimental design and Bayesian RL. We associated the transition model of an unknown MDP with the model parameter θ. In this framing, we have a new interpretation of objective functions for experimental design—they encourage the collection of data to improve knowledge of the transition model and are motivated by model-derived quantities, rather than by an external reward signal. Utility functions for experimental design can thus be reinterpreted as rewards for *exploration* behaviour that leads to improved knowledge in a model of the environment.

The experimental design scenario is most closely connected with model-based reinforcement learning (Sutton, 1990). Specifically, we consider reinforcement learning settings in which we have a Bayesian parametric model of the environment with parameter θ. A range of authors have considered 'intrinsic rewards' (Singh et al., 2005)—unlike external rewards which are separate from the model and environment dynamics, intrinsic rewards encourage behaviour to learn about the environment. For example, Itti and Baldi (2006) used surprisal as an intrinsic reward—agents are encouraged to take actions for which the outcome is not predictable, and hence will be surprising. Mathematically, surprisal can be defined using predictive entropy. Empowerment (Klyubin et al., 2005; Salge et al., 2014; Mohamed and Rezende, 2015) is another intrinsic reward signal that is based on conditional mutual information between state and action variables. Sajid et al. (2021) studied curiosity-driven exploration and the connection with free energy minimisation.

One line of research uses EIG as an intrinsic reward signal (Storck et al., 1995). This curiosity-driven exploration (Schmidhuber, 2010; Sun et al., 2011) is therefore the closest part of the RL literature to sequential experimental design. Specifically, Sun et al. (2011) utilise information gain as a reward. Given history h and h' such that h is a prefix of h' they define

$$\mathrm{IG}(h'\|h) = \mathrm{KL}\left(p(\theta|h')\|p(\theta|h)\right). \tag{131}$$

To motivate this choice, Sun et al. (2011) proved the following result (a more formal version of Example 12)

Proposition 19 (Sun et al. (2011))**.** *Let $h \subseteq h' \subseteq h''$ be histories such that h a prefix of h' and h' a prefix of h''. Suppose h' has been observed. Then,*

$$\mathbb{E}_{h''|h'}[\mathrm{IG}(h''\|h)] = \mathrm{IG}(h'\|h) + \mathbb{E}_{h''|h'}[\mathrm{IG}(h''\|h')] \tag{132}$$

so the information gain is additive in expectation.

Information gain for exploration was applied to robotics by Fung et al. (2016).

9.2.1 Computational approaches to exploration in Bayesian RL with EIG

To utilise information gain as an intrinsic reward for exploration requires approximation and optimisation of this quantity. Storck et al. (1995) focused on the tabular setting with finite states and actions, in which the transition model can be described with a finite number of parameters. Sun et al. (2011) also focused on the finite space case for their computations. Houthooft et al. (2016) tackled the continuous space problem. They used variational inference (Rezende et al., 2014; Kingma and Welling, 2014) to estimate the posterior distributions $p(\theta|\mathcal{D}_t)$. They then used the variational approximate posterior as a surrogate for the true posterior when computing the information gain reward. Information gain was combined with an external reward signal to balance exploration and exploitation. Shyam et al. (2019) used an ensemble to approximate

the distribution $p(\theta|\mathcal{D}_t)$, to estimate information gain they replaced Shannon entropy with Rényi entropy which can be calculated for a mixture of Gaussians. Sekar et al. (2020) used a closely related approach. Rather than the Rényi entropy, they used the empirical variance of ensemble means as a way of estimating the intractable marginal entropy that occurs in the EIG.

References

Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3): 4440–4458, 2019.

Billy Amzal, Frédéric Y Bois, Eric Parent, and Christian P Robert. Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical association*, 101(474):773–785, 2006.

Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50(1):5–43, 2003.

AC Atkinson and VV Fedorov. The design of experiments for discriminating between two rival models. *Biometrika*, 62(1):57–70, 1975.

Anthony Atkinson, Alexander Donev, and Randall Tobias. *Optimum experimental designs, with SAS*, volume 34. Oxford University Press, 2007.

Tim Bakker, Herke van Hoof, and Max Welling. Experimental design for mri by greedy policy search. *Advances in Neural Information Processing Systems*, 33, 2020.

Maximilian Balandat, Brian Karrer, Daniel Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. Botorch: A framework for efficient monte-carlo bayesian optimization. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

George A Barnard, Gs M Jenkins, and CB Winsten. Likelihood inference and time series. *Journal of the Royal Statistical Society: Series A (General)*, 125(3):321–352, 1962.

Joakim Beck, Ben Mansour Dia, Luis FR Espath, Quan Long, and Raul Tempone. Fast bayesian experimental design: Laplace-based importance sampling for the expected information gain. *Computer Methods in Applied Mechanics and Engineering*, 334:523–553, 2018.

Richard Bellman. A markovian decision process. *Journal of mathematics and mechanics*, 6(5):679–684, 1957.

Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

William H Beluch, Tim Genewein, Andreas Nürnberger, and Jan M Köhler. The power of ensembles for active learning in image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9368–9377, 2018.

James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24, 2011.

José M Bernardo. Expected information as expected utility. *the Annals of Statistics*, pages 686–690, 1979.

Donald A Berry and Chih-Hsiang Ho. One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach. *Biometrics*, pages 219–227, 1988.

Concha Bielza, Peter Müller, and David Ríos Insua. Decision analysis by augmented probability simulation. *Management Science*, 45(7):995–1007, 1999.

Allan Birnbaum. On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298):269–306, 1962.

Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistic symmetry and invariant neural networks. *arXiv preprint arXiv:1901.06082*, 2019.

George EP Box. Choice of response surface design and alphabetic optimality. Technical report, Wisconsin University-Madison Mathematics Research Center, 1982.

Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. Constraining effective field theories with machine learning. *Physical review letters*, 121(11):111801, 2018.

Anthony E Brockwell and Joseph B Kadane. A gridding method for bayesian sequential decision problems. *Journal of Computational and Graphical Statistics*, 12(3):566–584, 2003.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pages 23–37. Springer, 2009.

Alberto Giovanni Busetto, Cheng Soon Ong, and Joachim M Buhmann. Optimized expected information gain for nonlinear dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 97–104, 2009.

Roberto Calandra, Jan Peters, Carl Edward Rasmussen, and Marc Peter Deisenroth. Manifold gaussian processes for regression. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 3338–3345. IEEE, 2016.

Bradley P Carlin, Joseph B Kadane, and Alan E Gelfand. Approaches for optimal sequential decision analysis in clinical trials. *Biometrics*, pages 964–975, 1998.

André Gustavo Carlon, Ben Mansour Dia, Luis Espath, Rafael Holdorf Lopez, and Raul Tempone. Nesterov-aided stochastic gradient methods using laplace approximation for bayesian design optimization. *Computer Methods in Applied Mechanics and Engineering*, 363:112909, 2020.

Daniel R Cavagnaro, Jay I Myung, Mark A Pitt, and Janne V Kujala. Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural computation*, 22 (4):887–905, 2010.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

Yuxin Chen, S Hamed Hassani, Amin Karbasi, and Andreas Krause. Sequential information maximization: When is greedy near-optimal? In *Conference on Learning Theory*, pages 338–363. PMLR, 2015.

Yuxin Chen, Hamed Hassani, and Andreas Krause. Near-optimal bayesian active learning with correlated and noisy tests. In *Artificial Intelligence and Statistics*, pages 223–231. PMLR, 2017.

Merlise A Clyde, Peter Müller, and Giovanni Parmigiani. Exploring expected utility surfaces by markov chains. Technical report, Institute of Statistics and Decision Sciences, Duke University, Durham, NC, USA, 1996.

Alex R Cook, Gavin J Gibson, and Christopher A Gilligan. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

David Roxbee Cox. *Principles of statistical inference*. Cambridge university press, 2006.

Katalin Csilléry, Michael GB Blum, Oscar E Gaggiotti, and Olivier François. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution*, 25(7):410–418, 2010.

Mahasen B Dehideniya, Christopher C Drovandi, and James M McGree. Optimal bayesian design for discriminating between models with intractable likelihoods in epidemiology. *Computational Statistics & Data Analysis*, 124:277–297, 2018.

Joseph J DiStefano III, Allen R Stubberud, and Ivan J Williams. *Schaum's outline of feedback and control systems*. McGraw-Hill Education, 2014.

Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

Christopher C Drovandi and Anthony N Pettitt. Bayesian experimental design for models with intractable likelihoods. *Biometrics*, 69(4):937–948, 2013.

Christopher C Drovandi, James M McGree, and Anthony N Pettitt. A sequential monte carlo algorithm to incorporate model uncertainty in bayesian sequential design. *Journal of Computational and Graphical Statistics*, 23(1):3–24, 2014.

Michael O'Gordon Duff. *Optimal Learning: Computational procedures for Bayes-adaptive Markov decision processes*. University of Massachusetts Amherst, 2002.

Sergey Dushenko, Kapildeb Ambal, and Robert D McMichael. Sequential bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers. *Physical Review Applied*, 14(5):054036, 2020.

Agoston E Eiben, James E Smith, et al. *Introduction to evolutionary computing*, volume 53. Springer, 2003.

Gustav Elfving. Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, pages 255–262, 1952.

Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems 32*, pages 14036–14047. Curran Associates, Inc., 2019.

Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, and Tom Rainforth. A unified stochastic gradient approach to designing bayesian-optimal experiments. In *International Conference on Artificial Intelligence and Statistics*, pages 2959–2969. PMLR, 2020.

Adam Foster, Desi R Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. *arXiv preprint arXiv:2103.02438*, 2021.

Linton C Freeman. *Elementary applied statistics: for students in behavioral science*. New York: Wiley, 1965.

Nicholas C Fung, Carlos Nieto-Granda, Jason M Gregory, and John G Rogers. Autonomous exploration using an information gain metric. Technical report, US Army Research Laboratory Adelphi United States, 2016.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pages 1183–1192. PMLR, 2017.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

Mohammad Ghavamzadeh, Shie Mannor, Joelle Pineau, and Aviv Tamar. Bayesian reinforcement learning: A survey. *arXiv preprint arXiv:1609.04436*, 2016.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

Daniel Golovin, Andreas Krause, and Debajyoti Ray. Near-optimal bayesian active learning with noisy observations. In *Advances in Neural Information Processing Systems*, pages 766–774, 2010.

Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A Haddad, Tim P Vogels, et al. Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife*, 9:e56261, 2020.

Wenbo Gong, Sebastian Tschiatschek, Richard Turner, Sebastian Nowozin, José Miguel Hernández-Lobato, and Cheng Zhang. Icebreaker: Element-wise active information acquisition with bayesian deep latent gaussian model. *arXiv preprint arXiv:1908.04537*, 2019.

Javier González, Michael Osborne, and Neil Lawrence. Glasses: Relieving the myopia of bayesian optimisation. In *Artificial Intelligence and Statistics*, pages 790–799. PMLR, 2016.

Arthur Guez, David Silver, and Peter Dayan. Efficient bayes-adaptive reinforcement learning using sample-based search. In *Advances in neural information processing systems*, pages 1025–1033, 2012.

Markus Hainy, Werner G Müller, and Helga Wagner. Likelihood-free simulation-based optimal design with an application to spatial extremes. *Stochastic Environmental Research and Risk Assessment*, 30(2):481–492, 2016.

M Hamada, HF Martz, CS Reese, and AG Wilson. Finding near-optimal bayesian experimental designs via genetic algorithms. *The American Statistician*, 55(3):175–181, 2001.

Cong Han and Kathryn Chaloner. Bayesian experimental design for nonlinear mixed-effects models with application to hiv dynamics. *Biometrics*, 60(1):25–33, 2004.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning (2nd edition)*. Springer-Verlag, 2009.

W Keith Hastings. *Monte Carlo sampling methods using Markov chains and their applications*. Oxford University Press, 1970.

Daniel W Heck and Edgar Erdfelder. Maximizing the expected information gain of cognitive modeling via design optimization. *Computational Brain & Behavior*, 2(3):202–209, 2019.

Frank Heinrich, Paul A Kienzle, David P Hoogerheide, and Mathias Lösche. Information gain from isotopic contrast variation in neutron reflectometry on protein–membrane complex structures. *Journal of applied crystallography*, 53(3):800–810, 2020.

Philipp Hennig and Christian J Schuler. Entropy search for information-efficient global optimization. *Journal of Machine Learning Research*, 13(Jun):1809–1837, 2012.

José Miguel Hernández-Lobato, Matthew W Hoffman, and Zoubin Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *Advances in neural information processing systems*, pages 918–926, 2014.

Matthew Hoffman, Bobak Shahriari, and Nando Freitas. On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning. In *Artificial Intelligence and Statistics*, pages 365–374. PMLR, 2014.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Rein Houthooft, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel. Vime: Variational information maximizing exploration. *arXiv preprint arXiv:1605.09674*, 2016.

Xun Huan and Youssef Marzouk. Gradient-based stochastic optimization methods in bayesian experimental design. *International Journal for Uncertainty Quantification*, 4(6), 2014.

Xun Huan and Youssef M Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, 2013.

Xun Huan and Youssef M Marzouk. Sequential bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*, 2016.

Frank Hutter, Holger Hoos, and Kevin Leyton-Brown. An evaluation of sequential model-based optimization for expensive blackbox functions. In *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*, pages 1209–1216, 2013.

Maximilian Igl, Luisa Zintgraf, Tuan Anh Le, Frank Wood, and Shimon Whiteson. Deep variational reinforcement learning for pomdps. In *International Conference on Machine Learning*, pages 2117–2126. PMLR, 2018.

Laurent Itti and Pierre F Baldi. Bayesian surprise attracts human attention. In *Advances in neural information processing systems*, pages 547–554. Citeseer, 2006.

Shali Jiang, Henry Chai, Javier Gonzalez, and Roman Garnett. Binoculars for efficient, nonmyopic sequential experimental design. In *International Conference on Machine Learning*, pages 4794–4803. PMLR, 2020.

Ashish Kapoor, Kristen Grauman, Raquel Urtasun, and Trevor Darrell. Active learning with gaussian processes for object categorization. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680*, 2015.

Jack Kiefer and Jacob Wolfowitz. Optimum designs in regression problems. *The annals of mathematical statistics*, pages 271–294, 1959.

Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. In *ICLR*, 2014.

Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32:7026–7037, 2019.

Steven Kleinegesse and Michael Gutmann. Efficient Bayesian experimental design for implicit models. *arXiv preprint arXiv:1810.09912*, 2018.

Alexander S Klyubin, Daniel Polani, and Chrystopher L Nehaniv. All else being equal be empowered. In *European Conference on Artificial Life*, pages 744–753. Springer, 2005.

Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability*, 3:71–104, 2014.

Andreas Krause and Carlos E Guestrin. Near-optimal nonmyopic value of information in graphical models. *arXiv preprint arXiv:1207.1394*, 2012.

John Kruschke. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press, 2014.

Hendrik Kuck, Nando de Freitas, and Arnaud Doucet. Smc samplers for bayesian optimal nonlinear design. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 99–102. IEEE, 2006.

Harold J Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J of Basic Engineering*, 1964.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

Miguel Lázaro-Gredilla, Joaquin Quinonero-Candela, Carl Edward Rasmussen, and Aníbal R Figueiras-Vidal. Sparse spectrum gaussian process regression. *The Journal of Machine Learning Research*, 11: 1865–1881, 2010.

Jeremy Lewi, Robert Butera, and Liam Paninski. Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21(3):619–687, 2009.

David D Lewis and William A Gale. A sequential algorithm for training text classifiers. In *SIGIR'94*, pages 3–12. Springer, 1994.

Roger J Lewis and Donald A Berry. Group sequential clinical trials: a classical evaluation of bayesian decision-theoretic designs. *Journal of the American Statistical Association*, 89(428):1528–1534, 1994.

Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.

Dennis V Lindley. *Bayesian statistics, a review*, volume 2. SIAM, 1972.

Quan Long. Multimodal information gain in bayesian design of experiments. *Computational Statistics*, pages 1–21, 2021.

Quan Long, Marco Scavino, Raúl Tempone, and Suojin Wang. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39, 2013.

Thomas J Loredo. Bayesian adaptive exploration. In *AIP Conference Proceedings*, volume 707, pages 330–346. American Institute of Physics, 2004.

Jiankun Lyu, Sheng Wang, Trent E Balius, Isha Singh, Anat Levit, Yurii S Moroz, Matthew J O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, et al. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224, 2019.

Chao Ma, Sebastian Tschiatschek, Konstantina Palla, José Miguel Hernández-Lobato, Sebastian Nowozin, and Cheng Zhang. EDDI: Efficient dynamic discovery of high-value information with Partial VAE. *arXiv preprint arXiv:1809.11142*, 2018.

David JC MacKay. Information-based objective functions for active data selection. *Neural computation*, 4 (4):590–604, 1992.

James Matthew McGree, Christopher C Drovandi, MH Thompson, JA Eccleston, SB Duffull, Kerrie Mengersen, Anthony N Pettitt, and Timothy Goggin. Adaptive bayesian compound designs for dose finding studies. *Journal of Statistical Planning and Inference*, 142(6):1480–1492, 2012.

Ruth K Meyer and Christopher J Nachtsheim. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69, 1995.

Thomas Peter Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Institute of Technology, 2001.

Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. The application of bayesian methods for seeking the extremum. *Towards global optimization*, 2(117-129):2, 1978.

Shakir Mohamed and Danilo Jimenez Rezende. Variational information maximisation for intrinsically motivated reinforcement learning. *arXiv preprint arXiv:1509.08731*, 2015.

Peter Müller. Simulation based optimal design. *Handbook of Statistics*, 25:509–518, 2005.

Peter Müller and Giovanni Parmigiani. Optimal design via curve fitting of monte carlo experiments. *Journal of the American Statistical Association*, 90(432):1322–1330, 1995.

Peter Müller, Bruno Sansó, and Maria De Iorio. Optimal bayesian design by inhomogeneous markov chain simulation. *Journal of the American Statistical Association*, 99(467):788–798, 2004.

Peter Müller, Don A Berry, Andrew P Grieve, and Michael Krams. A bayesian decision-theoretic dose-finding trial. *Decision analysis*, 3(4):197–207, 2006.

Jay I Myung, Daniel R Cavagnaro, and Mark A Pitt. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, 2013.

Willie Neiswanger, Kirthevasan Kandasamy, Barnabas Poczos, Jeff Schneider, and Eric Xing. Probo: Versatile bayesian optimization using any probabilistic programming language. *arXiv preprint arXiv:1901.11515*, 2019.

John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4): 308–313, 1965.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14(1):265–294, 1978.

James Neufeld, Andras Gyorgy, Csaba Szepesvári, and Dale Schuurmans. Adaptive monte carlo via bandit allocation. In *International Conference on Machine Learning*, pages 1944–1952. PMLR, 2014.

Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

Michael A Osborne, Roman Garnett, and Stephen J Roberts. Gaussian processes for global optimization. In *3rd international conference on learning and intelligent optimization (LION3)*, pages 1–15. Citeseer, 2009.

Antony Overstall and James McGree. Bayesian design of experiments for intractable likelihood models using coupled auxiliary models and multivariate emulation. *Bayesian Analysis*, 15(1):103–131, 2020.

Antony M Overstall and David C Woods. Bayesian design of experiments using approximate coordinate exchange. *Technometrics*, 59(4):458–470, 2017.

J Lynn Palmer and Peter Müller. Bayesian optimal design in population models for haematologic data. *Statistics in medicine*, 17(14):1613–1622, 1998.

Liam Paninski. Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17 (7):1480–1507, 2005.

Costas Papadimitriou. Optimal sensor placement methodology for parametric identification of structural systems. *Journal of sound and vibration*, 278(4-5):923–947, 2004.

Andy Pole, Mike West, and Jeff Harrison. *Applied Bayesian forecasting and time series analysis*. Chapman and Hall/CRC, 2018.

Remus Pop and Patric Fulop. Deep ensemble bayesian active learning: Addressing the mode collapse issue in monte carlo dropout via ensembles. *arXiv preprint arXiv:1811.03897*, 2018.

David J Price, Nigel G Bean, Joshua V Ross, and Jonathan Tuke. On the efficient determination of optimal bayesian experimental designs using abc: A case study in optimal observation of epidemics. *Journal of Statistical Planning and Inference*, 172:1–15, 2016.

David J Price, Nigel G Bean, Joshua V Ross, and Jonathan Tuke. An induced natural selection heuristic for finding optimal bayesian experimental designs. *Computational Statistics & Data Analysis*, 126:112–124, 2018.

Joaquin Quinonero-Candela and Carl Edward Rasmussen. A unifying view of sparse approximate gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959, 2005.

Tom Rainforth. *Automating Inference, Learning, and Design using Probabilistic Programming*. PhD thesis, University of Oxford, 2017.

Tom Rainforth, Rob Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting monte carlo estimators. In *International Conference on Machine Learning*, pages 4267–4276. PMLR, 2018.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 1278–1286, 2014.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.

Stephane Ross, Brahim Chaib-draa, and Joelle Pineau. Bayes-adaptive pomdps. In *NIPS*, pages 1225–1232, 2007.

N Roy and A McCallum. Toward optimal active learning through sampling estimation of error reduction. int. conf. on machine learning, 2001.

Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

Elizabeth G. Ryan, Christopher C. Drovandi, M. Helen Thompson, and Anthony N. Pettitt. Towards bayesian experimental design for nonlinear models that require a large number of sampling times. *Computational Statistics & Data Analysis*, 70:45–60, 2014. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2013.08.017. URL `https://www.sciencedirect.com/science/article/pii/S0167947313003149`.

Elizabeth G Ryan, Christopher C Drovandi, and Anthony N Pettitt. Fully Bayesian experimental design for pharmacokinetic studies. *Entropy*, 17(3):1063–1089, 2015.

Elizabeth G Ryan, Christopher C Drovandi, James M McGree, and Anthony N Pettitt. A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016.

Kenneth J Ryan. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12(3):585–603, 2003.

Noor Sajid, Philip J Ball, Thomas Parr, and Karl J Friston. Active inference: demystified and compared. *Neural Computation*, 33(3):674–712, 2021.

Christoph Salge, Cornelius Glackin, and Daniel Polani. Empowerment–an introduction. In *Guided Self-Organization: Inception*, pages 67–114. Springer, 2014.

Tobias Scheffer, Christian Decomain, and Stefan Wrobel. Active hidden markov models for information extraction. In *International Symposium on Intelligent Data Analysis*, pages 309–318. Springer, 2001.

Jürgen Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, 2010.

Paola Sebastiani and Henry P Wynn. Maximum entropy sampling and optimal Bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1), 2000.

Ramanan Sekar, Oleh Rybkin, Kostas Daniilidis, Pieter Abbeel, Danijar Hafner, and Deepak Pathak. Planning to explore via self-supervised world models. In *International Conference on Machine Learning*, pages 8583–8592. PMLR, 2020.

Burr Settles. *Active learning literature survey*. PhD thesis, University of Wisconsin-Madison Department of Computer Sciences, 2009.

Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 1070–1079, 2008.

Ben Shababo, Brooks Paige, Ari Pakman, and Liam Paninski. Bayesian inference and online experimental design for mapping neural microcircuits. In *Advances in Neural Information Processing Systems*, pages 1304–1312, 2013.

Amar Shah and Zoubin Ghahramani. Parallel predictive entropy search for batch global optimization of expensive objective functions. *arXiv preprint arXiv:1511.07130*, 2015.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27 (3):379–423, 1948.

Pranav Shyam, Wojciech Jaśkowski, and Faustino Gomez. Model-based active exploration. In *International conference on machine learning*, pages 5779–5788. PMLR, 2019.

Satinder Singh, Andrew G Barto, and Nuttapong Chentanez. Intrinsically motivated reinforcement learning. Technical report, MASSACHUSETTS UNIV AMHERST DEPT OF COMPUTER SCIENCE, 2005.

Scott A Sisson, Yanan Fan, and Mark Beaumont. *Handbook of approximate Bayesian computation*. CRC Press, 2018.

Lewis Smith and Yarin Gal. Understanding measures of uncertainty for adversarial example detection. *arXiv preprint arXiv:1803.08533*, 2018.

Edward Snelson and Zoubin Ghahramani. Sparse gaussian processes using pseudo-inputs. *Advances in neural information processing systems*, 18:1257, 2006.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat, and Ryan Adams. Scalable bayesian optimization using deep neural networks. In *International conference on machine learning*, pages 2171–2180. PMLR, 2015.

James C Spall. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins apl technical digest*, 19(4):482–492, 1998.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Jan Storck, Sepp Hochreiter, Jürgen Schmidhuber, et al. Reinforcement driven information acquisition in non-deterministic environments. In *Proceedings of the international conference on artificial neural networks, Paris*, volume 2, pages 159–164. Citeseer, 1995.

Jonathan R Stroud, Peter Müller, and Gary L Rosner. Optimal sampling times in population pharmacokinetic studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):345–359, 2001.

Yi Sun, Faustino Gomez, and Jürgen Schmidhuber. Planning to be surprised: Optimal bayesian exploration in dynamic environments. In *International Conference on Artificial General Intelligence*, pages 41–51. Springer, 2011.

Richard S Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Machine learning proceedings 1990*, pages 216–224. Elsevier, 1990.

Csaba Szepesvári. Algorithms for reinforcement learning. *Synthesis lectures on artificial intelligence and machine learning*, 4(1):1–103, 2010.

Gabriel Terejanu, Rochan R Upadhyay, and Kenji Miki. Bayesian experimental design for the active nitridation of graphite by atomic nitrogen. *Experimental Thermal and Fluid Science*, 36:178–193, 2012.

Owen Thomas, Ritabrata Dutta, Jukka Corander, Samuel Kaski, and Michael U Gutmann. Likelihood-free inference by ratio estimation. *arXiv preprint arXiv:1611.10242*, 2016.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Robert K Tsutakawa. Design of experiment for bioassay. *Journal of the American Statistical Association*, 67(339):584–590, 1972.

Jojanneke van den Berg, Andrew Curtis, and Jeannot Trampert. Optimal nonlinear bayesian experimental design: an application to amplitude versus offset experiments. *Geophysical Journal International*, 155(2): 411–421, 2003.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Peter JM Van Laarhoven and Emile HL Aarts. Simulated annealing. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987.

Joep Vanlier, Christian A Tiemann, Peter AJ Hilbers, and Natal AW van Riel. A Bayesian approach to targeted experiment design. *Bioinformatics*, 28(8):1136–1142, 2012.

Joep Vanlier, Christian A Tiemann, Peter AJ Hilbers, and Natal AW van Riel. Optimal experiment design for model selection in biochemical networks. *BMC systems biology*, 8(1):1–16, 2014.

Julien Villemonteix, Emmanuel Vazquez, and Eric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.

Benjamin T Vincent and Tom Rainforth. The DARC toolbox: automated, flexible, and efficient delayed and risky choice experiments using bayesian adaptive design. *Retrieved from psyarxiv.com/yehjb*, 2017.

Julius von Kügelgen, Paul K Rubenstein, Bernhard Schölkopf, and Adrian Weller. Optimal experimental design via bayesian optimization: active causal structure learning for gaussian process networks. *arXiv preprint arXiv:1910.03962*, 2019.

Jon Wakefield. An expected loss approach to the design of dosage regimens via sampling-based methods. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(1):13–29, 1994.

Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient bayesian optimization. In *International Conference on Machine Learning*, pages 3627–3635. PMLR, 2017.

Andrew B Watson. Quest+: A general multidimensional bayesian adaptive psychometric method. *Journal of Vision*, 17(3):10–10, 2017.

John Whitehead and Hazel Brunier. Bayesian decision procedures for dose determining experiments. *Statistics in medicine*, 14(9):885–893, 1995.

John Whitehead and David Williamson. Bayesian decision procedures based on logistic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics*, 8(3):445–467, 1998.

Christopher K Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

Liu Yang, Rong Jin, and Rahul Sukthankar. Bayesian active distance metric learning. *arXiv preprint arXiv:1206.5283*, 2012.

Sue Zheng, Jason Pacheco, and John Fisher. A robust approach to sequential information theoretic planning. In *International Conference on Machine Learning*, pages 5941–5949, 2018.

Luisa Zintgraf, Kyriacos Shiarlis, Maximilian Igl, Sebastian Schulze, Yarin Gal, Katja Hofmann, and Shimon Whiteson. Varibad: A very good method for bayes-adaptive deep rl via meta-learning. *arXiv preprint arXiv:1910.08348*, 2019.

Chapter 2

Variational Bayesian Optimal Experimental Design

This paper was published as the following

Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems 32*, pages 14036–14047. Curran Associates, Inc., 2019.

Variational Bayesian Optimal Experimental Design

Adam Foster[†*] **Martin Jankowiak**[‡] **Eli Bingham**[‡] **Paul Horsfall**[‡]
Yee Whye Teh[†] **Tom Rainforth**[†] **Noah Goodman**[‡§]
[†]Department of Statistics, University of Oxford, Oxford, UK
[‡]Uber AI Labs, Uber Technologies Inc., San Francisco, CA, USA
[§]Stanford University, Stanford, CA, USA
adam.foster@stats.ox.ac.uk

Abstract

Bayesian optimal experimental design (BOED) is a principled framework for making efficient use of limited experimental resources. Unfortunately, its applicability is hampered by the difficulty of obtaining accurate estimates of the expected information gain (EIG) of an experiment. To address this, we introduce several classes of fast EIG estimators by building on ideas from amortized variational inference. We show theoretically and empirically that these estimators can provide significant gains in speed and accuracy over previous approaches. We further demonstrate the practicality of our approach on a number of end-to-end experiments.

1 Introduction

Tasks as seemingly diverse as designing a study to elucidate human cognition, selecting the next query point in an active learning loop, and designing online feedback surveys all constitute the same underlying problem: designing an experiment to maximize the information gathered. Bayesian optimal experimental design (BOED) forms a powerful mathematical abstraction for tackling such problems [8, 23, 37, 43] and has been successfully applied in numerous settings, including psychology [30], Bayesian optimization [16], active learning [15], bioinformatics [42], and neuroscience [38].

In the BOED framework, we construct a predictive model $p(y|\theta, d)$ for possible experimental outcomes y, given a design d and a particular value of the parameters of interest θ. We then choose the design that optimizes the expected information gain (EIG) in θ from running the experiment,

$$\mathrm{EIG}(d) \triangleq \mathbb{E}_{p(y|d)}\big[H[p(\theta)] - H[p(\theta|y,d)]\big], \tag{1}$$

where $H[\cdot]$ represents the entropy and $p(\theta|y,d) \propto p(\theta)p(y|\theta,d)$ is the posterior resulting from running the experiment with design d and observing outcome y. In other words, we seek the design that, in expectation over possible experimental outcomes, most reduces the entropy of the posterior over our target latent variables. If the predictive model is correct, this forms a design strategy that is (one-step) optimal from an information-theoretic viewpoint [24, 37].

The BOED framework is particularly powerful in sequential contexts, where it allows the results of previous experiments to be used in guiding the designs for future experiments. For example, as we ask a participant a series of questions in a psychology trial, we can use the information gathered from previous responses to ask more pertinent questions in the future, that will, in turn, return more information. This ability to design experiments that are self-adaptive can substantially increase their efficiency: fewer iterations are required to uncover the same level of information.

In practice, however, the BOED approach is often hampered by the difficulty of obtaining fast and high-quality estimates of the EIG: due to the intractability of the posterior $p(\theta|y,d)$, it constitutes

[*] Part of this work was completed by AF during an internship with Uber AI Labs.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

a nested expectation problem and so conventional Monte Carlo (MC) estimation methods cannot be applied [33]. Moreover, existing methods for tackling nested expectations have, in general, far inferior convergence rates than those for conventional expectations [22, 30, 32]. For example, nested MC (NMC) can only achieve, at best, a rate of $\mathcal{O}(T^{-1/3})$ in the total computational cost T [33], compared with $\mathcal{O}(T^{-1/2})$ for conventional MC.

To address this, we propose a variational BOED approach that sidesteps the double intractability of the EIG in a principled manner and yields estimators with convergence rates in line with those for conventional estimation problems. To this end, we introduce four efficient and widely applicable variational estimators for the EIG. The different methods each present distinct advantages. For example, two allow training with implicit likelihood models, while one allows for asymptotic consistency even when the variational family does not contain the target distribution.

We theoretically confirm the advantages of our estimators, showing that they all have a convergence rate of $\mathcal{O}(T^{-1/2})$ when the variational family contains the target distribution. We further verify their practical utility using a number of experiment design problems inspired by applications from science and industry, showing that they provide significant empirical gains in EIG estimation over previous methods and that these gains lead, in turn, to improved end-to-end performance.

To maximize the space of potential applications and users for our estimators, we provide[2] a general-purpose implementation of them in the probabilistic programming system Pyro [5], exploiting Pyro's first-class support for neural networks and variational methods.

2 Background

The BOED framework is a model-based approach for choosing an experiment design d in a manner that optimizes the information gained about some parameters of interest θ from the outcome y of the experiment. For instance, we may wish to choose the question d in a psychology trial to maximize the information gained about an underlying psychological property of the participant θ from their answer y to the question. In general, we adopt a Bayesian modelling framework with a prior $p(\theta)$ and a predictive model $p(y|\theta, d)$. The information gained about θ from running experiment d and observing y is the reduction in entropy from the prior to the posterior:

$$\mathrm{IG}(y, d) = H[p(\theta)] - H[p(\theta|y, d)]\,. \tag{2}$$

At the point of choosing d, however, we are uncertain about the outcome. Thus, in order to define a metric to assess the utility of the design d we take the expectation of $\mathrm{IG}(y, d)$ under the marginal distribution over outcomes $p(y|d) = \mathbb{E}_{p(\theta)}[p(y|\theta, d)]$ as per (1). We can further rearrange this as

$$\mathrm{EIG}(d) = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(\theta|y,d)}{p(\theta)}\right] = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y,\theta|d)}{p(\theta)p(y|d)}\right] = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta,d)}{p(y|d)}\right] \tag{3}$$

with the result that the EIG can also be interpreted as the mutual information between θ and y given d, or the epistemic uncertainty in y averaged over the prior $p(\theta)$. The Bayesian optimal design is defined as $d^* \triangleq \arg\max_{d\in\mathcal{D}} \mathrm{EIG}(d)$, where $\mathcal{D}$ is the set of permissible designs.

Computing the EIG is challenging since neither $p(\theta|y, d)$ or $p(y|d)$ can, in general, be found in closed form. Consequently, the integrand is intractable and conventional MC methods are not applicable. One common way of getting around this is to employ a nested MC (NMC) estimator [30, 43]

$$\hat{\mu}_{\mathrm{NMC}}(d) \triangleq \frac{1}{N}\sum_{n=1}^{N} \log \frac{p(y_n|\theta_{n,0}, d)}{\frac{1}{M}\sum_{m=1}^{M} p(y_n|\theta_{n,m}, d)} \quad \text{where} \quad \theta_{n,m} \overset{\text{i.i.d.}}{\sim} p(\theta),\ y_n \sim p(y|\theta = \theta_{n,0}, d). \tag{4}$$

Rainforth et al. [33] showed that this estimator, which has a total computational cost $T = \mathcal{O}(NM)$, is consistent in the limit $N, M \to \infty$ with RMSE convergence rate $\mathcal{O}(N^{-1/2} + M^{-1})$, and that it is asymptotically optimal to set $M \propto \sqrt{N}$, yielding an overall rate of $\mathcal{O}(T^{-1/3})$.

Given a base EIG estimator, a variety of different methods can be used for the subsequent optimization over designs, including some specifically developed for BOED [1, 29, 32]. In our experiments, we

[2] Implementations of our methods are available at `http://docs.pyro.ai/en/stable/contrib.oed.html`. To reproduce the results in this paper, see `https://github.com/ae-foster/pyro/tree/vboed-reproduce`.

will adopt Bayesian optimization [39], due to its sample efficiency, robustness to multi-modality, and ability to deal naturally with noisy objective evaluations. However, we emphasize that our focus is on the base EIG estimator and that our estimators can be used more generally with different optimizers.

The static design setting we have implicitly assumed thus far in our discussion can be generalized to sequential contexts, in which we design T experiments $d_1, ..., d_T$ with outcomes $y_1, ..., y_T$. We assume experiment outcomes are conditionally independent given the latent variables and designs, i.e.

$$p(y_{1:T}, \theta | d_{1:T}) = p(\theta) \prod_{t=1}^{T} p(y_t | \theta, d_t). \tag{5}$$

Having conducted experiments $1, ..., t-1$, we can design d_t by incorporating data in the standard Bayesian fashion: at experiment iteration t, we replace the prior $p(\theta)$ in (3) with $p(\theta | d_{1:t-1}, y_{1:t-1})$, the posterior conditional on the first $t-1$ designs and outcomes. We can thus conduct an adaptive sequential experiment in which we optimize the choice of the design d_t at each iteration.

3 Variational Estimators

Though consistent, the convergence rate of the NMC estimator is prohibitively slow for many practical problems. As such, EIG estimation often becomes the bottleneck for BOED, particularly in sequential experiments where the BOED calculations must be fast enough to operate in real-time.

In this section we show how ideas from amortized variational inference [10, 17, 34, 40] can be used to sidestep the double intractability of the EIG, yielding estimators with much faster convergence rates thereby alleviating the EIG bottleneck. A key insight for realizing why such fundamental gains can be made is that the NMC estimator is inefficient because a *separate* estimate of the integrand in (3) is made for each y_n. The variational approaches we introduce instead look to directly learn a *functional approximation*—for example, an approximation of $y \mapsto p(y|d)$—and then evaluate this approximation at multiple points to estimate the integral, thereby allowing information to be shared across different values of y. If M evaluations are made in learning the approximation, the total computational cost is now $T = \mathcal{O}(N + M)$, yielding substantially improved convergence rates.

Variational posterior $\hat{\mu}_{\text{post}}$ Our first approach, which we refer to as the variational posterior estimator $\hat{\mu}_{\text{post}}$, is based on learning an amortized approximation $q_p(\theta|y, d)$ to the posterior $p(\theta|y, d)$ and then using this to estimate the EIG:

$$\text{EIG}(d) \approx \mathcal{L}_{\text{post}}(d) \triangleq \mathbb{E}_{p(y,\theta|d)} \left[\log \frac{q_p(\theta|y, d)}{p(\theta)} \right] \approx \hat{\mu}_{\text{post}}(d) \triangleq \frac{1}{N} \sum_{n=1}^{N} \log \frac{q_p(\theta_n|y_n, d)}{p(\theta_n)}, \tag{6}$$

where $y_n, \theta_n \overset{\text{i.i.d.}}{\sim} p(y, \theta|d)$ and $\hat{\mu}_{\text{post}}(d)$ is a MC estimator of $\mathcal{L}_{\text{post}}(d)$. We draw samples of $p(y, \theta|d)$ by sampling $\theta \sim p(\theta)$ and then $y|\theta \sim p(y|\theta, d)$. We can think of this approach as amortizing the cost of the inner expectation, instead of running inference separately for each y.

To learn a suitable $q_p(\theta|y, d)$, we show in Appendix A that $\mathcal{L}_{\text{post}}(d)$ forms a variational lower bound $\text{EIG}(d) \geq \mathcal{L}_{\text{post}}(d)$ that is tight if and only if $q_p(\theta|y, d) = p(\theta|y, d)$. Barber and Agakov [3] used this bound to estimate mutual information in the context of transmission over noisy channels, but the connection to experiment design has not previously been made.

This result means we can learn $q_p(\theta|y, d)$ by introducing a family of variational distributions $q_p(\theta|y, d, \phi)$ parameterized by ϕ and then maximizing the bound with respect to ϕ:

$$\phi^* = \arg\max_{\phi} \mathbb{E}_{p(y,\theta|d)} \left[\log \frac{q_p(\theta|y, d, \phi)}{p(\theta)} \right], \qquad \text{EIG}(d) \approx \mathcal{L}_{\text{post}}(d; \phi^*). \tag{7}$$

Provided that we can generate samples from the model, this maximization can be performed using stochastic gradient methods [35] and the unbiased gradient estimator

$$\nabla_\phi \mathcal{L}_{\text{post}}(d; \phi) \approx \tfrac{1}{S} \sum\nolimits_{i=1}^{S} \nabla_\phi \log q_p(\theta_i|y_i, d, \phi) \quad \text{where} \quad y_i, \theta_i \overset{\text{i.i.d.}}{\sim} p(y, \theta|d), \tag{8}$$

and we note that no reparameterization is required as $p(y, \theta|d)$ is independent of ϕ. After K gradient steps we obtain variational parameters ϕ_K that approximate ϕ^*, which we use to compute

a corresponding EIG estimator by constructing a MC estimator for $\mathcal{L}_{\text{post}}(d;\phi)$ as per (6) with $q_p(\theta_n|y_n,d) = q_p(\theta_n|y_n,d,\phi_K)$. Interestingly, the tightness of $\mathcal{L}_{\text{post}}(d)$ turns out to be equal to the expected *forward* KL divergence[3] $\mathbb{E}_{p(y|d)}\left[\text{KL}\left(p(\theta|y,d)||q_p(\theta|y,d,\phi)\right)\right]$ so we can view this approach as learning an amortized proposal by minimizing this expected KL divergence.

Variational marginal $\hat{\mu}_{\text{marg}}$ In some scenarios, θ may be high-dimensional, making it difficult to train a good variational posterior approximation. An alternative approach that can be attractive in such cases is to instead learn an approximation $q_m(y|d)$ to the marginal density $p(y|d)$ and substitute this into the final form of the EIG in (3). As shown in Appendix A, this yields an *upper bound*

$$\text{EIG}(d) \leq \mathcal{U}_{\text{marg}}(d) \triangleq \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta,d)}{q_m(y|d)}\right] \approx \hat{\mu}_{\text{marg}}(d) \triangleq \frac{1}{N}\sum_{n=1}^{N} \log \frac{p(y_n|\theta_n,d)}{q_m(y_n|d)}, \quad (9)$$

where again $y_n, \theta_n \overset{\text{i.i.d.}}{\sim} p(y,\theta|d)$ and the bound is tight when $q_m(y|d) = p(y|d)$. Analogously to $\hat{\mu}_{\text{post}}$, we can learn $q_m(y|d)$ by introducing a variational family $q_m(y|d,\phi)$ and then performing stochastic gradient descent to *minimize* $\mathcal{U}_{\text{marg}}(d,\phi)$. As with $\hat{\mu}_{\text{post}}$, this bound was studied in a mutual information context [31], but it has not been utilized for BOED before.

Variational NMC $\hat{\mu}_{\text{VNMC}}$ As we will show in Section 4, $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{marg}}$ can provide substantially faster convergence rates than NMC. However, this comes at the cost of converging towards a biased estimate if the variational family does not contain the target distribution. To address this, we propose another EIG estimator, $\hat{\mu}_{\text{VNMC}}$, which allows one to trade-off resources between the fast learning of a biased estimator permitted by variational approaches, and the ability of NMC to eliminate this bias.[4]

We can think of the NMC estimator as approximating $p(y|d)$ using M samples from the prior. At a high-level, $\hat{\mu}_{\text{VNMC}}$ is based around learning a proposal $q_v(\theta|y,d)$ and then using samples from this proposal to make an importance sampling estimate of $p(y|d)$, potentially requiring far fewer samples than NMC. Formally, it is based around a bound that can be arbitrarily tightened, namely

$$\text{EIG}(d) \leq \mathbb{E}\left[\log p(y|\theta_0,d) - \log \frac{1}{L}\sum_{\ell=1}^{L} \frac{p(y,\theta_\ell|d)}{q_v(\theta_\ell|y,d)}\right] \triangleq \mathcal{U}_{\text{VNMC}}(d,L) \quad (10)$$

where the expectation is taken over $y, \theta_{0:L} \sim p(y,\theta_0|d)\prod_{\ell=1}^{L} q_v(\theta_\ell|y,d)$, which corresponds to one sample y,θ_0 from the model and L samples from the approximate posterior conditioned on y. To the best of our knowledge, this bound has not previously been studied in the literature. As with $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{marg}}$, we can minimize this bound to train a variational approximation $q_v(\theta|y,d,\phi)$. Important features of $\mathcal{U}_{\text{VNMC}}(d,L)$ are summarized in the following lemma; see Appendix A for the proof.

Lemma 1. *For any given model $p(\theta)p(y|\theta,d)$ and valid $q_v(\theta|y,d)$,*

1. $\text{EIG}(d) = \lim_{L\to\infty} \mathcal{U}_{VNMC}(d,L) \leq \mathcal{U}_{VNMC}(d,L_2) \leq \mathcal{U}_{VNMC}(d,L_1) \quad \forall L_2 \geq L_1 \geq 1$,
2. $\mathcal{U}_{VNMC}(d,L) = \text{EIG}(d) \;\; \forall L \geq 1 \quad$ *if* $\quad q_v(\theta|y,d) = p(\theta|y,d) \;\; \forall y,\theta$,
3. $\mathcal{U}_{VNMC}(d,L) - \text{EIG}(d) = \mathbb{E}_{p(y|d)}\left[\text{KL}\left(\prod_{\ell=1}^{L} q_v(\theta_\ell|y,d) \middle|\middle| \frac{1}{L}\sum_{\ell=1}^{L} p(\theta_\ell|y,d)\prod_{k\neq\ell} q_v(\theta_k|y,d)\right)\right]$

Like the previous bounds, the VNMC bound is tight when $q_v(\theta|y,d) = p(\theta|y,d)$. Importantly, the bound is also tight as $L \to \infty$, even for imperfect q_v. This means we can obtain asymptotically unbiased EIG estimates even when the true posterior is not contained in the variational family.

Specifically, we first train ϕ using K steps of stochastic gradient on $\mathcal{U}_{\text{VNMC}}(d,L)$ with some fixed L. To form a final EIG estimator, however, we use a MC estimator of $\mathcal{U}_{\text{VNMC}}(d,M)$ where typically $M \gg L$. This final estimator is a NMC estimator that is consistent as $N, M \to \infty$ with ϕ_K fixed

$$\hat{\mu}_{\text{VNMC}}(d) \triangleq \frac{1}{N}\sum_{n=1}^{N}\left(\log p(y_n|\theta_{n,0},d) - \log \frac{1}{M}\sum_{m=1}^{M} \frac{p(y_n,\theta_{n,m}|d)}{q_v(\theta_{n,m}|y_n,d,\phi_K)}\right) \quad (11)$$

where $\theta_{n,0} \overset{\text{i.i.d.}}{\sim} p(\theta)$, $y_n \sim p(y|\theta=\theta_{n,0},d)$ and $\theta_{n,m} \sim q_v(\theta|y=y_n,d,\phi_K)$. In practice, performance is greatly enhanced when the proposal q_v is a good, if inexact, approximation to the posterior. This significantly improves upon traditional $\hat{\mu}_{\text{NMC}}$, which sets $q_v(\theta|y,d) = p(\theta)$ in (11).

[3] See Appendix A for a proof. A comparison with the reverse KL divergence can be found in Appendix G.

[4] In Appendix F, we describe a method using $q_m(y|d)$ as a control variate that can also eliminate this bias and lower the variance of NMC, requiring additional assumptions about the model and variational family.

Implicit likelihood and $\hat{\mu}_{\mathbf{m}+\ell}$ So far we have assumed that we can evaluate $p(y|\theta, d)$ pointwise. However, many models of interest have *implicit likelihoods* from which we can draw samples, but not evaluate directly. For example, models with nuisance latent variables ψ (such as a random effect models) are implicit likelihood models because $p(y|\theta, d) = \mathbb{E}_{p(\psi|\theta)}[p(y|\theta, \psi, d)]$ is intractable, but can still be straightforwardly sampled from.

In this setting, $\hat{\mu}_{\text{post}}$ is applicable without modification because it only requires samples from $p(y|\theta, d)$ and *not* evaluations of this density. Although $\hat{\mu}_{\text{marg}}$ is not directly applicable in this setting, it can be modified to accommodate implicit likelihoods. Specifically, we can utilize *two* approximate densities: $q_m(y|d)$ for the marginal and $q_\ell(y|\theta, d)$ for the likelihood. We then form the approximation

$$\text{EIG}(d) \approx \mathcal{I}_{\text{m}+\ell}(d) \triangleq \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{q_\ell(y|\theta, d)}{q_m(y|d)}\right] \approx \hat{\mu}_{\text{m}+\ell}(d) \triangleq \frac{1}{N}\sum_{n=1}^{N} \log \frac{q_\ell(y_n|\theta_n, d)}{q_m(y_n|d)}. \quad (12)$$

Unlike the previous three cases, $\mathcal{I}_{\text{m}+\ell}(d)$ is not a bound on $\text{EIG}(d)$, meaning it is not immediately clear how to train $q_m(y|d)$ and $q_\ell(y|\theta, d)$ to achieve an accurate EIG estimator. The following lemma shows that we can bound the EIG estimation *error* of $\mathcal{I}_{\text{m}+\ell}$. The proof is in Appendix A.

Lemma 2. *For any given model $p(\theta)p(y|\theta, d)$ and valid $q_m(y|d)$ and $q_\ell(y|\theta, d)$, we have*

$$|\mathcal{I}_{m+\ell}(d) - \text{EIG}(d)| \leq -\mathbb{E}_{p(y,\theta|d)}[\log q_m(y|d) + \log q_\ell(y|\theta, d)] + C, \quad (13)$$

where $C = -H[p(y|d)] - \mathbb{E}_{p(\theta)}[H(p(y|\theta, d)]$ does not depend on q_m or q_ℓ. Further, the RHS of (13) is 0 if and only if $q_m(y|d) = p(y|d)$ and $q_\ell(y|\theta, d) = p(y|\theta, d)$ for almost all y, θ.

This lemma implies that we can learn $q_m(y|d)$ and $q_\ell(y|\theta, d)$ by maximizing $\mathbb{E}_{p(y,\theta|d)}[\log q_m(y|d) + \log q_\ell(y|\theta, d)]$ using stochastic gradient ascent, and substituting these learned approximations into (12) for the final EIG estimator. To the best of our knowledge, this approach has not previously been considered in the literature. We note that, in general, q_m and q_ℓ are learned separately and there need not be any weight sharing between them. See Appendix A.4 for a discussion of the case when we couple q_m and q_ℓ so that $q_m(y|d) = \mathbb{E}_{p(\theta)}[q_\ell(y|\theta, d)]$.

Using estimators for sequential BOED In sequential settings, we also need to consider the implications of replacing $p(\theta)$ in the EIG with $p(\theta|d_{1:t-1}, y_{1:t-1})$. At first sight, it appears that, while $\hat{\mu}_{\text{marg}}$ and $\hat{\mu}_{\text{m}+\ell}$ only require samples from $p(\theta|d_{1:t-1}, y_{1:t-1})$, $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{VNMC}}$ also require its density to be evaluated, a potentially severe limitation. Fortunately, we can, in fact, avoid evaluating this posterior density. We note that, from (5), we have $p(\theta|y_{1:t-1}, d_{1:t-1}) = p(\theta)\prod_{i=1}^{t-1} p(y_i|\theta, d_i)/p(y_{1:t-1}|d_{1:t-1})$. Substituting this into the integrand of (6) gives

$$\mathcal{L}_{\text{post}}(d_t) = \mathbb{E}_{p(\theta|y_{1:t-1},d_{1:t-1})p(y_t|\theta,d_t)}\left[\log \frac{q_p(\theta|y_t, d_t)}{p(\theta)\prod_{i=1}^{t-1} p(y_i|\theta, d_i)}\right] + \log p(y_{1:t-1}|d_{1:t-1}) \quad (14)$$

where $p(\theta)\prod_{i=1}^{t-1} p(y_i|\theta, d_i)$ can be evaluated exactly and the additive constant $\log p(y_{1:t-1}|d_{1:t-1})$ does not depend on the new design d_t, θ, or any of the variational parameters, and so can be safely ignored. Making the same substitution in (11) shows that we can also estimate $\mathcal{U}_{\text{VNMC}}(d_t, L)$ up to a constant, which can then be similarly ignored. As such, any inference scheme for sampling $p(\theta|d_{1:t-1}, y_{1:t-1})$, approximate or exact, is compatible with all our approaches.

Selecting an estimator Having proposed four estimators, we briefly discuss how to choose between them in practice. For reference, a summary of our estimators is given in Table 1, along with several baseline approaches. First, $\hat{\mu}_{\text{marg}}$ and $\hat{\mu}_{\text{m}+\ell}$ rely on approximating a distribution over y; $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{VNMC}}$ approximate distributions over θ. We may prefer the former two estimators if $\dim(y) \ll \dim(\theta)$ as it leaves us with a simpler density estimation problem, and vice versa. Second, $\hat{\mu}_{\text{marg}}$ and $\hat{\mu}_{\text{VNMC}}$ require an

Table 1: Summary of EIG estimators. Baseline methods are explained in Section 5.

		Implicit	Bound	Consistent	Eq.
Ours	$\hat{\mu}_{\text{post}}$	✓	Lower	✗	(6)
	$\hat{\mu}_{\text{marg}}$	✗	Upper	✗	(9)
	$\hat{\mu}_{\text{VNMC}}$	✗	Upper	✓	(11)
	$\hat{\mu}_{\text{m}+\ell}$	✓	✗	✗	(12)
Baseline	$\hat{\mu}_{\text{NMC}}$	✗	Upper	✓	(4)
	$\hat{\mu}_{\text{laplace}}$	✗	✗	✗	(75)
	$\hat{\mu}_{\text{LFIRE}}$	✓	✗	✗	(76)
	$\hat{\mu}_{\text{DV}}$	✓	Lower	✗	(77)

explicit likelihood whereas $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{m}+\ell}$ do not. If an explicit likelihood is available, it typically makes sense to use it—one would never use $\hat{\mu}_{\text{m}+\ell}$ over $\hat{\mu}_{\text{marg}}$ for example. Finally, if the variational families do not contain the target densities, $\hat{\mu}_{\text{VNMC}}$ is the only method guaranteed to converge to the true $\text{EIG}(d)$ in the limit as the computational budget increases. So we might prefer $\hat{\mu}_{\text{VNMC}}$ when computation time and cost are not constrained.

4 Convergence rates

We now investigate the convergence of our estimators. We start by breaking the overall error down into three terms: I) variance in MC estimation of the bound; II) the gap between the bound and the tightest bound possible given the variational family; and III) the gap between the tightest possible bound and $\text{EIG}(d)$. With variational EIG approximation $\mathcal{B}(d) \in \{\mathcal{L}_{\text{post}}(d), \mathcal{U}_{\text{marg}}(d), \mathcal{U}_{\text{VNMC}}(d, L), \mathcal{I}_{\text{m}+\ell}(d)\}$, optimal variational parameters ϕ^*, learned variational parameters ϕ_K after K stochastic gradient iterations, and MC estimator $\hat{\mu}(d, \phi_K)$ we have, by the triangle inequality,

$$\|\hat{\mu}(d,\phi_K) - \text{EIG}(d)\|_2 \leq \underbrace{\|\hat{\mu}(d,\phi_K) - \mathcal{B}(d,\phi_K)\|_2}_{\text{I}} + \underbrace{\|\mathcal{B}(d,\phi_K) - \mathcal{B}(d,\phi^*)\|_2}_{\text{II}} + \underbrace{|\mathcal{B}(d,\phi^*) - \text{EIG}(d)|}_{\text{III}}$$

where we have used the notation $\|X\|_2 \triangleq \sqrt{\mathbb{E}[X^2]}$ to denote the L^2 norm of a random variable.

By the weak law of large numbers, term I scales as $N^{-1/2}$ and can thus be arbitrarily reduced by taking more MC samples. Provided that our stochastic gradient scheme converges, term II can be reduced by increasing the number of stochastic gradient steps K. Term III, however, is a constant that can only be reduced by expanding the variational family (or increasing L for $\hat{\mu}_{\text{VNMC}}$). Each approximation $\mathcal{B}(d)$ thus converges to a biased estimate of the $\text{EIG}(d)$, namely $\mathcal{B}(d, \phi^*)$. As established by the following Theorem, if we set $N \propto K$, the rate of convergence to this biased estimate is $\mathcal{O}(T^{-1/2})$, where T represents the total computational cost, with $T = \mathcal{O}(N + K)$.

Theorem 1. *Let $\mathcal{X}$ be a measurable space and Φ be a convex subset of a finite dimensional inner product space. Let $X_1, X_2, \ldots$ be i.i.d. random variables taking values in $\mathcal{X}$ and $f : \mathcal{X} \times \Phi \to \mathbb{R}$ be a measurable function. Let*

$$\mu(\phi) \triangleq \mathbb{E}[f(X_1, \phi)] \approx \hat{\mu}_N(\phi) \triangleq \frac{1}{N} \sum\nolimits_{n=1}^{N} f(X_n, \phi)$$

and suppose that $\sup_{\phi \in \Phi} \|f(X_1, \phi)\|_2 < \infty$. Then $\sup_{\phi \in \Phi} \|\hat{\mu}_N(\phi) - \mu(\phi)\|_2 = \mathcal{O}(N^{-1/2})$. Suppose further that Assumption 1 in Appendix B holds and that ϕ^* is the unique minimizer of μ. After K iterations of the Polyak-Ruppert averaged stochastic gradient descent algorithm of [28] with gradient estimator $\nabla_\phi f(X_t, \phi)$, we have $\|\mu(\phi_K) - \mu(\phi^*)\|_2 = \mathcal{O}(K^{-1/2})$ and, combining with the first result,

$$\|\hat{\mu}_N(\phi_K) - \mu(\phi^*)\|_2 = \mathcal{O}(N^{-1/2} + K^{-1/2}) = \mathcal{O}(T^{-1/2}) \text{ if } N \propto K.$$

The proof relies on standard results from MC and stochastic optimization theory; see Appendix B. We note that the assumptions required for the latter, though standard in the literature, are strong. In practice, ϕ can converge to a local optimum $\phi^\dagger$, rather than the global optimum ϕ^*, introducing an additional asymptotic bias $\left|\mathcal{B}(d, \phi^\dagger) - \mathcal{B}(d, \phi^*)\right|$ into term III.

Theorem 1 can be applied directly to $\hat{\mu}_{\text{marg}}$, $-\hat{\mu}_{\text{post}}$, and $\hat{\mu}_{\text{VNMC}}$ (with fixed $M = L$), showing that they converge respectively to $\mathcal{U}_{\text{marg}}(d, \phi^*)$, $-\mathcal{L}_{\text{post}}(d, \phi^*)$, and $\mathcal{U}_{\text{VNMC}}(d, L, \phi^*)$ at a rate $= \mathcal{O}(T^{-1/2})$ if $N \propto K$ and the assumptions are satisfied. For $\hat{\mu}_{\text{m}+\ell}$, we combine Theorem 1 and Lemma 2 to obtain the same $\mathcal{O}(T^{-1/2})$ convergence rates; see the supplementary material for further details.

The key property of $\hat{\mu}_{\text{VNMC}}$ is that we need not set $M = L$ and can remove the asymptotic bias by increasing M with N. We begin by training ϕ with a fixed value of L, decreasing the error term $\|\mathcal{U}_{\text{VNMC}}(d, L, \phi_K) - \mathcal{U}_{\text{VNMC}}(d, L, \phi^*)\|_2$ at the fast rate $\mathcal{O}(K^{-1/2})$ until $|\mathcal{U}_{\text{VNMC}}(d, L, \phi^*) - \text{EIG}(d)|$ becomes the dominant error term. At this point, we start to increase N, M. Using the NMC convergence results discussed in Sec. 2, if we set $M \propto \sqrt{N}$, then $\hat{\mu}_{\text{VNMC}}$ converges to $\text{EIG}(d)$ at a rate $\mathcal{O}((NM)^{-1/3})$. Note that the total cost of the $\hat{\mu}_{\text{VNMC}}$ estimator is $T = \mathcal{O}(KL + NM)$, where typically $M \gg L$. The first stage, costing KL, is fast variational training of an amortized importance sampling proposal for $p(y|d) = \mathbb{E}_{p(\theta)}[p(y|\theta, d)]$. The second stage, costing NM, is slower refinement to remove the asymptotic bias using the learned proposal in an NMC estimator.

Table 2: Bias squared and variance from 5 runs, averaged over designs, of EIG estimators applied to four benchmarks. We use - to denote that a method does not apply and $*$ when it is superseded by other methods. Bold indicates the estimator with the lowest empirical mean squared error.

	A/B test		**Preference**		**Mixed effects**		**Extrapolation**	
	Bias^2	Var	Bias^2	Var	Bias^2	Var	Bias^2	Var
$\hat{\mu}_{\text{post}}$	1.33×10^{-2}	7.15×10^{-3}	4.26×10^{-2}	8.53×10^{-3}	2.34×10^{-3}	2.92×10^{-3}	1.24×10^{-4}	5.16×10^{-5}
$\hat{\mu}_{\text{marg}}$	7.45×10^{-2}	6.41×10^{-3}	$\mathbf{1.10\times10^{-3}}$	$\mathbf{1.99\times10^{-3}}$	-	-	-	-
$\hat{\mu}_{\text{VNMC}}$	3.44×10^{-3}	3.38×10^{-3}	4.17×10^{-3}	9.04×10^{-3}	-	-	-	-
$\hat{\mu}_{\text{m}+\ell}$	$*$	$*$	$*$	$*$	$\mathbf{3.06\times10^{-3}}$	$\mathbf{5.94\times10^{-5}}$	$\mathbf{6.90\times10^{-6}}$	$\mathbf{1.84\times10^{-5}}$
$\hat{\mu}_{\text{NMC}}$	4.70×10^{0}	3.47×10^{-1}	7.60×10^{-2}	8.36×10^{-2}	-	-	-	-
$\hat{\mu}_{\text{laplace}}$	$\mathbf{1.92\times10^{-4}}$	$\mathbf{1.47\times10^{-3}}$	8.42×10^{-2}	9.70×10^{-2}	-	-	-	-
$\hat{\mu}_{\text{LFIRE}}$	2.29×10^{0}	6.20×10^{-1}	1.30×10^{-1}	1.41×10^{-2}	1.41×10^{-1}	6.67×10^{-2}	-	-
$\hat{\mu}_{\text{DV}}$	4.34×10^{0}	8.85×10^{-1}	9.23×10^{-2}	8.07×10^{-3}	9.10×10^{-3}	5.56×10^{-4}	7.84×10^{-6}	4.11×10^{-5}

One can think of the standard NMC approach as a special case of $\hat{\mu}_{\text{VNMC}}$ in which we naively choose $p(\theta)$ as the proposal. That is, standard NMC skips the first stage and hence does not benefit from the improved convergence rate of learning an amortized proposal. It typically requires a much higher total cost to achieve the same accuracy as VNMC.

5 Related work

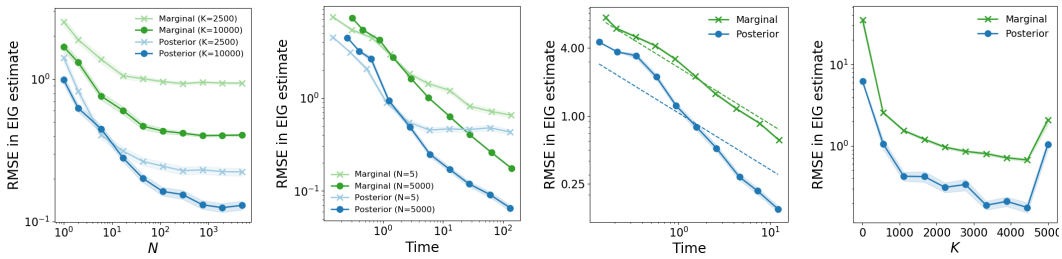
We briefly discuss alternative approaches to EIG estimation for BOED that will form our baselines for empirical comparisons. The **Nested Monte Carlo (NMC)** baseline was introduced in Sec. 2. Another established approach is to use a **Laplace approximation** to the posterior [22, 25]; this approach is fast but is limited to continuous variables and can exhibit large bias. Kleinegesse and Gutmann [18] recently suggested an implicit likelihood approach based on the Likelihood-Free Inference by Ratio Estimation **(LFIRE)** method of Thomas et al. [41]. We also consider a method based on the **Donsker-Varadhan (DV)** representation of the KL divergence [11] as used by Belghazi et al. [4] for mutual information estimation. Though not previously considered in BOED, we include it as a baseline for illustrative purposes. For a full discussion of the DV bound and a number of other variational bounds used in deep learning, we refer to the recent work of Poole et al. [31]. For further discussion of related work, see Appendix C.

6 Experiments

6.1 EIG estimation accuracy

We begin by benchmarking our EIG estimators against the aforementioned baselines. We consider four experiment design scenarios inspired by applications of Bayesian data analysis in science and industry. First, **A/B testing** is used across marketing and design [6, 19] to study population traits. Here, the design is the choice of the A and B group sizes and the Bayesian model is a Gaussian linear model. Second, revealed **preference** [36] is used in economics to understand consumer behaviour. We consider an experiment design setting in which we aim to learn the underlying utility function of an economic agent by presenting them with a proposal (such as offering them a price for a commodity) and observing their revealed preference. Third, fixed effects and random effects (nuisance variables) are combined in **mixed effects** models [14, 20]. We consider an example inspired by item-response theory [13] in psychology. We seek information only about the fixed effects, making this an implicit likelihood problem. Finally, we consider an experiment where labelled data from one region of design space must be used to predict labels in a target region by **extrapolation** [27]. In summary, we have two models with explicit likelihoods (A/B testing, preference) and two that are implicit (mixed effects, extrapolation). Full details of each model are presented in Appendix D.

For each scenario, we estimated the EIG across a grid of designs with a fixed computational budget for each estimator and calculated the true EIG analytically or with brute force computation as appropriate; see Table 2 for the results. Whilst the Laplace method, unsurprisingly, performed best for the Gaussian linear model where its approximation becomes exact, we see that our methods are otherwise more accurate. All our methods outperformed NMC.



(a) Convergence in N (b) Convergence in K (c) Convergence $N = K$ (d) Fixed budget $N + K$

Figure 1: Convergence of RMSE for $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{marg}}$. (a) Convergence in number of MC samples N with a fixed number K of gradient updates of the variational parameters. (b) Convergence in time when increasing K and with N fixed. (c) Convergence in time when setting $N = K$ and increasing both (dashed lines represent theoretical rates). (d) Final RMSE with $N + K = 5000$ fixed, for different K. Each graph shows the mean with shading representing ± 1 std. err. from 100 trials.

6.2 Convergence rates

We now investigate the empirical convergence characteristics of our estimators. Throughout, we consider a single design point from the A/B test example. We start by examining the convergence of $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{marg}}$ as we allocate the computational budget in different ways.

We first consider the convergence in N after a fixed number of K updates to the variational parameters. As shown in Figure 1a, the RMSE initially decreases as we increase N, before plateauing due to the bias in the estimator. We also see that $\hat{\mu}_{\text{post}}$ substantially outperforms $\hat{\mu}_{\text{marg}}$. We next consider the convergence as a function of wall-clock time when N is held fixed and we increase K. We see in Figure 1b that, as expected, the errors decrease with time and that when a small value of $N = 5$ is taken, we again see a plateauing effect, with the variance of the final MC estimator now becoming the limiting factor. In Figure 1c we take $N = K$ and increase both, obtaining the predicted convergence rate $\mathcal{O}(T^{-1/2})$ (shown by the dashed lines). We conjecture that the better performance of $\hat{\mu}_{\text{post}}$ is likely due to θ being lower dimensional ($\dim = 2$) than y ($\dim = 10$). In Figure 1d, we instead fix $T = N + K$ to investigate the optimal trade-off between optimization and MC error: it appears the range of K/T between 0.5 and 0.9 gives the lowest RMSE.

Finally, we show how $\hat{\mu}_{\text{VNMC}}$ can improve over NMC by using an improved variational proposal for estimating $p(y|d)$. In Figure 2, we plot the EIG estimates obtained by first running K steps of stochastic gradient with $L = 1$ to learn $q_v(\theta|y, d)$, before increasing M and N. We see that spending some of our time budget training $q_v(\theta|y, d)$ leads to noticeable improvements in the estimation, but also that it is important to increase N and M. Rather than plateauing like $\hat{\mu}_{\text{post}}$ and $\hat{\mu}_{\text{marg}}$, $\hat{\mu}_{\text{VNMC}}$ continues to improve after the initial training period as, albeit at a slower $O(T^{-1/3})$ rate.

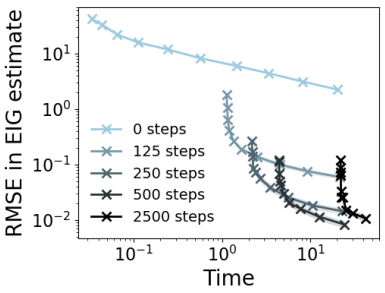
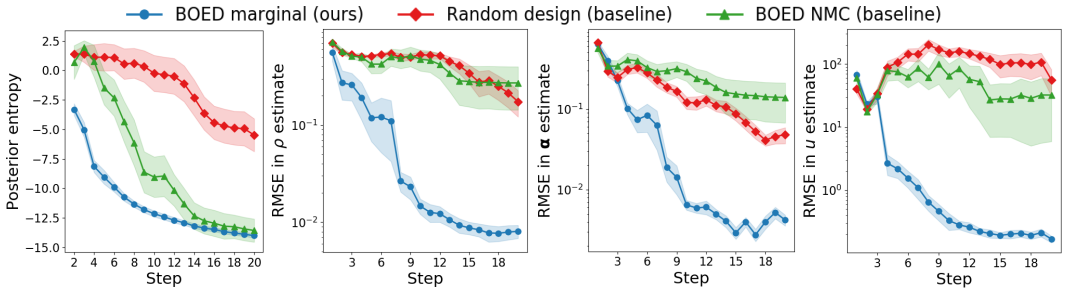


Figure 2: Convergence of $\hat{\mu}_{\text{VNMC}}$ taking $M = \sqrt{N}$. 'Steps' refers to pre-training of the variational posterior (i.e. K), with 0 steps corresponding to $\hat{\mu}_{\text{NMC}}$. Means and confidence intervals as per Fig. 1.

6.3 End-to-end sequential experiments

We now demonstrate the utility of our methods for designing sequential experiments. First, we demonstrate that our variational estimators are sufficiently robust and fast to be used for adaptive experiments with a class of models that are of practical importance in many scientific disciplines. To this end, we run an adaptive psychology experiment with human participants recruited from Amazon Mechanical Turk to study how humans respond to features of stylized faces. To account for fixed effects—those *common* across the population—as well as individual variations that we treat as nuisance variables, we use the mixed effects regression model introduced in Sec. 6.1. See Appendix D for full details of the experiment.

To estimate the EIG for different designs, we use $\hat{\mu}_{\text{m}+\ell}$, since it yields the best performance on our mixed effects model benchmark (see Table 2). Our EIG estimator is integrated into a system that



(a) Entropy (b) Posterior RMSE of ρ (c) Posterior RMSE of $\boldsymbol{\alpha}$ (d) Posterior RMSE of u

Figure 4: Evolution of the posterior in the sequential CES experiment. (a) Total entropy of a mean-field variational approximation of the posterior. (b)(c)(d) The RMSE of the posterior approximations of ρ, $\boldsymbol{\alpha}$ and u as compared to the true values used to simulate agent responses. Note the scale of the vertical axis is logarithmic. All plots show the mean and ± 1 std. err. from 10 independent runs.

presents participants with a stimulus, receives their response, learns an updated model, and designs the next stimulus, all online. Despite the relative simplicity of the design problem (with 36 possible designs) using BOED with $\hat{\mu}_{\mathrm{m}+\ell}$ leads to a more certain (i.e. lower entropy) posterior than random design; see Figure 3.

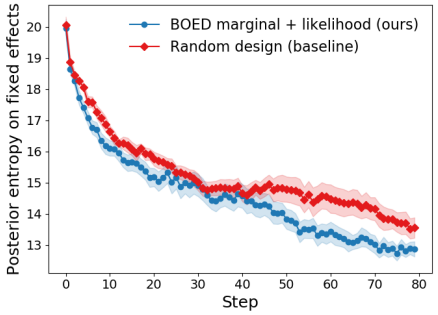


Figure 3: Evolution of the posterior entropy of the fixed effects in the Mechanical Turk experiment in Sec. 6.3. We depict the mean and ± 1 std. err. from 10 experimental trials.

Second, we consider a more challenging scenario in which a random design strategy gleans very little. We compare random design against two BOED strategies: $\hat{\mu}_{\mathrm{marg}}$ and $\hat{\mu}_{\mathrm{NMC}}$. Building on the revealed preference example in Sec. 6.1, we consider an experiment to infer an agent's utility function which we model using the Constant Elasticity of Substitution (CES) model [2] with latent variables $\rho, \boldsymbol{\alpha}, u$. We seek designs for which the agent's response will be informative about $\theta = (\rho, \boldsymbol{\alpha}, u)$. See Appendix D for full details. We estimate the EIG using $\hat{\mu}_{\mathrm{marg}}$ because the dimension of y is smaller than that of θ, and select designs $d \in [0, 100]^6$ using Bayesian optimization. To investigate parameter recovery we simulate agent responses from the model with fixed values of $\rho, \boldsymbol{\alpha}, u$. Figure 4 shows that using BOED with our marginal estimator reduces posterior entropy *and* concentrates more quickly on the true parameter values than both baselines. Random design makes no inroads into the learning problem, while BOED based on NMC particularly struggles at the outset when $p(\theta|d_{1:t-1}, y_{1:t-1})$, the prior at iteration t, is high variance. Our method selects informative designs throughout.

7 Discussion

We have developed efficient EIG estimators that are applicable to a wide range of experimental design problems. By tackling the double intractability of the EIG in a principled manner, they provide substantially improved convergence rates relative to previous approaches, and our experiments show that these theoretical advantages translate into significant practical gains. Our estimators are well-suited to modern deep probabilistic programming languages and we have provided an implementation in Pyro. We note that the interplay between variational and MC methods in EIG estimation is not directly analogous to those in standard inference settings because the NMC EIG estimator is itself inherently biased. Our $\hat{\mu}_{\mathrm{VNMC}}$ estimator allows one to play off the advantages of these approaches, namely the fast learning of variational approaches and asymptotic consistency of NMC.

Acknowledgements

We gratefully acknowledge research funding from Uber AI Labs. MJ would like to thank Paul Szerlip for help generating the sprites used in the Mechanical Turk experiment. AF would like to thank Patrick Rebeschini, Dominic Richards and Emile Mathieu for their help and support. AF gratefully acknowledges funding from EPSRC grant no. EP/N509711/1. YWT's and TR's research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071.

References

[1] Billy Amzal, Frédéric Y Bois, Eric Parent, and Christian P Robert. Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical association*, 101(474):773–785, 2006.

[2] Kenneth J Arrow, Hollis B Chenery, Bagicha S Minhas, and Robert M Solow. Capital-labor substitution and economic efficiency. *The review of Economics and Statistics*, pages 225–250, 1961.

[3] David Barber and Felix Agakov. The IM algorithm: a variational approach to information maximization. *Advances in Neural Information Processing Systems*, 16:201–208, 2003.

[4] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Devon Hjelm, and Aaron Courville. Mutual information neural estimation. In *International Conference on Machine Learning*, pages 530–539, 2018.

[5] Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1): 973–978, 2019.

[6] George EP Box, J Stuart Hunter, and William G Hunter. Statistics for experimenters. In *Wiley Series in Probability and Statistics*. Wiley Hoboken, NJ, 2005.

[7] Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders. In *4th International Conference on Learning Representations, ICLR*, 2016.

[8] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

[9] Alex R Cook, Gavin J Gibson, and Christopher A Gilligan. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

[10] Peter Dayan, Geoffrey E Hinton, Radford M Neal, and Richard S Zemel. The Helmholtz machine. *Neural computation*, 7(5):889–904, 1995.

[11] Monroe D Donsker and SR Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. *Communications on Pure and Applied Mathematics*, 28(1): 1–47, 1975.

[12] Sylvain Ehrenfeld. Some experimental design problems in attribute life testing. *Journal of the American Statistical Association*, 57(299):668–679, 1962.

[13] Susan E Embretson and Steven P Reise. *Item response theory*. Psychology Press, 2013.

[14] Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[15] Daniel Golovin, Andreas Krause, and Debajyoti Ray. Near-optimal bayesian active learning with noisy observations. In *Advances in Neural Information Processing Systems*, pages 766–774, 2010.

[16] José Miguel Hernández-Lobato, Matthew W Hoffman, and Zoubin Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *Advances in neural information processing systems*, pages 918–926, 2014.

[17] Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. In *ICLR*, 2014.

[18] Steven Kleinegesse and Michael U Gutmann. Efficient bayesian experimental design for implicit models. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 476–485, 2019.

[19] Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M Henne. Controlled experiments on the web: survey and practical guide. *Data mining and knowledge discovery*, 18(1): 140–181, 2009.

[20] John Kruschke. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2014.

[21] Tuan Anh Le, Maximilian Igl, Tom Rainforth, Tom Jin, and Frank Wood. Auto-Encoding Sequential Monte Carlo. *International Conference on Learning Representations (ICLR)*, 2018.

[22] Jeremy Lewi, Robert Butera, and Liam Paninski. Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21(3):619–687, 2009.

[23] Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.

[24] Dennis V Lindley. *Bayesian statistics, a review*, volume 2. SIAM, 1972.

[25] Quan Long, Marco Scavino, Raúl Tempone, and Suojin Wang. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39, 2013.

[26] Chao Ma, Sebastian Tschiatschek, Konstantina Palla, Jose Miguel Hernandez Lobato, Sebastian Nowozin, and Cheng Zhang. EDDI: Efficient dynamic discovery of high-value information with partial VAE. *arXiv preprint arXiv:1809.11142*, 2018.

[27] David JC MacKay. Information-based objective functions for active data selection. *Neural computation*, 4(4):590–604, 1992.

[28] Eric Moulines and Francis R Bach. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in Neural Information Processing Systems*, pages 451–459, 2011.

[29] Peter Müller. Simulation based optimal design. *Handbook of Statistics*, 25:509–518, 2005.

[30] Jay I Myung, Daniel R Cavagnaro, and Mark A Pitt. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, 2013.

[31] Ben Poole, Sherjil Ozair, Aäron van den Oord, Alexander A Alemi, and George Tucker. On variational lower bounds of mutual information. *NeurIPS Workshop on Bayesian Deep Learning*, 2018.

[32] Tom Rainforth. *Automating Inference, Learning, and Design using Probabilistic Programming*. PhD thesis, University of Oxford, 2017.

[33] Tom Rainforth, Robert Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting Monte Carlo estimators. In *International Conference on Machine Learning*, pages 4264–4273, 2018.

[34] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 1278–1286, 2014.

[35] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[36] Paul A Samuelson. Consumption theory in terms of revealed preference. *Economica*, 15(60): 243–253, 1948.

[37] Paola Sebastiani and Henry P Wynn. Maximum entropy sampling and optimal Bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1), 2000.

[38] Ben Shababo, Brooks Paige, Ari Pakman, and Liam Paninski. Bayesian inference and online experimental design for mapping neural microcircuits. In *Advances in Neural Information Processing Systems*, pages 1304–1312, 2013.

[39] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

[40] Andreas Stuhlmüller, Jacob Taylor, and Noah Goodman. Learning stochastic inverses. In *Advances in neural information processing systems*, pages 3048–3056, 2013.

[41] Owen Thomas, Ritabrata Dutta, Jukka Corander, Samuel Kaski, and Michael U Gutmann. Likelihood-free inference by ratio estimation. *arXiv preprint arXiv:1611.10242*, 2016.

[42] Joep Vanlier, Christian A Tiemann, Peter AJ Hilbers, and Natal AW van Riel. A Bayesian approach to targeted experiment design. *Bioinformatics*, 28(8):1136–1142, 2012.

[43] Benjamin T Vincent and Tom Rainforth. The DARC toolbox: automated, flexible, and efficient delayed and risky choice experiments using bayesian adaptive design. 2017.

A Details for variational estimators

The proofs in A.1 and A.2 are included for completeness.

A.1 Variational posterior $\hat{\mu}_{\text{post}}$

We require valid approximations $q_p(\theta|y,d)$ to have the same support as $p(\theta|y,d)$. Recall

$$\mathcal{L}_{\text{post}}(d) = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{q_p(\theta|y,d)}{p(\theta)}\right] \tag{16}$$

and

$$\text{EIG}(d) = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(\theta|y,d)}{p(\theta)}\right] \tag{17}$$

We aim to show $\text{EIG}(d) \geq \mathcal{L}_{\text{post}}(d)$. Following [3], we have

$$\begin{aligned}
\text{EIG}(d) - \mathcal{L}_{\text{post}}(d) =& \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(\theta|y,d)}{p(\theta)} - \log \frac{q_p(\theta|y,d)}{p(\theta)}\right] && (18)\\
=& \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(\theta|y,d)p(\theta)}{p(\theta)q_p(\theta|y,d)}\right] && (19)\\
=& \mathbb{E}_{p(y|d)}\left[\mathbb{E}_{p(\theta|y,d)}\left[\log \frac{p(\theta|y,d)}{q_p(\theta|y,d)}\right]\right] && (20)\\
=& \mathbb{E}_{p(y|d)}\left[\text{KL}\left(p(\theta|y,d)||q_p(\theta|y,d)\right)\right] && (21)\\
\geq& 0. && (22)
\end{aligned}$$

To further prove that the bound is tight, we note that the penultimate term $\mathbb{E}_{p(y|d)}\left[\text{KL}\left(p(\theta|y,d)||q_p(\theta|y,d)\right)\right]$ equals 0 if and only if $\text{KL}\left(p(\theta|y,d)||q_p(\theta|y,d)\right) = 0$ for almost all y (i.e. the union of all y for which this does not hold has measure zero). The occurs if and only if $q_p(\theta|y,d) = p(\theta|y,d)$ for almost all y, θ.

A.2 Variational marginal $\hat{\mu}_{\text{marg}}$

We now demonstrate that $\mathcal{U}_{\text{marg}}(d)$ is an upper bound on $\text{EIG}(d)$. Proceeding in the same manner as for $\hat{\mu}_{\text{post}}$, we find

$$\begin{aligned}
\mathcal{U}_{\text{marg}}(d) - \text{EIG}(d) =& \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta,d)}{q_m(y|d)} - \log \frac{p(y|\theta,d)}{p(y|d)}\right] && (23)\\
=& \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta,d)p(y|d)}{q_m(y|d)p(y|\theta,d)}\right] && (24)\\
=& \mathbb{E}_{p(y|d)}\left[\log \frac{p(y|d)}{q_m(y|d)}\right] && (25)\\
=& \text{KL}\left(p(y|d)||q_m(y|d)\right) && (26)\\
\geq& 0. && (27)
\end{aligned}$$

Again, the bound is tight if and only if $q_m(y|d) = p(y|d)$ almost everywhere.

A.3 Variational NMC $\hat{\mu}_{\text{VNMC}}$

We now prove Lemma 1 from the main paper, duplicating the Lemma itself below for convenience.

Lemma 1. *For any given model $p(\theta)p(y|\theta,d)$ and valid $q_v(\theta|y,d)$,*

1. $\text{EIG}(d) = \lim_{L\to\infty} \mathcal{U}_{VNMC}(d,L) \leq \mathcal{U}_{VNMC}(d,L_2) \leq \mathcal{U}_{VNMC}(d,L_1) \quad \forall L_2 \geq L_1 \geq 1$,
2. $\mathcal{U}_{VNMC}(d,L) = \text{EIG}(d) \quad \forall L \geq 1 \quad \textit{if} \quad q_v(\theta|y,d) = p(\theta|y,d) \quad \forall y,\theta$,
3. $\mathcal{U}_{VNMC}(d,L) - \text{EIG}(d) = \mathbb{E}_{p(y|d)}\left[\text{KL}\left(\prod_{\ell=1}^{L} q_v(\theta_\ell|y,d)\,\middle|\middle|\,\frac{1}{L}\sum_{\ell=1}^{L} p(\theta_\ell|y,d)\prod_{k\neq\ell} q_v(\theta_k|y,d)\right)\right]$

Proof. Starting with proving the first result in lemma, we first recall the definition of $\mathcal{U}_{\text{VNMC}}(d, L)$ itself,

$$\mathcal{U}_{\text{VNMC}}(d, L) = \mathbb{E}\left[\log p(y|\theta_0, d) - \log \frac{1}{L}\sum_{\ell=1}^{L} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right] \tag{28}$$

where the expectation is taken over $y, \theta_{0:L} \sim p(y, \theta_0|d)\prod_{\ell=1}^{L} q_v(\theta_\ell|y, d)$. We consider positive integers $L_2 \geq L_1$. We let $\delta = \mathcal{U}_{\text{VNMC}}(d, L_1) - \mathcal{U}_{\text{VNMC}}(d, L_2)$. Then,

$$\delta = \mathbb{E}\left[\log \frac{1}{L_2}\sum_{\ell=1}^{L_2} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right] - \mathbb{E}\left[\log \frac{1}{L_1}\sum_{\ell=1}^{L_1} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right]. \tag{29}$$

We now proceed as in [7]. Let $I_1, ..., I_{L_1}$ be distinct indices drawn uniformly from $1, ..., L_2$. Then,

$$\frac{1}{L_2}\sum_{\ell=1}^{L_2} \frac{p(y, \theta_\ell)}{q_v(\theta_\ell|y, d)} = \mathbb{E}_{I_1,...,I_{L_1}}\left[\frac{1}{L_1}\sum_{j=1}^{L_1} \frac{p(y, \theta_{I_j})}{q_v(\theta_{I_j}|y, d)}\right] \tag{30}$$

So

$$\delta = \mathbb{E}\left[\log\left(\mathbb{E}_{I_{1:L_1}}\left[\frac{1}{L_1}\sum_{j=1}^{L_1} \frac{p(y, \theta_{I_j})}{q_v(\theta_{I_j}|y, d)}\right]\right)\right] - \mathbb{E}\left[\log \frac{1}{L_1}\sum_{\ell=1}^{L_1} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right], \tag{31}$$

then by Jensen's Inequality

$$\delta \geq \mathbb{E}\left[\mathbb{E}_{I_{1:L_1}}\left[\log\left(\frac{1}{L_1}\sum_{j=1}^{L_1} \frac{p(y, \theta_{I_j})}{q_v(\theta_{I_j}|y, d)}\right)\right]\right] - \mathbb{E}\left[\log \frac{1}{L_1}\sum_{\ell=1}^{L_1} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right] \tag{32}$$

$$\geq \mathbb{E}\left[\log \frac{1}{L_1}\sum_{\ell=1}^{L_1} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right] - \mathbb{E}\left[\log \frac{1}{L_1}\sum_{\ell=1}^{L_1} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right] \tag{33}$$

$$\geq 0 \tag{34}$$

where we have used that $\theta_{I_1}, ..., \theta_{I_{L_1}} \overset{d}{=} \theta_1, ..., \theta_{L_1}$. This shows that $\mathcal{U}_{\text{VNMC}}(d, L_1) \geq \mathcal{U}_{\text{VNMC}}(d, L_2)$. For the limit $\lim_{L\to\infty} \mathcal{U}_{\text{VNMC}}(d, L)$ we first fix some y for which $p(y|d) > 0$ and consider

$$\mathcal{U}_{\text{VNMC}}(d, L, y) = \mathbb{E}\left[\log p(y|\theta_0, d) - \log \frac{1}{L}\sum_{\ell=1}^{L} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right]. \tag{35}$$

with the expectation taken over $p(\theta_0|y, d)\prod_{\ell=1}^{L} q_v(\theta_\ell|y, d)$. Since $p(y, \theta|d)/q_v(\theta|y, d)$ is bounded by assumption, the Strong Law of Large Numbers implies that, in limit of large L,

$$\frac{1}{L}\sum_{\ell=1}^{L} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)} \to p(y|d) \; a.s. \tag{36}$$

Furthermore, using the same argument as before, $\mathcal{U}_{\text{VNMC}}(d, L_1, y) \geq \mathcal{U}_{\text{VNMC}}(d, L_2, y)$ whenever $L_2 \geq L_1$. Thus the Bounded Convergence Theorem implies

$$\mathcal{U}_{\text{VNMC}}(d, L, y) \downarrow \mathbb{E}_{p(\theta_0|y,d)}[\log p(y|\theta_0, d) - \log p(y|d)] \text{ as } L \to \infty \tag{37}$$

so, taking expectations of $p(y|d)$, by the Monotone Convergence Theorem

$$\mathcal{U}_{\text{VNMC}}(d, L) \downarrow \mathbb{E}_{p(y,\theta_0|d)}[\log p(y|\theta_0, d) - \log p(y|d)] = \text{EIG}(d) \text{ as } L \to \infty. \tag{38}$$

For the second result, we simply note that

$$\frac{p(y, \theta|d)}{p(\theta|y, d)} = \frac{p(y, \theta|d)}{\frac{p(y,\theta|d)}{p(y|d)}} = p(y|d) \tag{39}$$

Finally, for the third result, we proceed as in [21]. We have

$$\mathcal{U}_{\text{VNMC}}(d, L) - \text{EIG}(d) = \mathbb{E}\left[\log p(y|d) - \log \frac{1}{L}\sum_{\ell=1}^{l} \frac{p(y, \theta_\ell|d)}{q_v(\theta_\ell|y, d)}\right] \tag{40}$$

where the expectation is over $p(y, \theta_0|d) \prod_{\ell=1}^{L} q_v(\theta_\ell|y, d)$.

Then

$$\mathcal{U}_{\text{VNMC}}(d, L) - \text{EIG}(d) = \mathbb{E}\left[-\log \frac{1}{L}\sum_{\ell=1}^{L} \frac{p(\theta_\ell|y, d)}{q_v(\theta_\ell|y, d)}\right] \tag{41}$$

$$= \mathbb{E}\left[\log \frac{\prod_{\ell=1}^{L} q_v(\theta_\ell|y, d)}{\frac{1}{L}\sum_{\ell=1}^{L} p(\theta_\ell|y, d) \prod_{k\neq\ell} q_v(\theta_k|y, d)}\right] \tag{42}$$

$$= \mathbb{E}\left[\log \frac{\prod_{\ell=1}^{L} q_v(\theta_\ell|y, d)}{P(\theta_{1:L}|y, d)}\right] \tag{43}$$

$$= \mathbb{E}_{p(y|d)}\left[\text{KL}\left(\prod_{\ell=1}^{L} q_v(\theta_\ell|y, d)||P(\theta_{1:L}|y, d)\right)\right] \tag{44}$$

where $P(\theta_{1:L}|y, d) = \frac{1}{L}\sum_{\ell=1}^{L} p(\theta_\ell|y, d) \prod_{k\neq\ell} q_v(\theta_k|y, d)$. □

A.4 Variational marginal + likelihood $\hat{\mu}_{\mathbf{m}+\ell}$

We now prove Lemma 2 from the main paper, duplicating the Lemma itself below for convenience.

Lemma 2. *For any given model $p(\theta)p(y|\theta, d)$ and valid $q_m(y|d)$ and $q_\ell(y|\theta, d)$, we have*

$$|\mathcal{I}_{m+\ell}(d) - \text{EIG}(d)| \leq -\mathbb{E}_{p(y,\theta|d)}[\log q_m(y|d) + \log q_\ell(y|\theta, d)] + C, \tag{13}$$

where $C = -H[p(y|d)] - \mathbb{E}_{p(\theta)}\left[H(p(y|\theta, d)\right]$ does not depend on q_m or q_ℓ. Further, the RHS of (13) is 0 if and only if $q_m(y|d) = p(y|d)$ and $q_\ell(y|\theta, d) = p(y|\theta, d)$ for almost all y, θ.

Proof. We aim to bound $|\mathcal{I}_{\text{m}+\ell}(d) - \text{EIG}(d)|$. Let $\delta = \mathcal{I}_{\text{m}+\ell}(d) - \text{EIG}(d)$. We have

$$\delta = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{q_\ell(y|\theta, d)}{q_m(y|d)}\right] - \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta, d)}{p(y|d)}\right] \tag{45}$$

$$= \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{q_\ell(y|\theta, d)}{q_m(y|d)} - \log \frac{p(y|\theta, d)}{p(y|d)}\right] \tag{46}$$

$$= \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{q_\ell(y|\theta, d)}{q_m(y|d)} - \log \frac{p(y|\theta, d)}{q_m(y|d)} + \log \frac{p(y|\theta, d)}{q_m(y|d)} - \log \frac{p(y|\theta, d)}{p(y|d)}\right] \tag{47}$$

$$= -\mathbb{E}_{p(y,\theta|d)}\left[\log \frac{q_m(y|d)p(y|\theta, d)}{q_\ell(y|\theta, d)q_m(y|d)}\right] + \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta, d)p(y|d)}{q_m(y|d)p(y|\theta, d)}\right] \tag{48}$$

$$= -\mathbb{E}_{p(\theta)}\left[\mathbb{E}_{p(y|\theta,d)}\left[\log \frac{p(y|\theta, d)}{q_\ell(y|\theta, d)}\right]\right] + \mathbb{E}_{p(y|d)}\left[\log \frac{p(y|d)}{q_m(y|d)}\right] \tag{49}$$

$$= -\mathbb{E}_{p(\theta)}\left[\text{KL}(p(y|\theta, d)||q_\ell(y|\theta, d))\right] + \text{KL}(p(y|d)||q_m(y|d)). \tag{50}$$

So, by the triangle inequality

$$|\delta| \leq \mathbb{E}_{p(\theta)}\left[\text{KL}(p(y|\theta, d)||q_\ell(y|\theta, d))\right] + \text{KL}(p(y|d)||q_m(y|d)). \tag{51}$$

We can rewrite the RHS using the following relation

$$\text{KL}(p(x)||q(x)) = \mathbb{E}_{p(x)}\left[\log \frac{p(x)}{q(x)}\right] \tag{52}$$

$$= \mathbb{E}_{p(x)}[\log p(x)] - \mathbb{E}_{p(x)}[\log q(x)] \tag{53}$$

$$= -H[p(x)] - \mathbb{E}_{p(x)}[\log q(x)]. \tag{54}$$

This gives us

$$|\delta| \leq \mathbb{E}_{p(\theta)}\left[-H(p(y|\theta, d)] - \mathbb{E}_{p(y,\theta|d)}[\log q_\ell(y|\theta, d)] - H[p(y|d)] - \mathbb{E}_{p(y,|d)}[\log q_m(y|d)]\right. \tag{55}$$

$$\leq -\mathbb{E}_{p(y,\theta|d)}[\log q_m(y|d) + \log q_\ell(y|\theta, d)] - H[p(y|d)] - \mathbb{E}_{p(\theta)}\left[H(p(y|\theta, d)\right] \tag{56}$$

as required.

Finally, from (51) we see that the error bound is tight if and only if both KL-divergences are 0 if and only if $q_\ell(y|\theta, d) = p(y|\theta, d)$ and $q_m(y|d) = p(y|d)$ for almost all y, θ. □

We conclude with an additional observation. Suppose that we set $q_m(y|d) = \mathbb{E}_{p(\theta)}[q_\ell(y|\theta, d)]$. This could be possible for instance when θ takes finitely many values. In this case, $\mathcal{I}_{\mathrm{m}+\ell}(d)$ is actually a lower bound on $\mathrm{EIG}(d)$. This is in contrast to the general case when q_m and q_ℓ are learned separately, in which it is neither an upper nor a lower bound.

To show that $\mathcal{I}_{\mathrm{m}+\ell}(d)$ is a lower bound when $q_m(y|d) = \mathbb{E}_{p(\theta)}[q_\ell(y|\theta, d)]$, we begin with the Donsker-Varadhan bound [11]

$$\mathrm{EIG}(d) \geq \mathbb{E}_{p(y,\theta|d)}[T(y,\theta)] - \log\left(\mathbb{E}_{p(\theta)p(y|d)}[e^{T(y,\theta)}]\right). \tag{57}$$

Substituting $T(y,\theta) = \log(q_\ell(y|\theta,d)/q_m(y|d))$ we have

$$\begin{aligned}
\mathrm{EIG}(d) &\geq \mathbb{E}_{p(y,\theta|d)}\left[\log\frac{q_\ell(y|\theta,d)}{q_m(y|d)}\right] - \log\left(\mathbb{E}_{p(\theta)p(y|d)}\left[\frac{q_\ell(y|\theta,d)}{q_m(y|d)}\right]\right) && (58)\\
&\geq \mathcal{I}_{\mathrm{m}+\ell}(d) - \log\left(\mathbb{E}_{p(y|d)}\left[\mathbb{E}_{p(\theta)}\left\{\frac{q_\ell(y|\theta,d)}{q_m(y|d)}\right\}\right]\right) && (59)\\
&\geq \mathcal{I}_{\mathrm{m}+\ell}(d) - \log\left(\mathbb{E}_{p(y|d)}\left[\frac{\mathbb{E}_{p(\theta)}\{q_\ell(y|\theta,d)\}}{q_m(y|d)}\right]\right) && (60)\\
&\geq \mathcal{I}_{\mathrm{m}+\ell}(d) - \log\left(\mathbb{E}_{p(y|d)}\left[\frac{q_m(y|d)}{q_m(y|d)}\right]\right) && (61)\\
&\geq \mathcal{I}_{\mathrm{m}+\ell}(d). && (62)
\end{aligned}$$

B Details for convergence rates

We now provide the details for Theorem 1. Key to proving the aspect of the Theorem relating to the convergence of the variational parameter ϕ_K to ϕ^* is Assumption 1. Points 1-5 correspond to assumptions H2', H3, H4, H6, and H7 of [28]; our proof will rely heavily on theirs. We note that also that our measurability assumption made in the Theorem itself means that their assumption H1 is automatically satisfied.

Assumption 1. Assume:

1. The function $\phi \mapsto f(X,\phi)$ is almost surely convex in its second argument and differentiable with Lipschitz continuous gradient, i.e. $\forall \phi_1, \phi_2 \in \Phi$:

   $$\mathbb{E}(\|\nabla f(X,\phi_1) - \nabla f(X,\phi_2)\|^2) \leq C\|\phi_1 - \phi_2\|$$

   with probability 1 for some C.

2. The function f is ν-strongly convex; that is, for all $\phi_1, \phi_2 \in \Phi$:

   $$\begin{aligned} f(X,\phi_1) \geq f(X,\phi_2) &+ \nabla f(X,\phi_2)^T(\phi_1 - \phi_2) \\ &+ \tfrac{\nu}{2}\|\phi_1 - \phi_2\|^2 \end{aligned}$$

3. There exists $\sigma > 0$ such that $\mathbb{E}[\|\nabla f(X,\phi^*)\|^2) \leq \sigma^2$

4. The function $\phi \mapsto f(X,\phi)$ is almost surely twice differentiable with Lipschitz continuous Hessian Hf, i.e. $\forall \phi_1, \phi_2 \in \Phi$:

   $$\mathbb{E}(\|(Hf)(X,\phi_1) - (Hf)(X,\phi_2)\|) \leq C'\|\phi_1 - \phi_2\|$$

5. There exists $\tau > 0$ such that $\mathbb{E}[\|\nabla f(X,\phi^*)\|^4] \leq \tau^4$ and there exists a positive definite operator Σ such that $\mathbb{E}[\nabla f(X,\phi^*) \otimes \nabla f(X,\phi^*)] \preccurlyeq \Sigma$

6. The function μ is Lipschitz continuous

It should be noted that, though relatively standard, these assumptions are also quite strong, particularly the assumption of strong convexity of f, and may well not hold in practice. In short, the stochastic gradient scheme used in optimizing the bounds may only converge toward a local optimum of

the bound $\phi^\dagger$, rather than the global optimum ϕ^*. When this happens the behavior and rates of convergence will generally be the same, but the error breakdown will become

$$\begin{aligned}
&\|\hat{\mu}(d,\phi_K) - \text{EIG}(d)\|_2 \\
&\qquad \leq \|\hat{\mu}(d,\phi_K) - \mathcal{B}(d,\phi_K)\|_2 && \text{(63a)}\\
&\qquad\quad + \left\|\mathcal{B}(d,\phi_K) - \mathcal{B}(d,\phi^\dagger)\right\|_2 && \text{(63b)}\\
&\qquad\quad + \left|\mathcal{B}(d,\phi^\dagger) - \text{EIG}(d)\right|. && \text{(63c)}
\end{aligned}$$

where

$$\left|\mathcal{B}(d,\phi^\dagger) - \text{EIG}(d)\right| \geq \left|\mathcal{B}(d,\phi^*) - \text{EIG}(d)\right|.$$

We now present our proof for the result, repeating the Theorem itself for convenience.

Theorem 1. *Let $\mathcal{X}$ be a measurable space and Φ be a convex subset of a finite dimensional inner product space. Let $X_1, X_2, ...$ be i.i.d. random variables taking values in $\mathcal{X}$ and $f : \mathcal{X} \times \Phi \to \mathbb{R}$ be a measurable function. Let*

$$\mu(\phi) \triangleq \mathbb{E}[f(X_1,\phi)] \approx \hat{\mu}_N(\phi) \triangleq \frac{1}{N}\sum\nolimits_{n=1}^{N} f(X_n,\phi)$$

and suppose that $\sup_{\phi\in\Phi} \|f(X_1,\phi)\|_2 < \infty$. Then $\sup_{\phi\in\Phi} \|\hat{\mu}_N(\phi) - \mu(\phi)\|_2 = \mathcal{O}(N^{-1/2})$. Suppose further that Assumption 1 in Appendix B holds and that ϕ^* is the unique minimizer of μ. After K iterations of the Polyak-Ruppert averaged stochastic gradient descent algorithm of [28] with gradient estimator $\nabla_\phi f(X_t,\phi)$, we have $\|\mu(\phi_K) - \mu(\phi^*)\|_2 = \mathcal{O}(K^{-1/2})$ and, combining with the first result,

$$\|\hat{\mu}_N(\phi_K) - \mu(\phi^*)\|_2 = \mathcal{O}(N^{-1/2} + K^{-1/2}) = \mathcal{O}(T^{-1/2}) \text{ if } N \propto K.$$

Proof of Theorem 1

Proof. We begin by establishing the uniform convergence of $\hat{\mu}_N(\phi)$ to $\mu(\phi)$, for which we simply use the L^2 weak law of large numbers. Specifically, we let $Y_n = f(X_n,\phi)$ and $\varepsilon_N(\phi) = \|\hat{\mu}_N(\phi) - \mu(\phi)\|_2$, then

$$\begin{aligned}
\varepsilon_N^2(\phi) &= \mathbb{E}\left(\left[\frac{1}{N}\sum_{n=1}^{N}(Y_n - \mathbb{E}Y_n)\right]^2\right) && \text{(64)}\\
&= \mathbb{E}\left(\frac{1}{N^2}\sum_{n=1}^{N}(Y_n - \mathbb{E}Y_n)^2\right) && \text{(65)}\\
&= \frac{1}{N^2}\cdot N\text{Var}(Y_n) && \text{(66)}\\
&\leq \frac{1}{N}\sup_{\phi\in\Phi}\|f(X_1,\phi)\|_2^2 && \text{(67)}
\end{aligned}$$

which is bounded by assumption. Thus

$$\sup_{\phi\in\Phi}\varepsilon_N(\phi) = \mathcal{O}(N^{-1/2}) \tag{68}$$

as required.

We turn now to the stochastic gradient descent convergence. We begin by applying Theorem 3 of [28] using points 1-5 of Assumption 1 to give

$$\|\phi_K - \phi^*\|_2 = \mathcal{O}(K^{-1/2}) \tag{69}$$

and (see [28] page 4)

$$\mathbb{E}\mu(\phi_K) - \mu(\phi^*) = \mathcal{O}(K^{-1/2}). \tag{70}$$

To establish L^2 convergence of the function values, it remains to control the variance of $\mu(\phi_K)$. We now invoke point 6 of Assumption 1 to see that, for some constant B (namely the Lipschitz constant for μ),

$$\text{Var}[\mu(\phi_K)] = \mathbb{E}\left[\left(\mu(\phi_K) - \mathbb{E}\left[\mu(\phi_K)\right]\right)^2\right] \tag{71}$$

$$\leq \mathbb{E}\left[(\mu(\phi_K) - \mu(\mathbb{E}\phi_K))^2\right] \tag{72}$$
$$\leq B^2\mathbb{E}\left[(\phi_t - \mathbb{E}\phi_t)^2\right] \tag{73}$$
$$\leq B^2\|\phi_K - \phi^*\|_2^2. \tag{74}$$

By (69) we conclude $\sqrt{\text{Var}[\mu(\phi_K)]} = \mathcal{O}(K^{-1/2})$. Thus $\mu(\phi_K)$ converges in L^2 at the required rate. Finally, if $\epsilon_K = \|\hat{\mu}_K(\phi_K) - \mu(\phi^*)\|_2$ then

$$\begin{aligned}
\epsilon_K &\leq \|\hat{\mu}_K(\phi_K) - \mu(\phi_K)\|_2 + \|\mu_K(\phi_K) - \mu(\phi^*)\|_2 \\
&\leq \|\hat{\mu}_K(\phi_K) - \mu(\phi_K)\|_2 + \sup_{\phi\in\Phi}\|\hat{\mu}_K(\phi) - \mu(\phi)\|_2 \\
&= \mathcal{O}(N^{-1/2} + K^{-1/2}) \\
&= \mathcal{O}(T^{-1/2})
\end{aligned}$$

as required. □

Finally, we discuss the necessary extensions for $\mathcal{I}_{\text{m}+\ell}$. The assumptions of the Theorem are subtly different in this case. Specifically, we require Assumption 1 to hold for the integrand of $\mathcal{F}$ rather than the integrand of $\mathcal{I}_{\text{m}+\ell}$, where $\mathcal{F}(d,\phi) = -\mathbb{E}[\log q_m(y|d) + \log q_\ell(y|\theta,d)] + C$ is the loss function that we use to train ϕ, and require $\mathcal{I}_{\text{m}+\ell}$ to be Lipschitz continuous in ϕ.

The Monte Carlo error is no different in this setting. However, ϕ^* is optimal with respect to $\mathcal{F}(d,\phi)$ rather than $\mathcal{I}_{\text{m}+\ell}$ and the asymptotic bias term is $|\mathcal{I}_{\text{m}+\ell}(d,\phi^*) - \text{EIG}(d)| \leq \mathcal{F}(d,\phi^*)$ by Lemma 2. For the optimization term, we have from equation (69) that $\|\phi_K - \phi^*\|_2 = \mathcal{O}(K^{-1/2})$. Then by the Lipschitz assumption on $\mathcal{I}_{\text{m}+\ell}$, we have $\|\mathcal{I}_{\text{m}+\ell}(d,\phi_k) - \mathcal{I}_{\text{m}+\ell}(d,\phi^*)\|_2 = \mathcal{O}(K^{-1/2})$. The rest of the proof now goes through as above.

C Related work

In this section, we provide a more detailed discussion of existing techniques for EIG estimation to complement Sec. 5 in the main text.

One established approach is to use a **Laplace approximation** to the posterior to make fast approximations of EIG [22, 25]

$$\hat{\mu}_{\text{laplace}}(d) \triangleq \frac{1}{N}\sum_{n=1}^{N}\left[H[p(\theta)] - H[q(\theta|y_n,d)]\right] \tag{75}$$

where $q(\theta|y_n,d)$ is a Laplace approximation to $p(\theta|y_n,d)$ that is computed once for each $y_n \sim p(y|d)$.

Kleinegesse and Gutmann [18] recently suggested an implicit likelihood approach that directly approximates the ratio $r(d,\theta,y) = p(y|\theta,d)/p(y|d)$ using samples from $p(y|\theta,d)$ and $p(y|d)$ and the **Likelihood-Free Inference by Ratio Estimation (LFIRE)** method suggested by [41], which is itself based around logistic regression. This yields the estimator

$$\hat{\mu}_{\text{LFIRE}}(d) \triangleq \frac{1}{N}\sum_{n=1}^{N}\log\hat{r}(d,\theta_n,y_n) \tag{76}$$

where $\log\hat{r}(d,\theta_n,y_n)$ is estimated separately for each pairs of samples y_n,θ_n.

In principal one could also exploit the equivalence between EIG and MI and use other existing MI estimation methods, a number of which were recently summarized by [31]. Of particular note, Belghazi et al. [4] use a bound on MI in the context of generative adversarial neural network training that is based on the **Donsker-Varadhan (DV)** representation of the KL divergence [11]. Specifically, they introduce a parametrized approximation $T(y,\theta|d,\phi)$ to $\log\frac{p(y,\theta|d)}{p(\theta)p(y|d)}$ and then optimize the lower bound

$$\mathcal{L}_{\text{DV}}(d) \triangleq \mathbb{E}_{p(y,\theta|d)}[T(y,\theta|d,\phi)] - \log\left(\mathbb{E}_{p(\theta)p(y|d)}[e^{T(y,\theta|d,\phi)}]\right). \tag{77}$$

The estimator $\hat{\mu}_{\text{DV}}$ is then produced in an analogous manner to $\hat{\mu}_{\text{post}}$.

The EIG has been applied by a number of authors in specific contexts. For instance, the EIG has been used to formulate acquisition functions in Bayesian optimization [16]. More recently, Ma et al. [26] used an EIG-type objective to select features rather than designs for a partial VAE model. The EIG estimation exploits the model structure of the partial VAE. Additionally, and in contrast to this paper, approximations learned using the ELBO are used rather than approximations that are trained using variational objectives that are directly tied to EIG estimation. For further discussion on the implications of using the ELBO (i.e. the reverse KL divergence) in EIG estimation settings, see Appendix G.

As mentioned previously, mutual information bounds are of interest in traditional signal processing [3] and of increasing interest in the deep learning community [31]—although to the best of our knowledge they have not been applied to BOED before. Interestingly, it is lower bounds that are of primary importance in the deep learning setting because of the interplay between MI estimation and the subsequent gradient-based optimization over parameters. This is in contrast to this work, in which we maximize EIG over designs using Bayesian optimization—allowing the use of estimators such as $\hat{\mu}_{\mathrm{m}+\ell}$ that are not, in expectation, bounds.

D Experiment details

Computing All experiments were run on a machine with 32818560 kB memory, 8 Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz processors, running Fedora 28, Python 3.6.8, Pytorch 1.1.0. To reproduce the results presented in the paper, see `https://github.com/ae-foster/pyro/tree/vboed-reproduce`. The methods in this paper form part of Pyro's OED support, the documentation for which is provided at `http://docs.pyro.ai/en/stable/contrib.oed.html`.

D.1 EIG estimation accuracy

A/B test We consider a classical A/B test, commonly used in marketing and design applications. Here the experiment design is the choice of group sizes: n participants are split between groups A and B of size n_A and $n - n_A$, respectively. For each participant we measure a continuous response y. We consider a linear data analysis model

$$\theta \sim N(0, \Sigma_\theta) \qquad y|\theta, d \sim N(X_d\theta, I) \tag{78}$$

where X_d is the $n \times 2$ design matrix with $(1\ 0)$ for the first n_A rows and $(0\ 1)$ for the remainder.

In this example we set the number of participants to be $n = 10$ with 11 designs ($n_A = 0, ..., 10$) and the prior covariance matrix to be

$$\Sigma_\theta = \begin{pmatrix} 10^2 & 0 \\ 0 & 1.82^2 \end{pmatrix} \tag{79}$$

We chose families of variational distributions that include the true posterior (or true marginal). For the amortised posterior, we set $\phi = (A, \Sigma_{\mathrm{p}})$ with ϕ trained separately for each d and let

$$q_p(\theta|y, d, \phi) \sim N(Ay, \Sigma_{\mathrm{p}}) \tag{80}$$

where A is a 10×2 matrix and Σ_{p} is positive definite. For the marginal, we simply take $\phi = (\mu_{\mathrm{m}}, \Sigma_{\mathrm{m}})$ and

$$q_m(y|d, \phi) \sim N(\mu_{\mathrm{m}}, \Sigma_{\mathrm{m}}). \tag{81}$$

For NMC and Laplace, no variational families need to be specified.

For LFIRE, we used a parametrization $\phi = (b, \delta, \Lambda)$ and used the ratio estimate

$$\log \hat{r}(y|\theta, d, \phi) = b - (y - \delta)^T \Lambda (y - \delta) \tag{82}$$

where Λ is positive definite. This form was chosen to mimic the approximation made by the posterior method, and so reduce the effect of architecture on performance.

For DV, we used a similar critic, namely we set $\phi = (A, \Lambda)$ and

$$T(y, \theta|d, \phi) = -(\theta - Ay)^T \Lambda (\theta - Ay) \tag{83}$$

where Λ is positive definite.

The ground truth EIG(d) was computed analytically. In Table 2, each estimator was allowed 10 seconds computation.

Preference We consider searching for an agent's utility indifference point, using responses that are both *censored* and *corrupted* with non-uniform noise. Let $d \in \mathbb{R}$ and

$$\begin{aligned} \theta &\sim N(\mu_\theta, \sigma_\theta^2) \\ \eta|\theta, d &\sim N(d - \theta, \sigma_\eta^2(1 + |d|)^2) \\ y &= f(\eta) \end{aligned} \tag{84}$$

where

$$f : \mathbb{R} \to [\epsilon, 1 - \epsilon] \tag{85}$$

$$x \mapsto \begin{cases} \epsilon & \text{if } x \leq \text{logit}(\epsilon) \\ 1 - \epsilon & \text{if } x \geq \text{logit}(1 - \epsilon) \\ \frac{1}{1-e^{-x}} & \text{otherwise} \end{cases} \tag{86}$$

and $\text{logit}(p) = \log p - \log(1 - p)$.

For this example we set $\mu_\theta = -20, \sigma_\theta = 20$ and $\sigma_\eta = 1$. We took designs on a linearly spaced grid in $[-80, 80]$. For the variational family for the posterior, we took $\phi = (w, \sigma, \mu_0, \sigma_0, \mu_1, \sigma_1)$ and then

$$q_p(\theta|y, d, \phi) \sim N(\mu_\text{p}, \sigma_\text{p}^2) \qquad \text{where} \qquad \hat{\eta} = d - \text{logit}(y) \tag{87}$$

$$\mu_\text{p} = w\hat{\eta} + (1 - w)\mu_\theta + \mu_0 \, \mathbf{1}_{\{y=\epsilon\}} + \mu_1 \, \mathbf{1}_{\{y=1-\epsilon\}} \tag{88}$$

$$\sigma_\text{p}^2 = \sigma^2 + \sigma_0^2 \, \mathbf{1}_{\{y=\epsilon\}} + \sigma_1^2 \, \mathbf{1}_{\{y=1-\epsilon\}} \tag{89}$$

For the marginal, we simply took $\phi = (\mu_\text{m}, \sigma_\text{m})$ and

$$q_m(y|d, \phi) \sim f \# N(\mu_\text{m}, \sigma_\text{m}^2). \tag{90}$$

where $\#$ denotes the push-forward measure. We note that this variational family contains the true marginal.

For LFIRE, we used the parametrization $\phi = (b, b_0, b_1, \delta, \lambda)$ with ratio estimate

$$\hat{\eta} = d - \text{logit}(y) \tag{91}$$

$$\log \hat{r}(y|\theta, d, \phi) = b - \lambda(\hat{\eta} - \delta)^2 + b_0 \, \mathbf{1}_{\{y=\epsilon\}} + b_1 \, \mathbf{1}_{\{y=1-\epsilon\}} \tag{92}$$

For DV, the critic had parametrization $\phi = (b_0, b_1, \delta_i, \delta_0, \delta_1, \lambda_i, \lambda_0, \lambda_1)$ and we set

$$\hat{\eta} = d - \text{logit}(y) \tag{93}$$

$$\lambda = \lambda_i + \lambda_0 \, \mathbf{1}_{\{y=\epsilon\}} + \lambda_1 \, \mathbf{1}_{\{y=1-\epsilon\}} \tag{94}$$

$$\delta = \delta_i + \delta_0 \, \mathbf{1}_{\{y=\epsilon\}} + \delta_1 \, \mathbf{1}_{\{y=1-\epsilon\}} \tag{95}$$

$$T(y, \theta|d, \phi) = -\lambda(\hat{\eta} - \delta)^2 + b_0 \, \mathbf{1}_{\{y=\epsilon\}} + b_1 \, \mathbf{1}_{\{y=1-\epsilon\}} \tag{96}$$

Both these forms were chosen to minimize the differences between the functional forms used for different methods.

The ground truth $\text{EIG}(d)$ was computed by running the marginal method, which is statistically consistent for this example because the true marginal is contained in the variational family, to convergence. The posterior and Laplace methods are both asymptotically biased (see Figure 5) and in this case both make the same (Gaussian) distributional assumption. The posterior method, however, produces better EIG estimates. For the benchmarking results in Table 2, 10 seconds computation was allowed.

Mixed Effects Regression We consider BOED for a mixed effects regression model with a non-linear linking function that will also serve as the basis for the adaptive experiment we run in Sec. 6.3. This class of models is commonly used for analyzing data in a variety of scientific disciplines, where including nuisance variables can be a critical component of the model. In our adaptive experiment, the nuisance variables—i.e. the random effects—are used to account for the variability of individual human participants. Because of the presence of nuisance variables these implicit likelihood models represent a significant challenge for BOED.

We begin by describing the experiment set-up. Participants were presented with a question of the form seen in Figure 6 with the possible images shown in Figure 7. There were two image feature

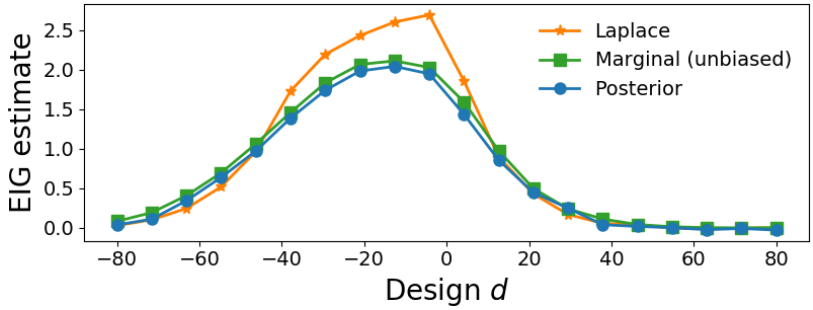


Figure 5: EIG curves for the Preference example, with estimators run until variance is negligible and iterates of ϕ are stable to highlight the asymptotic bias.

dimensions with 3 levels each. A single image i could therefore be represented as a 1×6 matrix X_i with two entries 1 and the rest 0. With the left image i_1 and right image i_2, the question was represented as $X_d = X_{i_1} - X_{i_2}$ encoding the assumed left-right symmetry. We then considered a model for the ith participant

$$\theta \sim N(0, \Sigma_\theta) \tag{97}$$
$$\sigma_\psi^{-2} \sim \Gamma(\alpha_\psi, \beta_\psi) \tag{98}$$
$$\psi_i | \sigma_\psi \sim N(0, \sigma_\psi^2 I_6) \tag{99}$$
$$\sigma_k^{-2} \sim \Gamma(\alpha_k, \beta_k) \tag{100}$$
$$\log k_i | \sigma_k \sim N(0, \sigma_k^2) \tag{101}$$
$$\eta | \theta, \psi_i, k_i, d \sim N(k_i(X_d\theta + X_d\psi_i), \sigma_\eta^2) \tag{102}$$
$$y = f(\eta) \tag{103}$$

where f is the censored sigmoid defined in (86) and $i \in \{1, ..., 8\}$ as there were 8 different participants.

The actual prior values of the parameters used were

$$\Sigma_\theta = 100 I_6 \qquad \sigma_\eta = 10 \tag{104}$$
$$\alpha_\psi = \beta_\psi = \alpha_k = \beta_k = 2 \tag{105}$$

We begin by discussing the variational families used to estimate the EIG.

For the posterior estimator of EIG, we took $\phi = (A, \Sigma_\text{p})$ and

$$\hat{\eta} = \text{logit}(y) \tag{106}$$
$$q_p(\theta | y, d, \phi) \sim N(A\hat{\eta}, \Sigma_\text{p}) \tag{107}$$

For the marginal + likelihood estimator, we set $\phi = (\mu_\text{m}, \sigma_\text{m}, \mu_\ell, \sigma_\ell, \xi)$ and took

$$q_m(y | d, \phi) \sim f\#N(\mu_\text{m}, \sigma_\text{m}^2) \tag{108}$$
$$q_\ell(y | \theta, d, \phi) \sim f\#N(e^\xi X_d\theta + \mu_\ell, \sigma_\ell^2) \tag{109}$$

For LFIRE, we used $\phi = (b, \delta, \lambda)$ and then took

$$\hat{\eta} = \text{logit}(y) \tag{110}$$
$$\log \hat{r}(y | \theta, d, \phi) = b - \lambda(\hat{\eta} - \delta)^2 \tag{111}$$

For DV, we used $\phi = (\lambda, \xi)$ and

$$\hat{\eta} = \text{logit}(y) \tag{112}$$
$$T(y, \theta | d, \phi) = -\lambda(\hat{\eta} - e^\xi X_d\theta)^2 \tag{113}$$

For benchmarking, we computed the ground truth using a variant of NMC. Specifically, we note that

$$p(y|d) = \mathbb{E}_{p(\theta,\psi,k)}[p(y|\theta,\psi,k,d)] \tag{114}$$

$$p(y|\theta,d) = \mathbb{E}_{p(\psi,k)}[p(y|\theta,\psi,k,d)] \tag{115}$$

and for this model, we can sample directly from $p(\psi, k)$. These identities allow us to estimate the marginal and likelihood by Monte Carlo, and then combine in a NMC estimator for EIG(d). Whilst inefficient, this estimator is statistically consistent.

We allowed 60 seconds computation per estimator to compute the results of Table 2. Encouragingly, we find that our variational estimators outperform the LFIRE and DV baselines on this model and exhibit low errors even though they both make suboptimal distributional assumptions about the posterior/marginal.

Extrapolation We consider designing experiments to reduce posterior uncertainty in the model prediction at another point in design space—a point that we cannot experiment on directly. For this example, we take $\psi \sim N(\mu_\psi, \Sigma_\psi)$ and

$$\theta|\psi \sim \text{Bernoulli}(\text{logit}^{-1}(X_\theta \psi))$$

$$y|\psi, d \sim \text{Bernoulli}(\text{logit}^{-1}(X_d \psi))$$

where $X_\theta = \begin{pmatrix} 1 & -\frac{1}{2} \end{pmatrix}$ and $X_d = (-1 \quad d)$ for $d \in \mathbb{R}$. Interestingly, this model admits efficient sampling of $y, \theta \sim p(y, \theta|d)$ but *not* $y \sim p(y|\theta, d)$. Therefore, whilst the posterior, marginal + likelihood and DV methods are all applicable, LFIRE is not.

For the posterior method we set $\phi = (l_0, l_1)$ and

$$l_{\text{p}}(y) = l_1 y + l_0(1-y) \tag{116}$$

$$q_p(\theta|y,d,\phi) \sim \text{Bernoulli}(\text{logit}^{-1}(l_{\text{p}}(y))). \tag{117}$$

We computed the prior entropy, which is not analytically tractable here, using a MC estimator, noting that θ has a finite sample space.

For the marginal + likelihood method, we let $\phi = (l, l_0, l_1)$ and then

$$q_m(y|d,\phi) \sim \text{Bernoulli}(\text{logit}^{-1}(l)) \tag{118}$$

$$l_\ell(\theta) = l_1\theta + l_0(1-\theta) \tag{119}$$

$$q_\ell(y|\theta,d,\phi) \sim \text{Bernoulli}(\text{logit}^{-1}(l_\ell(\theta))). \tag{120}$$

Finally, for DV, we let $\phi = (w_y, w_\theta, w_{y\theta})$ and took

$$T(\theta, y|d,\phi) = w_y y + w_\theta \theta + w_{y\theta} y\theta. \tag{121}$$

The ground truth EIG was computed using MC, noting that the sample spaces for y, θ are finite in this example. 10 seconds computation per methods was allowed for the results in Table 2.

D.2 End-to-end sequential experiments

Mechanical Turk experiment We begin by describing the experiment itself. Participants were presented with a question of the form seen in Figure 6 with the possible images shown in Figure 7. There were two image feature dimensions with 3 levels each. A single image i could therefore be represented as a 1×6 matrix X_i with two entries 1 and the rest 0. With the left image i_1 and right image i_2, the question was represented as $X_d = X_{i_1} - X_{i_2}$ encoding the assumed left-right symmetry.

The model and EIG estimation were the same as the mixed effects model in Sec. D.1. When optimizing the EIG to select designs d_t, we estimated EIG across all candidate designs. We allowed a 30s turnaround to learn the posterior from the previous data, estimate the EIG, select the next design, and present it to the user. We estimated the EIG in parallel for all 36 designs to select the best design at each step. For each independent run of the experiment there were 8 participants, each answering 10 questions. This allowed the interplay between fixed effects and random effects to be apparent.

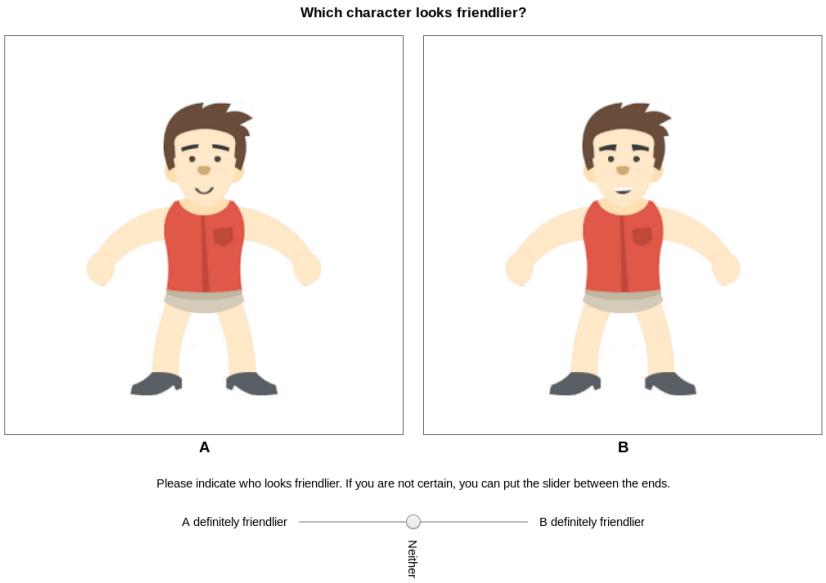


Figure 6: A screenshot of the question answering interface used by human participants in the adaptive experiment in Sec. 6.3.

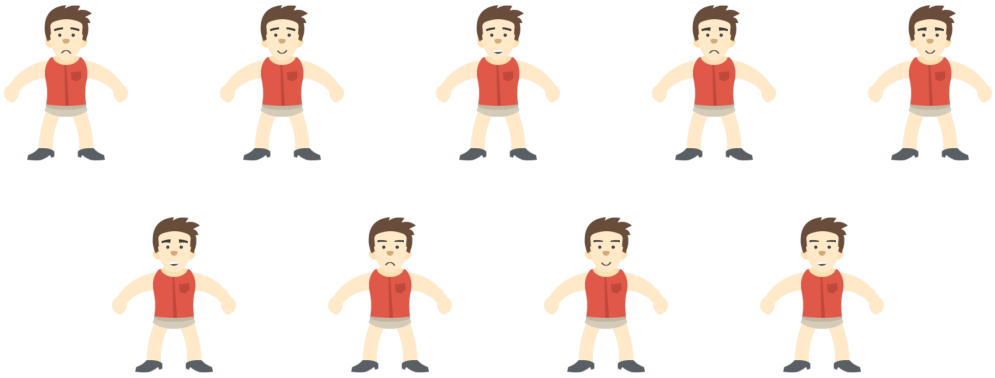


Figure 7: The nine characters we used in the adaptive experiment in Sec. 6.3. They vary along two feature dimensions: the mouth (smile, frown, showing teeth) and eyebrows.

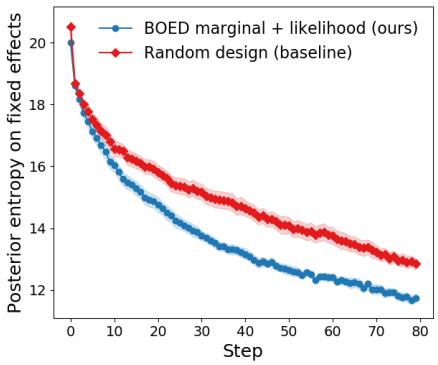


Figure 8: Evolution of the posterior entropy of the fixed effects in the Mechanical Turk experiment in Sec. 6.3 with simulated data. We depict the mean and ± 1 std. err. from 10 experimental trials.

Because we used this model to run an adaptive experiment, we required a variational family to learn the full posterior (over random effects and hyperparameters as well as θ).

For the full variational inference of the posterior used when we receive actual data, we used a partial mean-field approximation. Specifically, we set $q(\theta, \sigma_\psi, (\psi_i)_{i=1}^8, \sigma_k, (k_i)_{i=1}^8)$ to be

$$\theta \sim N(\mu_\theta, \Sigma_\theta) \tag{122}$$
$$\sigma_\psi^{-2} \sim \Gamma(\alpha_\psi, \beta_\psi) \tag{123}$$
$$\psi_i|\theta \sim N(A(\theta - \mu_\theta) + \mu_{\psi_i}, \Sigma_{\psi_i}) \tag{124}$$
$$\sigma_k^{-2} \sim \Gamma(\alpha_k, \beta_k) \tag{125}$$
$$\log k_i \sim N(\mu_{k_i}, \sigma_{k_i}^2) \tag{126}$$

and we learned the variational parameters $\mu_\theta, \Sigma_\theta, \alpha_\psi, \beta_\psi, A, \mu_{\psi_i}, \Sigma_{\psi_i}, \alpha_k, \beta_k, \mu_{k_i}, \sigma_{k_i}$ by conventional (not amortized) variational inference. Note that, under this approximate posterior, θ is multivariate Gaussian so we can compute its entropy analytically.

Finally we ran an additional experiment identical to the first, but using simulated data rather than human responses. We took

$$\theta = (-30 \quad 30 \quad 0 \quad -12 \quad -6 \quad 18)\,. \tag{127}$$

We simulated the random effects ψ, k from the prior and used the prior value $\sigma_\eta = 10$. The entropy results are presented in Figure 8. As expected, BOED decreases posterior uncertainty more quickly.

D.3 Constant Elasticity of Substitution (CES) experiment

We begin by describing the experiment set-up. The economic agent is presented with a sequence of designs d. Each designs comprises two baskets $\mathbf{x}$ and $\mathbf{x}'$ of goods. The agent then indicates which basket they prefer on a one-dimensional slider—they may indicate a strong preference, weak preference, or indifference.

To model the agent's responses, we use the CES utility model [2] which defines a utility

$$U(\mathbf{x}) = \left(\sum_i x_i^\rho \alpha_i \right)^{1/\rho} \tag{128}$$

for a basket of goods $\mathbf{x}$. In this experiment, we took baskets $\mathbf{x} \in [0, 100]^3$ representing non-negative quantities of three commodities.

Extending the preference example in the previous section, we assume the agent, when asked to compare baskets $\mathbf{x}$ and $\mathbf{x}'$ and indicate their preference on a slider, base their response on $U(\mathbf{x}) - U(\mathbf{x}')$. Specifically, we use the following likelihood model

$$\rho \sim \text{Beta}(a_\rho, b_\rho) \tag{129}$$
$$\boldsymbol{\alpha} \sim \text{Dirichlet}(\boldsymbol{c_\alpha}) \tag{130}$$
$$\log u \sim N(\mu_u, \sigma_u^2) \tag{131}$$
$$\eta|\rho, \boldsymbol{\alpha}, d \sim N(u \cdot (U(\mathbf{x}) - U(\mathbf{x}')), \sigma_\eta^2 u^2 (1 + \|\mathbf{x} - \mathbf{x}'\|)^2) \tag{132}$$
$$y = f(\eta) \tag{133}$$

This represents a challenging experiment design problem for a number of reasons. First, for large values of $U(\mathbf{x}) - U(\mathbf{x}')$ the agent's response will be predictable gaining little information. For very different baskets ($\|\mathbf{x} - \mathbf{x}'\|$ large) the responses will be noisy indicating our intuition that it is more difficult to compare very different baskets. However, very similar baskets will have similar utilities and the agent will be predictably indifferent. Optimal designs therefore lie in a sweet spot where: i) baskets are similar to avoid high noise regions, but dissimilar enough to be informative; and ii) the difference in utility is close to 0 under the current posterior. BOED is able to trade off these considerations in a principled manner.

For this specific example we took

$$a_\rho = b_\rho = 1 \qquad\qquad \boldsymbol{c_\alpha} = (1, 1, 1) \tag{134}$$

$$\mu_u = 1 \qquad \sigma_u = 3 \tag{135}$$

$$\sigma_\eta = 0.005 \tag{136}$$

To estimate the EIG, we used a marginal guide based on the one used in the preference example. Specifically, we set $\phi = (\mu_\text{m}, \sigma_\text{m}, p_0, p_1)$ and

$$r(y|d,\phi) \sim f\#N(\mu_\text{m}, \sigma_\text{m}^2), \tag{137}$$

$$q_p(y|d,\phi) = \begin{cases} \epsilon & \text{with probability } p_0 \\ 1-\epsilon & \text{with probability } p_1 \\ r(y|d,\phi) & \text{with probability } 1-p_0-p_1 \end{cases} \tag{138}$$

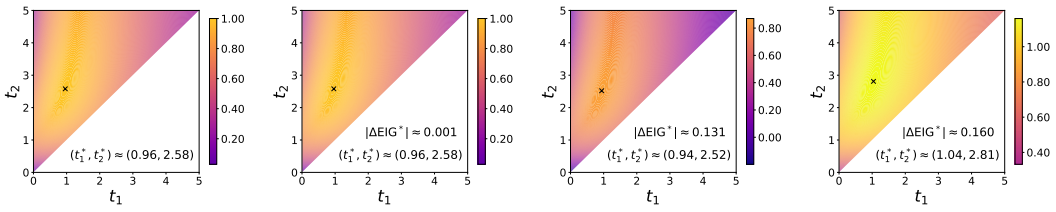
where $\#$ denotes the push-forward measure. This is simply a mixture of a discrete distribution on end-points with a sigmoid transformed Gaussian.

To select designs, we used Bayesian optimization with a Matern52 kernel with lengthscale 20 and variance set empirically. Both $\hat{\mu}_\text{marg}$ and $\hat{\mu}_\text{NMC}$ were allowed the same time budget to select designs and used an identical Bayesian optimization procedure. Random designs were chosen uniformly on $[0, 100]^6$.

To learn the posterior at subsequent steps we used a mean-field variational approximation with the same families as the prior. That is, we updated the parameters $a_\rho, b_\rho, \boldsymbol{c}_{\boldsymbol{\alpha}}, \mu_u, \sigma_u$ and left the structure otherwise intact. The RMSEs of Figure 4 were expectations over the posterior: $\left(\mathbb{E}_{p(\theta|d_{1:t}, y_{1:t})}[\|\theta - \theta^*\|^2]\right)^{1/2}$.

E Additional experiments

E.1 Death process

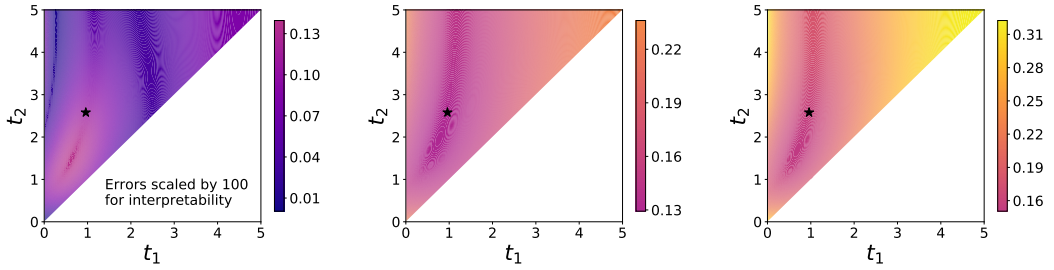


(a) Exact EIG (b) Posterior LogNormal (c) Truncated Normal (d) Laplace

Figure 9: EIG surfaces estimated by four methods for the two-dimensional design (t_1, t_2) for the continuous time model described in Sec. E.1. The optimal design (t_1^*, t_2^*) determined by each method is indicated with a cross. The posterior method with a LogNormal variational distribution yields nearly exact results. The posterior method with a Truncated Normal distribution and the Laplace method are not as accurate but still result in designs with large EIG. Note that the EIG has been scaled for interpretability and that all four figures use a common scale. The errors of these estimators are examined more closely in Figure 10.

We examine experimental design for the simple continuous time process considered in [9] and [18], arising in epidemiology. Consider a population with fixed size N that is initially healthy at time $t = 0$, with individuals becoming infected at a constant rate b as time evolves. We consider a design space $d = (t_1, t_2)$, where $0 \leq t_1 \leq t_2$, corresponding to the times at which we measure the number of infected individuals. We place a log-normal prior on the infection rate b.

For this example, we investigate how the choice of variational family affects the asymptotic bias. In Fig. 9 we compare the EIG surfaces obtained using four estimators: i) an exact method that uses brute force quadrature; ii) $\hat{\mu}_\text{post}$ with a log-normal variational distribution; iii) $\hat{\mu}_\text{post}$ with a truncated normal variational distribution; and iv) the Laplace approximation $\hat{\mu}_\text{laplace}$. The log-normal family matches the true posterior best, giving mean absolute errors $\sim 10^{-3}$. The second posterior method and the Laplace approximation both make the same distributional assumption, but Laplace results in absolute errors that are about 30% higher than for the posterior method. See Fig. 10 for a closer analysis of the errors of the approximate methods.



(a) Posterior LogNormal (b) Posterior Truncated Normal (c) Laplace

Figure 10: Absolute EIG errors corresponding to the estimates depicted in Fig. 9. The optimal design (t_1^*, t_2^*) determined by an exact method is indicated with a star. The absolute error of the LogNormal Posterior estimate is $\sim 10^{-3}$ across the design space. The mean absolute error of the Laplace EIG estimates across the design space is about 30% higher than for the Posterior method with a Truncated Normal variational distribution. In this case the Laplace method results in an upper bound, while (as always) both Posterior methods yield a lower bound. All three figures have the same scale as Fig. 9, except for the LogNormal errors, which have been scaled by an additional factor of 100.

Experimental details The likelihood for observing (I_1, I_2) infected individuals from a population of size N at times (t_1, t_2) is given by [12]:

$$p(I_1, I_2|b, t_1, t_2) = \frac{N!}{I_1!(I_2 - I_1)!(N - I_2)!} \left[1 - e^{-bt_1}\right]^{I_1} \times \left[1 - e^{-b(t_2 - t_1)}\right]^{I_2 - I_1} \left[e^{-bt_1}\right]^{I_2 - I_1} \left[e^{-bt_2}\right]^{N - I_2} \tag{139}$$

The prior over the infection rate $b > 0$ is taken to be

$$\log b \sim N(\mu_b, \sigma_b) \tag{140}$$

so that the joint density is given by

$$p(I_1, I_2, b|t_1, t_2) = p(I_1, I_2|b, t_1, t_2)p(b) \tag{141}$$

In our experiment we choose $N = 10$, $\mu_b = 0$, and $\sigma_b = 0.25$. The figures are scaled such that the maximum EIG over the design space (as computed with the exact method) is 1.0. For all four EIG estimation methods we use quadrature and exact summation over the outcomes (I_1, I_2) where appropriate to obtain maximally accurate results. That is, the obtained results are only constrained by the methods themselves and not the computational budget used. Note that we do not make use of any kind of amortization.

F Consistent EIG estimation with control variates

In this section, we show that an approximation to the marginal density $q_m(y|d)$ can be used a control variate. Control variates are a means to reduce the variance of Monte Carlo estimators by using expectations which can be computed analytically. Here, we assume that, for every θ, the KL divergence $\mathrm{KL}\left(\, p(y|\theta, d) \,||\, q_m(y|d) \,\right)$ can be computed analytically. For example, this would be the case if both $p(y|\theta, d)$ and $q_m(y|d)$ were Gaussian.

We begin by writing the EIG as

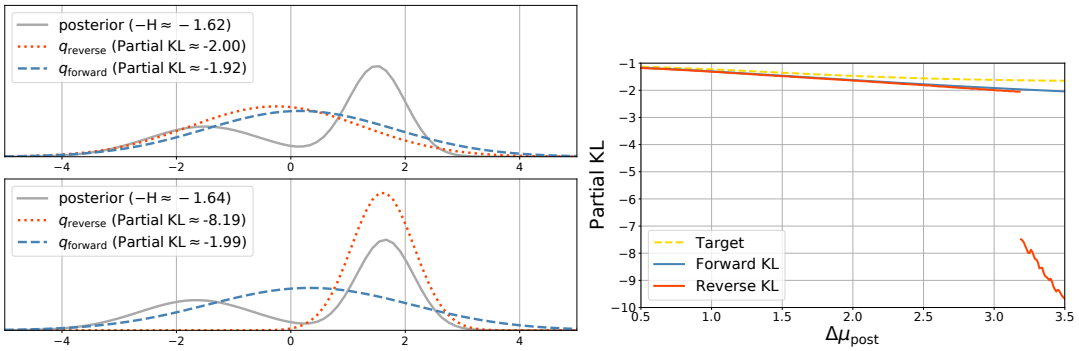
$$\mathrm{EIG}(d) = \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta, d)}{p(y|d)}\right] \tag{142}$$

$$= \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{p(y|\theta, d)}{q_m(y|d)}\right] + \mathbb{E}_{p(y,\theta|d)}\left[\log \frac{q_m(y|d)}{p(y|d)}\right] \tag{143}$$

$$= \mathbb{E}_{p(\theta)}\left[\mathrm{KL}\left(\, p(y|\theta, d) \,||\, q_m(y|d) \,\right)\right] - \mathrm{KL}\left(\, p(y|d) \,||\, q_m(y|d) \,\right). \tag{144}$$

We can now use our assumption on the first term,

$$\mathbb{E}_{p(\theta)}\left[\mathrm{KL}\left(\, p(y|\theta, d) \,||\, q_m(y|d) \,\right)\right] \to \mathbb{E}_{p(\theta)}\left[\text{analytic function of } \theta\right] \tag{145}$$



(a) Forward and reverse KL minimization

(b) Discontinuity in the partial KL

Figure 11: (a) Normal variational distributions found by fitting to a target posterior that is a mixture with two distinct Normal components. In both plots, the target posterior is a mixture of $N(\mu_1, 0.5^2)$ and $N(\mu_2, 1.0^2)$ and we vary $\Delta\mu_{\text{post}} = \mu_1 - \mu_2$. In the top plot, the gap between the two components is $\Delta\mu_{\text{post}} = 3.0$, while in the bottom plot $\Delta\mu_{\text{post}} = 3.3$. In contrast to the behaviour resulting from forward KL minimization, the mode-seeking behaviour of reverse KL minimization leads to a large change in the corresponding optimal variational distribution from top to bottom. (b) We plot the partial KL as we vary $\Delta\mu_{\text{post}}$ for the target posterior described in (a). The partial KL as estimated by reverse KL minimization exhibits a sharp discontinuity as the gap between the two components crosses $\Delta\mu_{\text{post}} \approx 3.18$.

and this expectation can be computed efficiently with conventional Monte Carlo. For the second term, we use Nested Monte Carlo

$$\text{KL}\left(\, p(y|d) \,||\, q_m(y|d) \,\right) \approx \frac{1}{N}\sum_{n=1}^{N} \log \frac{\frac{1}{M}\sum_{m=1}^{M} p(y_n|\theta_m, d)}{q_m(y_n|d)} \tag{146}$$

where $y_n \overset{\text{i.i.d.}}{\sim} p(y|d)$ and $\theta_m \overset{\text{i.i.d.}}{\sim} p(\theta)$. The key benefit of this approach is that this estimator may have lower variance than a direct NMC estimator of EIG(d). Indeed, if we let $A = \log\left(\frac{1}{M}\sum_{m=1}^{M}(y|\theta_m, d)\right)$ and $B = \log q_m(y|d)$ then the variance of the estimator in (146) is

$$\text{Var}(A - B) = \text{Var}(A) + \text{Var}(B) - 2\,\text{Cov}(A, B) \tag{147}$$

so the variance will be low when $\text{Cov}(A, B)$ is large. We can expect this to happen when $q_m(y|d)$ is a good approximation to the true marginal density $p(y|d)$.

Finally, note that just like $\hat{\mu}_{\text{VNMC}}$, this estimator is consistent, i.e. it will converge to the EIG as $N, M \to \infty$.

G KL $(\, q \,||\, p \,)$ versus KL $(\, p \,||\, q \,)$

In Appendix A.1, we showed that our posterior estimator is implicitly minimizing the following expected KL divergence

$$\text{EIG}(d) - \mathcal{L}_{\text{post}}(d) = \mathbb{E}_{p(y|d)}\left[\text{KL}\left(\, p(\theta|y, d) \,||\, q_p(\theta|y, d) \,\right)\right]. \tag{148}$$

In variational inference, the inner KL divergence is referred to as the *forward* KL. In this section, we compare our approach with a similar approach which also uses a posterior approximation, but instead minimize the *reverse* KL divergence, KL $(\, q_p(\theta|y, d) \,||\, p(\theta|y, d) \,)$.

Specifically, we explore how the reverse KL divergence exhibits discontinuous behaviour that could be problematic in the context of EIG estimation. We begin by writing the posterior estimator as

$$\mathcal{L}_{\text{post}}(d) = \mathbb{E}_{p(y|d)}\left[\mathbb{E}_{p(\theta|y)}[\log q_p(\theta|y, d)]\right] + H[p(\theta)]. \tag{149}$$

The term involving q_p is the expectation of the partial KL, $\mathbb{E}_{p(\theta|y)}\left[\log q_p(\theta|y, d)\right]$. We will show that reverse KL minimization can lead to a discontinuity in the partial KL.

We consider two possible methods for choosing q_p. We know from (148) that the optimal choice of q_p within a variational family $\mathcal{Q}$ is

$$q_{\text{forward}}(\theta|y,d) \triangleq \underset{q\in\mathcal{Q}}{\arg\min}\, \text{KL}\left(\, p(\theta|y,d) \,||\, q(\theta) \,\right). \tag{150}$$

An alternative choice is

$$q_{\text{reverse}}(\theta|y,d) \triangleq \underset{q\in\mathcal{Q}}{\arg\min}\, \text{KL}\left(\, q(\theta) \,||\, p(\theta|y,d) \,\right) \tag{151}$$

which is the form usually seen in variational inference. The posterior method outlined in Section 3 attempts to learn q_{forward} for each y by maximizing the bound $\mathcal{L}_{\text{post}}$. In this appendix, we show that the alternative q_{reverse}, as well as resulting in less accurate EIG estimates in light of (148), can lead to discontinuities in the partial KL.

Minimizing the reverse KL can result in the well-known behaviour of mode-locking—and thus mode-dropping—which in our context can result in significant misestimates of the EIG. Furthermore, since this mode-locking behaviour is discontinuous (so that it can occur for a particular design d but not for a neighbouring design d') it can potentially result in large design-dependent bias in EIG estimation. For a quantitative exploration of this phenomenon for two bimodal posteriors and a Normal family of variational distributions $\mathcal{Q}$ see Figure 11.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Variational Bayesian Optimal Experimental Design
Publication Status	Published
Publication Details	Adam Foster, Martin Jankowiak, Eli Bingham, Paul Horsfall, Yee Whye Teh, Tom Rainforth, Noah Goodman (2019). Variational Bayesian Optimal Experimental Design. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Student Confirmation

Student Name:	Adam Foster		
Contribution to the Paper	First author. Led development of the methodology, theory and conducted all the numerical experiments presented in the main paper. Led the writing of the manuscript.		
Signature		Date	25/10/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Dr Tom Rainforth		
Supervisor comments I verify Adam's above account		
Signature 	Date	26/10/21

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 3

A Unified Stochastic Gradient Approach to Designing Bayesian-Optimal Experiments

This paper was published as the following

> Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, and Thomas Rainforth. A Unified Stochastic Gradient Approach to Designing Bayesian-Optimal Experiments. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 2959–2969. PMLR, 2020.

A Unified Stochastic Gradient Approach to Designing Bayesian-Optimal Experiments

Adam Foster[†] **Martin Jankowiak**[‡] **Matthew O'Meara**[§] **Yee Whye Teh**[†] **Tom Rainforth**[†✠]

[†]Department of Statistics, University of Oxford, Oxford, UK
[‡]Uber AI, San Francisco, CA, USA
[§]University of Michigan, Ann Arbor, MI, USA
[✠]Christ Church, University of Oxford, Oxford, UK
adam.foster@stats.ox.ac.uk

Abstract

We introduce a fully stochastic gradient based approach to Bayesian optimal experimental design (BOED). Our approach utilizes variational lower bounds on the expected information gain (EIG) of an experiment that can be simultaneously optimized with respect to both the variational and design parameters. This allows the design process to be carried out through a single unified stochastic gradient ascent procedure, in contrast to existing approaches that typically construct a pointwise EIG estimator, before passing this estimator to a separate optimizer. We provide a number of different variational objectives including the novel adaptive contrastive estimation (ACE) bound. Finally, we show that our gradient-based approaches are able to provide effective design optimization in substantially higher dimensional settings than existing approaches.

1 INTRODUCTION

The design of experiments is a key problem in almost every scientific discipline. Namely, one wishes to construct an experiment that is most informative about the investigated process, while minimizing its cost. For example, in a psychological trial, we want to ensure questions posed to participants are pertinent and do not have predictable responses. In a pharmaceutical trial, we want to minimize the number of participants needed to test our hypotheses. In an online automated help system, we want to ensure we ask questions that identify the user's problem as quickly as possible.

In all these scenarios, our ultimate high-level aim is to choose designs that maximize the information gathered by the experiment. A powerful and broadly used approach for formalizing this aim is Bayesian optimal experimental design (BOED) (Chaloner and Verdinelli, 1995; Lindley, 1956; Myung et al., 2013). In BOED, we specify a Bayesian model for the experiment and then choose the design that maximizes the expected information gain (EIG) from running it. More specifically, let θ denote the latent variables we wish to learn about from running the experiment and let $\xi \in \Xi$ represent the experimental design. By introducing a prior $p(\theta)$ and a predictive distribution $p(y|\theta, \xi)$ for experiment outcomes y, we can calculate the EIG under this model by taking the expected reduction in posterior entropy

$$I(\xi) \triangleq \mathbb{E}_{p(y|\xi)} \left[H[p(\theta)] - H[p(\theta|y, \xi)] \right], \qquad (1)$$

where $H[\cdot]$ represents the entropy of a distribution and $p(\theta|y, \xi) \propto p(\theta)p(y|\theta, \xi)$. Our experimental design process now becomes that of the finding the design ξ^* that maximizes $I(\xi)$.

Unfortunately, finding ξ^* is typically a very challenging problem in practice. Even evaluating $I(\xi)$ for a single design is computationally difficult because it represents a *nested* expectation and thus has no direct Monte Carlo estimator (Rainforth et al., 2018; Zheng et al., 2018). Though a large variety of approaches for performing this estimation have been suggested (Myung et al., 2013; Watson, 2017; Kleinegesse and Gutmann, 2018; Foster et al., 2019), the resulting BOED strategies share a critical common feature: they estimate $I(\xi)$ on a point-by-point basis and feed this estimator to an outer-level optimizer that selects the design.

This framework can be highly inefficient for a number of reasons. For example, it adds an extra level of nest-

Proceedings of the 23[rd] International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

ing to the overall computation process: $I(\xi)$ must be separately estimated for each ξ, substantially increasing the overall computational cost. Furthermore, one must typically resort to gradient-free methods to carry out the resulting optimization, which means it is difficult to scale the overall BOED process to high dimensional design settings due to a dearth of optimization schemes which remain effective in such settings.

To alleviate these inefficiencies and open the door to applying BOED in high-dimensional settings, we introduce an alternative to this two-stage framework by introducing unified objectives that can be directly maximized to simultaneously estimate $I(\xi)$ and optimize ξ. Specifically, by building on the work of Foster et al. (2019), we construct variational lower bounds to $I(\xi)$ that can be simultaneously optimized with respect to both the variational and design parameters. Optimizing the former ensures that we achieve a tight bound that in turn gives accurate estimates of $I(\xi)$, while simultaneously optimizing the latter circumvents the need for an expensive outer optimization process. Critically, this approach allows the optimization to be performed using stochastic gradient ascent (SGA) (Robbins and Monro, 1951) and therefore scaled to substantially higher dimensional design problems than existing approaches.

To account for the varying needs of different problem settings, we introduce several classes of suitable variational lower bounds. Most notably, we introduce the adaptive contrastive estimation (ACE) bound: an EIG variational lower bound that can be made arbitrarily tight, while remaining amenable to simultaneous SGA on both the variational parameters and designs.

We demonstrate[1] the applicability of our unified gradient approach using a wide range of experimental design problems, including a real-world high-dimensional example from the pharmacology literature (Lyu et al., 2019). We find that our approaches are able to effectively optimize the EIG, consistently outperforming baseline two-stage approaches, with particularly large gains achieved for high-dimensional problems. These gains lead, in turn, to improved designs and more informative experiments.

2 BACKGROUND

2.1 Bayesian optimal experimental design

When experimentation is costly, time consuming, or dangerous, it is essential to design experiments to learn the most from them. To choose between potential designs, we require a metric of the quality of a candidate design. In the BOED framework dating back to Lindley (1956), this metric represents how much more certain we will become in our knowledge of the world after doing the experiment and analyzing the data. We prefer designs that will lead to strong conclusions even if we are not yet sure what those conclusions will be.

Specifically, we consider an experiment with design ξ, latent variable θ and outcome y. For example, ξ may represent the question posed to a participant in a psychology trial, y their answer, and θ their underlying psychological characteristic which is being studied. The BOED framework begins with a Bayesian model of the experimental process. This model consists of a likelihood $p(y|\theta, \xi)$ that predicts the experimental outcome under design ξ and latent variable θ and a prior $p(\theta)$ which incorporates initial beliefs about the unknown θ. After conducting the experiment, our beliefs about θ are updated to the posterior $p(\theta|y, \xi)$. The information gained about θ from doing the experiment with design ξ and obtaining outcome y is the reduction in entropy from the prior to the posterior

$$\mathrm{IG}(y, \xi) = H[p(\theta)] - H[p(\theta|y, \xi)]. \tag{2}$$

As it stands, information gain cannot be evaluated until after the experiment. To define a metric that will let us choose between designs before experimentation, we can use the *expected* information gain (EIG), $I(\xi)$, by taking the expectation of IG over hypothesized outcomes y using the marginal distribution under our model, $p(y|\xi)$, to give

$$I(\xi) \triangleq \mathbb{E}_{p(y|\xi)}\left[H[p(\theta)] - H[p(\theta|y, \xi)]\right] \tag{3}$$

which can be rewritten in the form of a mutual information between θ and y with ξ fixed, namely

$$I(\xi) = \mathrm{MI}_{\xi}(\theta; y) = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(y|\theta, \xi)}{p(y|\xi)}\right]. \tag{4}$$

The Bayesian optimal design, ξ^*, is now the one which maximizes EIG over the set of feasible designs Ξ

$$\xi^* = \arg\max_{\xi \in \Xi} I(\xi). \tag{5}$$

In *iterated* experimental design, we design a sequence $\xi_1, ..., \xi_T$ of experiments. At time t, the prior $p(\theta)$ in (4) is replaced by the posterior given the previous experiment designs and observed outcomes, namely

$$p(\theta|\xi_{1:t-1}, y_{1:t-1}) \propto p(\theta) \prod_{\tau=1}^{t-1} p(y_\tau|\theta, \xi_\tau). \tag{6}$$

This now allows us to construct adaptive experiments, wherein we use information gathered from previous iterations to select the designs used at future iterations.

[1] Supporting code is provided at `https://github.com/ae-foster/pyro/tree/sgboed-reproduce`.

2.2 Estimating expected information gain

Making even a single point estimate of EIG when solving (5) can be challenging because we must first estimate the unknown $p(y|\xi)$ or $p(\theta|y,\xi)$, and then take an expectation over $p(\theta)p(y|\theta,\xi)$. Nested Monte Carlo (NMC) estimators (Rainforth et al., 2018), which make a Monte Carlo approximation of both the inner and outer integrals, converge relatively slowly: at a rate $\mathcal{O}(T^{-1/3})$ in the total computational budget T.

Foster et al. (2019) noted that this approach is inefficient because it makes a separate Monte Carlo approximation of the integrand for every sample of the outer integral. To share information between different samples, they proposed a number of variational estimators that used amortization, i.e. they attempted to learn the functional form of the integrand rather than approximating it afresh each time. One of their approaches was based on amortized variational inference and required an *inference network* $q_\phi(\theta|y)$ which takes as input ϕ, y and outputs a distribution over θ. For any $q_\phi(\theta|y)$, we can construct a lower bound on $I(\xi)$. This is the Barber-Agakov (BA), or posterior, lower bound (Barber and Agakov, 2003)

$$I_{BA}(\xi,\phi) \triangleq \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[\log q_\phi(\theta|y)] + H[p(\theta)]\,, \quad (7)$$

which was also used by (Pacheco and Fisher, 2019) and which has found use representation learning (Poole et al., 2019) and maximizing information transmission over noisy channels (Barber and Agakov, 2003).

To make high-quality approximations to $I(\xi)$, and simultaneously learn a good posterior approximation, Foster et al. (2019) maximize this bound with respect to ϕ. This approach is most effective when the bound is tight, i.e. $\max_\phi I_{BA}(\xi,\phi) = I(\xi)$. For $I_{BA}(\xi,\phi)$, this occurs when it is possible to have $q_\phi(\theta|y) = p(y|\theta,\xi)$, i.e. when the inference network is powerful enough to find the true posterior distribution for every y.

To obtain high-quality approximations of $I(\xi)$ even when the inference network cannot capture the true posterior, Foster et al. (2019) also considered another variational estimator: variational nested Monte Carlo (VNMC). This uses the inference network $q_\phi(\theta|y)$ in conjunction with additional samples to improve the estimate of the integrand. They showed that this leads to the following *upper* bound on $I(\xi)$

$$I_{VNMC}(\xi,\phi,L) \triangleq \mathbb{E}\left[\log \frac{p(y|\theta_0,\xi)}{\frac{1}{L}\sum_{\ell=1}^{L}\frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q_\phi(\theta_\ell|y)}}\right], \quad (8)$$

where the expectation is over $p(\theta_0)p(y|\theta_0,\xi)q_\phi(\theta_{1:L}|y)$. The inference network in VNMC is trained by minimization, in the same way I_{BA} is trained by maximization. I_{VNMC} has the attractive feature that the bound becomes tight as $L \to \infty$, even if $q_\phi(\theta_\ell|y)$ is not powerful enough to directly represent the true posterior.

2.3 Optimizing the EIG

The experimental design problem is to find the design that maximizes the EIG. Therefore, as well as finding a way to estimate EIG, existing approaches subsequently need to find a way of searching across Ξ to find promising designs. At a high-level, most existing approaches propose a two-stage procedure in which noisy estimates of $I(\xi)$ are made, and a separate optimization procedure selects the candidate design ξ to evaluate next.

Kleinegesse and Gutmann (2018) and Foster et al. (2019) both use Bayesian optimization (BO) for this outer optimization step, a black-box optimization method that is tolerant to noise in the estimates of the objective function (Snoek et al., 2012), in this case $I(\xi)$. Some approaches (Watson, 2017; Lyu et al., 2019) instead select a finite number of candidate designs in Ξ and estimate $I(\xi)$ at each candidate, with some refining this process further by adaptively allocating computational resources between these designs (Vincent and Rainforth, 2017; Rainforth, 2017). Another suggested approach is to use MCMC methods to carry out this outer optimization (Amzal et al., 2006; Müller, 2005).

3 GRADIENT-BASED BOED

Our central proposal is to replace the two-stage procedure outlined above with a single stage that simultaneously estimates $I(\xi)$ and optimizes ξ. This has the critical advantage of allowing SGA to be directly applied to the design optimization. Not only does this provide substantial computational gains over approaches which must construct separate estimates for each design considered, but it also provides the potential to scale to substantially higher dimensional design problems than those which can be effectively tackled with existing approaches. Since we take gradients with respect to ξ, we henceforth assume that Ξ is continuous.

In our approach, we utilize variational *lower bounds* on I. Specifically, suppose we have a bound $\mathcal{L}(\xi,\phi) \leq I(\xi)$ with variational parameters ϕ. For fixed ξ, the estimate of $I(\xi)$ improves as we maximize with respect to ϕ. We propose to maximize $\mathcal{L}$ *jointly* with respect to (ξ,ϕ). As we train ϕ, the variational approximation improves; as we train ξ our design moves to regions where the lower bound on EIG is largest. By tackling this as a single optimization problem over (ξ,ϕ), we obviate the need to have an outer optimizer for ξ. Using a lower bound is important because it allows us to perform a single maximization over (ξ,ϕ), rather than a more complex

optimization such as the max-min optimization that would result if we used an upper bound.

In practice, we do not have lower bounds on I that we can evaluate and differentiate in closed form. Instead, we have bounds that are expectations over $p(\theta)p(y|\theta,\xi)$. Fortunately, we can still maximize these lower bounds with respect to (ξ,ϕ) by using SGA, which is known to remain effective in high dimensions (Bottou, 2010).

3.1 Barber-Agakov (BA)

We now make our first concrete proposal for the lower bound $\mathcal{L}(\xi,\phi)$: the BA bound I_{BA}, as defined in (7). The difference is we will now optimize (ξ,ϕ) jointly whereas previously only ϕ was trained using gradients. To perform SGA, we use the following unbiased estimators for $\partial I_{BA}/\partial\phi$ and $\partial I_{BA}/\partial\xi$

$$\widehat{\frac{\partial I_{BA}}{\partial\phi}} = \frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\phi}\log q_\phi(\theta_n|y_n), \tag{9}$$

$$\widehat{\frac{\partial I_{BA}}{\partial\xi}} = \frac{1}{N}\sum_{n=1}^{N}\log q_\phi(\theta_n|y_n)\frac{\partial}{\partial\xi}\log p(y_n|\theta_n,\xi) \tag{10}$$

where $\theta_n, y_n \overset{\text{i.i.d.}}{\sim} p(\theta)p(y|\theta,\xi)$. The estimator of $\partial I_{BA}/\partial\xi$ is a score function estimator, other possibilities are discussed in Section 3.6.

3.2 Adaptive contrastive estimation (ACE)

The BA bound provides one specific case of our one-stage procedure for optimal experimental design. We now introduce a new lower bound that improves upon I_{BA}. The potential issue with the BA bound is that it may not be sufficiently tight, which happens when the inference network cannot represent the true posterior. One possible solution is to introduce additional samples, as in the VNMC estimator (8). However, we cannot use VNMC directly for a one-stage procedure: since it is an upper bound, we must minimize it with respect to ϕ, but we still wish to maximize with respect to ξ.

Looking more closely at the VNMC bound, we see that its main failure case is when the denominator strongly *under-estimates* $p(y|\xi)$, which can happen when all the inner samples $\theta_1,...,\theta_L$ miss regions where the joint $p(\theta_\ell)p(y|\theta_\ell,\xi)$ is large. In addition to the samples $\theta_{1:L}$, we also have the original sample θ_0 from which y was sampled. One way to avoid the under-estimation in the denominator would be to include this sample, giving

$$I_{ACE}(\xi,\phi,L) = \mathbb{E}\left[\log\frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q_\phi(\theta_\ell|y)}}\right] \tag{11}$$

where the expectation is with respect to $p(\theta_0)p(y|\theta_0,\xi)q(\theta_{1:L}|y)$. In fact, by including θ_0 we cause the denominator to now *over-estimate* $p(y|\xi)$ which results in a new **lower bound** on $I(\xi)$ which can be jointly maximized with respect to (ξ,ϕ). The samples $\theta_{1:L}$ can now be seen as contrasts to the original sample θ_0. For this reason, we call $\theta_{1:L}$ *contrastive samples* and we call (11) the **adaptive contrastive estimate (ACE)** of EIG. The following theorem establishes that I_{ACE} is a valid lower bound on the EIG which becomes tight as $L\to\infty$.

Theorem 1. *For any model $p(\theta)p(y|\theta,\xi)$ and inference network $q_\phi(\theta|y)$, we have the following:*

1. *I_{ACE} is a lower bound on $I(\xi)$ and we can characterize the error term as an expected KL divergence:*

$$I(\xi) - I_{ACE}(\xi,\phi,L) = \mathbb{E}_{p(y|\xi)}\left[KL\left(P(\theta_{0:L}|y)\middle\|\prod_\ell q_\phi(\theta_\ell|y)\right)\right] \geq 0,$$

$$P(\theta_{0:L}|y) = \frac{1}{L+1}\sum_{\ell=0}^{L}p(\theta_\ell|y,\xi)\prod_{k\neq\ell}q_\phi(\theta_k|y).$$

2. *As $L\to\infty$, we recover the true EIG: $\lim_{L\to\infty} I_{ACE}(\xi,\phi,L) = I(\xi)$.*
3. *The ACE bound is monotonically increasing in L: $I_{ACE}(\xi,\phi,L_2) \geq I_{ACE}(\xi,\phi,L_1)$ for $L_2 \geq L_1 \geq 0$.*
4. *If the inference network equals the true posterior $q_\phi(\theta|y) = p(\theta|y,\xi)$, then $I_{ACE}(\xi,\phi,L) = I(\xi), \forall L$.*

See Appendix A for the proof and additional results. Gradient estimation for ACE is discussed in Section 3.6. We note that, to the best of our knowledge, I_{ACE} has not previously appeared in the BOED literature.[2]

3.3 Prior contrastive estimation (PCE)

Theorem 1 tells us that I_{ACE} can become close to $I(\xi)$ if either: 1) the inference network becomes close to the true posterior $p(\theta|y,\xi)$, 2) we increase the number of contrastive samples L. The BA bound only becomes tight in case 1). A special case of ACE is to replace the inference network $q_\phi(\theta|y)$ with a fixed distribution and rely on the contrastive samples to make good estimates of $I(\xi)$, only becoming tight in case 2), i.e. as $L\to\infty$. This simplification can speed up training, since we no longer need to learn additional parameters ϕ.

To explore this, we propose the **prior contrastive estimation (PCE)** bound, in which the prior $p(\theta)$ is used to generate contrastive samples:

$$I_{PCE}(\xi,L) \triangleq \mathbb{E}\left[\log\frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}p(y|\theta_\ell,\xi)}\right], \tag{12}$$

[2] Aside from a recent blog post (Sobolev, 2019) we believe this bound has not previously been suggested in any context.

where the expectation is over $p(\theta_0)p(y|\theta_0,\xi)p(\theta_{1:L})$. Whilst inherently less powerful than ACE, PCE can be effective when the prior and posterior are similar, such that $p(\theta)$ is a suitable proposal to estimate $p(y|\xi)$.

Though, to the best of our knowledge, this bound has not been applied to BOED before, we note that it shares a connection to the information noise contrastive estimation (InfoNCE) bound on mutual information used in representation learning (van den Oord et al., 2018). Given K data samples x_k, corresponding representations z_k, and a critic $f_\psi(x,z) \geq 0$, we have

$$\mathrm{MI}(x;z) \geq \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K} \log \frac{f_\psi(x_k,z_k)}{\frac{1}{K}\sum_{\ell=1}^{K} f_\psi(x_\ell,z_k)}\right] \quad (13)$$

where the expectation is over $p(x)p(z|x)$, $p(x)$ is the data distribution, and $p(z|x)$ is the encoder. Poole et al. (2019) showed that the encoder density $p(z|x)$ is the optimal critic, although it is rarely known in closed form in the representation learning context. Writing θ for x and y for z, we note the mathematical connection between this optimal case and I_{PCE}.

3.4 Likelihood-free ACE

In some models such as random effects models, the likelihood $p(y|\theta,\xi)$ is not known in closed form but can be sampled from. This presents a problem when computing I_{ACE} or its derivatives because the likelihood appears in (11). To allow ACE to be used for these kinds of models, we now show that using a unnormalized approximation to the likelihood still results in a valid lower bound on the EIG. In fact, if using a parametrized likelihood approximation f_ψ, it is then possible to train ψ jointly with (ξ,ϕ) to approximate the likelihood, learn an inference network, and find the optimal design through the solution to a single optimization problem. The following theorem, whose proof is presented in Appendix A, shows that replacing the likelihood with an unnormalized approximation does result in a valid lower bound on EIG.

Theorem 2. *Consider a model $p(\theta)p(y|\theta,\xi)$ and inference network $q_\phi(\theta|y)$. Let $f_\psi(\theta,y) \geq 0$ be an unnormalized likelihood approximation. Then,*

$$I(\xi) \geq \mathbb{E}\left[\log \frac{f_\psi(\theta_0,y)}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell)f_\psi(\theta_\ell,y)}{q_\phi(\theta_\ell|y)}}\right] \quad (14)$$

where the expectation is over $p(\theta_0)p(y|\theta_0,\xi)q_\phi(\theta_{1:L}|y)$.

3.5 Iterated experimental design with ACE

In iterated experimental design, we replace $p(\theta)$ by $p(\theta|y_{1:t-1},\xi_{1:t-1})$ as per (6). We can sample $p(\theta|y_{1:t-1},\xi_{1:t-1})$ by performing inference. Whilst variational inference also provides a closed form estimate of the posterior density, some other inference methods do not. This is problematic because the prior density appears in (11). Fortunately, it is sufficient to know the density *up to proportionality* (Foster et al., 2019). Indeed if $p(\theta) = A \cdot \gamma(\theta)$ where A does not depend on (ξ,ϕ,y) and γ is an unnormalized density, then

$$I(\xi) \geq \mathbb{E}\left[\log \frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{\gamma(\theta_\ell)p(y|\theta_\ell,\xi)}{q_\phi(\theta_\ell|y)}}\right] - \log A \quad (15)$$

and the derivatives of $\log A$ are simply zero.

3.6 Gradient estimation for ACE

To optimize the ACE bound with respect to (ξ,ϕ) we need unbiased gradient estimators of $\partial I_{ACE}/\partial\xi$ and $\partial I_{ACE}/\partial\phi$. The simplest form of the ξ-gradient is

$$\frac{\partial I_{ACE}}{\partial \xi} = \mathbb{E}\left[\frac{\partial g}{\partial \xi} + g \cdot \frac{\partial}{\partial \xi}\log p(y|\theta_0,\xi)\right] \quad (16)$$

where the expectation is with respect to $p(\theta_0)p(y|\theta,\xi)q(\theta_{1:L}|y)$, and

$$g(y,\theta_{0:L},\phi,\xi) = \log \frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q_\phi(\theta_\ell|y)}}. \quad (17)$$

Estimating the expectation (16) directly using Monte Carlo gives the score function, or REINFORCE, estimator. Unfortunately, this is often high variance, and reducing gradient estimator variance is often important in solving challenging experimental design problems.

One variance reduction method is reparameterization. For this, we introduce random variables $\epsilon, \epsilon'_{1:L}$ which do not depend on (ξ,ϕ) along with representations of y and θ as deterministic functions of these variables: $y = y(\theta_0,\xi,\epsilon)$ and $\theta_\ell = \theta(y,\phi,\epsilon'_\ell)$. This now permits the reparameterized gradient

$$\frac{\partial I_{ACE}}{\partial \xi} = \mathbb{E}\left[\frac{\partial g}{\partial \xi} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial \xi} + \sum_{\ell=1}^{L} \frac{\partial g}{\partial \theta_\ell}\frac{\partial \theta_\ell}{\partial y}\frac{\partial y}{\partial \xi}\right] \quad (18)$$

where the expectation is over $p(\theta_0)p(\epsilon)p(\epsilon'_{1:L})$. A Monte Carlo approximation of this expectation is typically a much lower variance estimator for the true ξ-gradient.

Alternatively, if y is a discrete random variable we can sum over the possible values $\mathcal{Y}$. This approach is known as Rao-Blackwellization and gives

$$\frac{\partial I_{ACE}}{\partial \xi} = \sum_{y\in\mathcal{Y}} \mathbb{E}\left[\frac{\partial g}{\partial \xi}p(y|\theta_0,\xi) + g\,\frac{\partial}{\partial \xi}p(y|\theta_0,\xi)\right] \quad (19)$$

where the expectation is now over $p(\theta_0)\prod_{\ell=1}^{L} q_\phi(\theta_\ell|y)$.

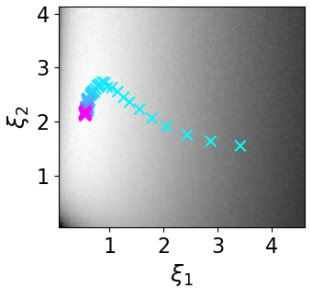


Figure 1: A sample trajectory for the death process. The grayscale shows the EIG surface (white is maximal), whilst crosses show the optimization trajectory of ξ using ACE with pink representing later steps. See Sec. 4.2 for details.

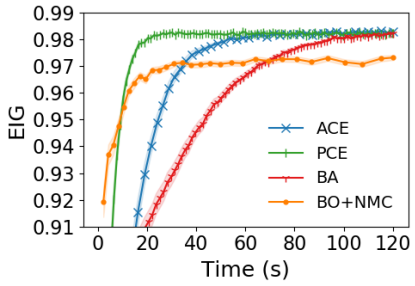


Figure 2: Optimization of EIG for the death process as a function of wall clock time. We depict the mean and ± 1 standard error (s.e.) from 100 runs. The final EIG values (rightmost points) are as follows: [ACE] $\mathbf{0.9830 \pm 0.0001}$, [PCE] 0.9822 ± 0.0001, [BA] 0.9822 ± 0.0002, [BO] 0.9732 ± 0.0009. See Sec. 4.2 for details.

Turning to $\partial I_{ACE}/\partial\phi$, we note that if $\theta_{1:L}$ are reparameterizable (i.e. can be expressed $\theta_\ell = \theta(y, \phi, \epsilon'_\ell)$), then we can utilize the double reparameterization of Tucker et al. (2018); for full details see Appendix A.1.

4 EXPERIMENTS

We now learn optimal experimental designs in five scenarios: the **death process**, a well known two-dimensional design problem from epidemiology; a non-conjugate **regression** model with a 400-dimensional design; an ablation study in the setting of **advertising**; a real-world **biomolecular docking** problem from pharmacology in 100 dimensions; and a **constant elasticity of substitution** iterated design problem in behavioural economics with 6 dimensional designs.

4.1 Evaluating experimental designs

We first discuss which metrics we will use to judge the quality of the designs we obtain. Our primary metric on designs is, of course, the EIG. We prefer designs with high EIGs. In some cases, we can evaluate the EIG analytically. In other cases, we can use a sufficiently large number of samples in a NMC (Rainforth et al., 2018) estimator to be sure that we have estimates that are sufficiently accurate to compare designs.

To explore the limits of our methods, we will also consider scenarios where neither of these approaches is suitable. In these cases, we pair the ACE lower bound (with ξ fixed for evaluation) with the VNMC upper bound (Foster et al., 2019) to trap the true EIG value—if the lower bound of one design is higher than the upper bound for another, we can be sure that the first design is superior (noting that the bounds themselves can be tractably estimated to a very high accuracy).

In some settings, when we know the true optimal design ξ^*, we will also consider the *design error* $\|\xi^* - \xi\|$, i.e. how close our design is to the optimal design.

In iterated experiment design, as well as designing experiments, we must also perform inference on the latent variable θ after each iteration. Here, we also investigate the quality of the final posterior. Specifically, if $p(\theta|y_{1:t}, \xi_{1:t})$ is the posterior after t experiments, we use the posterior entropy, and the posterior RMSE $\mathbb{E}_{\theta\sim p(\theta|y_{1:t},\xi_{1:t})}\left[(\theta-\theta^*)^2\right]^{1/2}$. We prefer low entropies and low RMSE values.

4.2 Death process

We consider an example from epidemiology, the death process (Cook et al., 2008; Kleinegesse and Gutmann, 2018), in which a population of $N = 10$ individuals transitions from healthy to infected states at a constant but unknown rate θ. We can measure the number of infected individuals at two different times ξ_1 and $\xi_1 + \xi_2$ where $\xi_1, \xi_2 \geq 0$. Our aim is to infer the infection rate θ from these observations. For full details of the prior and likelihood used, see Appendix B.2.

On this problem, we apply gradient methods with Rao-Blackwellization over the 66 possible outcomes. Figure 1 shows a sample optimization trajectory with the approximate EIG surface. We compare against BO using the Rao-Blackwellized NMC estimator of Vincent and Rainforth (2017). Figure 2 shows that, for the allowed time budget, all gradient methods perform better than BO even on this two-dimensional problem.

4.3 Regression

We now compare our one-stage gradient approaches to experimental design against a two-stage baseline on a high-dimensional design problem. We choose a general purpose Bayesian linear regression model with n observations and p features. The design ξ is an $n \times p$ matrix; the latent variables are $\theta = (\mathbf{w}, \sigma)$, where $\mathbf{w}$ is the p dimensional regression coefficient and σ^2 is the scalar variance. The n outcomes are generated using a Normal likelihood $y_i \sim N(\boldsymbol{\xi}_i \cdot \mathbf{w}, \sigma)$ for $i = 1, ..., n$. Here $\boldsymbol{\xi}_i$ is the ith row of ξ. To avoid trivial solutions,

Table 1: Regression results. We estimate lower and upper bounds on the final EIG and present the mean and ±1 s.e. from 10 runs. See Sec. 4.3 for details.

Method	EIG l.b.	EIG u.b.
ACE	16.1 ± 0.1	20.7 ± 0.2
PCE	16.6 ± 0.1	21.5 ± 0.2
BA	16.4 ± 0.2	21.1 ± 0.2
BO + VNMC	7.3 ± 0.1	9.6 ± 0.1
Random Search + VNMC	7.1 ± 0.1	9.4 ± 0.1

we enforce the constraint $\|\boldsymbol{\xi}_i\|_1 = 1$ for all i. We use independent priors $w_j \sim \text{Laplace}(1)$ for $j = 1, ..., p$ and $\sigma \sim \text{Exp}(1)$. See Appendix B.3 for complete details.

We set $n = p = 20$ and applied five methods to this 400 dimensional design problem: BA, ACE and PCE, as well as the VNMC estimator of Foster et al. (2019), with both BO and random search to optimize over Ξ. The results are presented in Table 1. We note that the gradient methods strongly outperform the gradient-free baselines, with about double the final EIG.

4.4 Advertising

We now conduct a detailed ablation study on the effects of dimension on the quality of experimental designs produced using our gradient approaches and BO. To further isolate the distinction between one-stage and two-stage approaches to BOED, we choose a setting in which we can compute $I(\xi)$ analytically. We give BO, but not the gradient methods, access to a EIG oracle when making point evaluations of $I(\xi)$, i.e. our two-stage baseline is spared the need to estimate $I(\xi)$. Thus we put BO in the best possible position and ensure any gains are due to improvements from using gradient-based optimization.

Suppose that we are given an advertising budget of B dollars that we need to allocate among D regions, i.e. we choose $\boldsymbol{\xi} \geq \mathbf{0}$ with $\sum_{i=1}^{D} \xi_i = B$. After conducting an ad campaign, we observe a vector of sales $\boldsymbol{y}$. We use this data to make inferences about the underlying market opportunities $\boldsymbol{\theta}$ in each region. Our prior incorporates the knowledge that neighbouring regions are more correlated than distant ones—this leads to an interesting experimental design problem because information can be pooled between regions. We can also compute the true EIG and optimal design ξ^* analytically. For full details, see Appendix B.4.

We compare the performance of four estimation and optimization methods on this problem, see Fig. 3 for the results. The three gradient-based methods (ACE, PCE, BA) perform best, with the BO baseline struggling in dimensions $D \geq 6$, even though the latter has access to an EIG oracle. PCE performed well in low dimensions, but degraded as the dimension increases

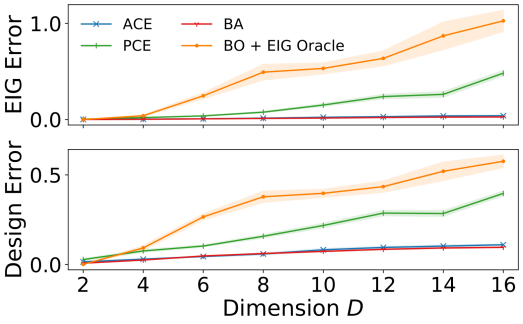


Figure 3: Mean absolute EIG and design errors for the marketing model in Sec. 4.4 averaged over 10 runs. The EIG is normalized such that an EIG error of unity corresponds to doing no better than a uniform budget, i.e. $\xi_i = B/D$ for $i = 1, ..., D$.

and sampling from the prior becomes increasingly inefficient, ACE and BA avoid this by learning adaptive proposal distributions. We note that since in this case the family of variational distributions used in ACE and BA include the true posterior, both methods yield similar performance.

4.5 Biomolecular docking

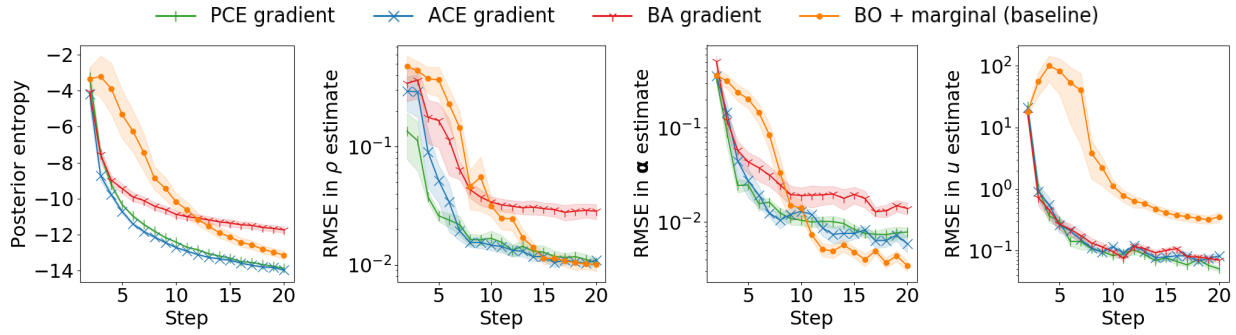
We now consider an experimental design problem of interest to the pharmacology community. Having demonstrated that our one-stage gradient methods compare favourably with two stage approaches, we now compare against designs crafted by domain experts.

In molecular docking, computational techniques are used to predict the binding affinity between a compound and a receptor. When synthesized in the lab, the two may bind—this is called a *hit*. Learning a well-calibrated hit-rate model can guide how many compounds to evaluate for additional objectives, such as drug-likeness or toxicity, before experimental testing. Lyu et al. (2019) modelled the probability of outcome y_i being a hit, given the predicted binding affinity, or docking score $\xi_i \in [-75, 0]$, as

$$p(y_i = 1|\theta, \xi) = \text{bottom} + \frac{\text{top} - \text{bottom}}{1 + e^{-(\xi_i - \text{ee50}) \times \text{slope}}} \quad (20)$$

where $\theta = (\text{top}, \text{bottom}, \text{ee50}, \text{slope})$ with priors given in Appendix B.5.

Of 150 million compounds, Lyu et al. (2019) selected a batch of compounds to experimentally test to best fit the sigmoid hit-rate model. They considered 6 candidate designs and selected one that maximized the EIG estimated by NMC. Here, we instead apply gradient-based BOED to search across candidate designs which consist of 100 docking scores $\xi_1, ..., \xi_{100}$. To evaluate our final designs, we present upper and lower bounds on the final EIG: see Table 2. We see that all gradient methods are able to outperform experts in terms of



(a) Total Entropy (b) Posterior RMSE of ρ (c) Posterior RMSE of $\boldsymbol{\alpha}$ (d) Posterior RMSE of u

Figure 4: Improvement in the posterior in the sequential CES experiment. Each step took 120 seconds for each method. We present the mean and ± 1 standard error from 10 runs. See Sec. 4.6 for details.

Table 2: Biomolecular docking results showing the mean and ± 1 s.e. from 10 runs. For the expert, we took the best design of Lyu et al. (2019) appropriately rescaled to consist of 100 docking scores for comparison.

Method	EIG lower bound	EIG upper bound
ACE	$\mathbf{1.0835 \pm 0.0003}$	$\mathbf{1.0852 \pm 0.0001}$
PCE	1.0825 ± 0.0002	1.0839 ± 0.0002
BA	1.0780 ± 0.0003	1.0794 ± 0.0003
Expert	1.0191	1.0227

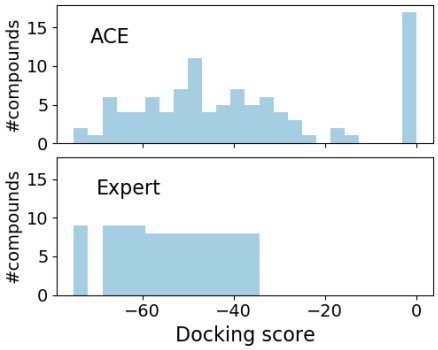


Figure 5: Designs for the biomolecular docking problem obtained by ACE and by Lyu et al. (2019). Designs consist of 100 docking scores at which to test compounds.

EIG, and that ACE appears the best of the gradient methods. Figure 5 shows our designs are qualitatively different to those produced by experts.

4.6 Constant elasticity of substitution

We finally turn to *iterated* experimental design in which we produce designs, generate data and make inference repeatedly. This problem therefore captures the end-to-end-process of experimentation and inference.

We consider an experiment in behavioural economics that was previously also considered by Foster et al. (2019). In this experiment, a participant is asked to compare baskets $\mathbf{x}, \mathbf{x}'$ of goods. The model assumes that their response (on a slider) is based on the difference in utility of the baskets, and the constant elasticity of substitution (CES) model (Arrow et al., 1961) governed by latent variables $(\rho, \boldsymbol{\alpha}, u)$ is then used for this utility. The aim is to learn $(\rho, \boldsymbol{\alpha}, u)$ characterizing the participant's utility. In the experiment, there are 20 sequential steps of experimentation with the same participant. We compare our gradient-based approach against the most successful approach of Foster et al. (2019) that approximates the marginal density to form an upper bound on EIG, and BO to optimize ξ. For full details, see Appendix B.6.

Figure 4 shows that gradient-based methods are effective on this problem; both ACE and PCE decrease the posterior entropy and RMSEs on the latent variables faster and further than the baseline, whereas BA does not do so well. We suggest that the similar performance of ACE and PCE is due to the smaller changes in the posterior at middle and late steps, after much data has been accumulated: when the posterior does not change much at each step, $p(\theta|y_{1:t-1}, \xi_{1:t-1})$ forms an effective proposal for estimating $p(y_t|\xi_t)$.

5 CONCLUSIONS

We have introduced a new approach for Bayesian experimental design that does away with the two stages of estimating EIG and separately optimizing over Ξ. We use stochastic gradients to maximize a lower bound on $I(\xi)$ and so find optimal designs by solving a single optimization problem. This unification leads to substantially improved performance, especially on high-dimensional design problems.

Of the three lower bounds, I_{BA}, I_{ACE} and I_{PCE}, we note that in all five experiments ACE generally did as well as the better of BA and PCE: we therefore recommend it as the default choice. BA performed well when the inference network could closely approximate the true posterior; PCE performed well when the prior was an adequate proposal for estimating $p(y|\xi)$ and does not require the training of variational parameters.

Acknowledgements

AF gratefully acknowledges funding from EPSRC grant no. EP/N509711/1. YWT's research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071. TR gratefully acknowledges funding from Tencent AI Labs and a junior research fellowship supported by Christ Church, Oxford.

References

Billy Amzal, Frédéric Y Bois, Eric Parent, and Christian P Robert. Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical association*, 101(474):773–785, 2006.

Kenneth J Arrow, Hollis B Chenery, Bagicha S Minhas, and Robert M Solow. Capital-labor substitution and economic efficiency. *The review of Economics and Statistics*, pages 225–250, 1961.

David Barber and Felix Agakov. The IM algorithm: a variational approach to information maximization. *Advances in Neural Information Processing Systems*, 16:201–208, 2003.

Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, 2018.

Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders. *arXiv preprint arXiv:1509.00519*, 2015.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

Alex R Cook, Gavin J Gibson, and Christopher A Gilligan. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

Monroe D Donsker and SR Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.

Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems 32*, pages 14036–14047. Curran Associates, Inc., 2019.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Steven Kleinegesse and Michael Gutmann. Efficient Bayesian experimental design for implicit models. *arXiv preprint arXiv:1810.09912*, 2018.

Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.

Jiankun Lyu, Sheng Wang, Trent E Balius, Isha Singh, Anat Levit, Yurii S Moroz, Matthew J O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, et al. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224, 2019.

Peter Müller. Simulation based optimal design. *Handbook of Statistics*, 25:509–518, 2005.

Jay I Myung, Daniel R Cavagnaro, and Mark A Pitt. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, 2013.

Jason Pacheco and John Fisher. Variational information planning for sequential decision making. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2028–2036, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

Ben Poole, Sherjil Ozair, Aäron van den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International Conference on Machine Learning*, pages 5171–5180, 2019.

Tom Rainforth. *Automating Inference, Learning, and Design using Probabilistic Programming*. PhD thesis, University of Oxford, 2017.

Tom Rainforth, Robert Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting Monte Carlo estimators. In *International Conference on Machine Learning*, pages 4264–4273, 2018.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

Artem Sobolev. Thoughts on mutual information: More estimators, 2019. URL http://artem.sobolev.name/posts/2019-08-10-thoughts-on-mutual-information-more-estimators.html.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

George Tucker, Dieterich Lawson, Shixiang Gu, and Chris J Maddison. Doubly reparameterized gradient estimators for monte carlo objectives. *arXiv preprint arXiv:1810.04152*, 2018.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Benjamin T Vincent and Tom Rainforth. The DARC toolbox: automated, flexible, and efficient delayed and risky choice experiments using bayesian adaptive design. 2017.

Andrew B Watson. Quest+: A general multidimensional bayesian adaptive psychometric method. *Journal of Vision*, 17(3):10–10, 2017.

Sue Zheng, Jason Pacheco, and John Fisher. A robust approach to sequential information theoretic planning. In *International Conference on Machine Learning*, pages 5941–5949, 2018.

A GRADIENT-BASED BOED

We begin with the proof of Theorem 1, which we restate for convenience.

Theorem 1. *For any model $p(\theta)p(y|\theta,\xi)$ and inference network $q_\phi(\theta|y)$, we have the following:*

1. *I_{ACE} is a lower bound on $I(\xi)$ and we can characterize the error term as an expected KL divergence:*

$$I(\xi) - I_{ACE}(\xi,\phi,L) = \mathbb{E}_{p(y|\xi)}\left[KL\left(P(\theta_{0:L}|y)\middle|\middle|\prod_\ell q_\phi(\theta_\ell|y)\right)\right] \geq 0,$$

$$P(\theta_{0:L}|y) = \frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|y,\xi)\prod_{k\neq\ell} q_\phi(\theta_k|y).$$

2. *As $L\to\infty$, we recover the true EIG: $\lim_{L\to\infty} I_{ACE}(\xi,\phi,L) = I(\xi)$.*
3. *The ACE bound is monotonically increasing in L: $I_{ACE}(\xi,\phi,L_2) \geq I_{ACE}(\xi,\phi,L_1)$ for $L_2 \geq L_1 \geq 0$.*
4. *If the inference network equals the true posterior $q_\phi(\theta|y) = p(\theta|y,\xi)$, then $I_{ACE}(\xi,\phi,L) = I(\xi), \forall L$.*

We add the further technical assumption that $p(\theta)p(y|\theta,\xi)/q_\phi(\theta|y)$ is bounded.

Proof. To begin with 1., we have the error term $\delta = I(\xi) - I_{ACE}(\xi,\phi,L)$ which can be written

$$\delta = \mathbb{E}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q_\phi(\theta_\ell|y)}}{p(y|\xi)}\right] \tag{21}$$

$$= \mathbb{E}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|y)\prod_{k\neq\ell} q_\phi(\theta_k|y)}{\prod_{\ell=0}^{L} q_\phi(\theta_\ell|y)}\right] \tag{22}$$

$$= \mathbb{E}\left[\log \frac{P(\theta_{0:\ell}|y)}{\prod_{\ell=0}^{L} q_\phi(\theta_\ell|y)}\right] \tag{23}$$

where the expectation is over $p(y|\xi)p(\theta_0|y,\xi)\prod_{\ell=1}^{L} q_\phi(\theta_\ell|y)$. Note that the integrand is symmetric under a permutation of the labels $0,...,L$, so its expectation will be the same over the distribution $p(y|\xi)p(\theta_\ell|y,\xi)\prod_{k\neq\ell} q_\phi(\theta_k|y)$. Since $P(\theta_{0:L})$ is a mixture of distributions of this form, this then implies that the expectation will be the same if it is taken over the distribution $p(y|\xi)P(\theta_{0:L})$, yielding

$$\delta = \mathbb{E}_{p(y|\xi)P(\theta_{0:L}|y)}\left[\log \frac{P(\theta_{0:L}|y)}{\prod_{\ell=0}^{L} q_\phi(\theta_\ell|y)}\right] \tag{24}$$

which is the expected KL divergence required. We therefore have $\delta \geq 0$.

For 2., we use that $p(\theta)p(y|\theta,\xi)/q_\phi(\theta|y)$ is bounded. The ACE denominator is a consistent estimator of the marginal likelihood. Indeed,

$$\frac{1}{L+1}\frac{p(\theta_0)p(y|\theta_0,\xi)}{q_\phi(\theta_0|y)} \to 0 \tag{25}$$

and

$$\frac{1}{L+1}\sum_{\ell=1}^{L}\frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q_\phi(\theta_\ell|y)} \to p(y|\xi) \text{ a.s.} \tag{26}$$

as $L\to\infty$ by the Strong Law of Large Numbers, since

$$\mathbb{E}_{q_\phi(\theta|y)}\left[\frac{p(\theta)p(y|\theta,\xi)}{q_\phi(\theta|y)}\right] = p(y|\xi). \tag{27}$$

This establishes the a.s. pointwise convergence of the ACE integrand to $\log p(y|\theta_0,\xi)/p(y|\xi)$. Hence by Bounded Convergence Theorem,

$$\hat{I}_{ACE}(\xi,\phi,L) \to I(\xi) \tag{28}$$

as $L\to\infty$.

To establish 3., we use a similar approach to 1. We let $\varepsilon = I_{ACE}(\xi,\phi,L_2) - I_{ACE}(\xi,\phi,L_1)$. Then

$$\varepsilon = \mathbb{E}\left[\log \frac{\frac{1}{L_1+1}\sum_{\ell=0}^{L_1}\frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q(\theta_\ell|y)}}{\frac{1}{L_2+1}\sum_{\ell=0}^{L_2}\frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q(\theta_\ell|y)}}\right] \tag{29}$$

$$= \mathbb{E}\left[\log \frac{Q(\theta_{0:L_2}|y)}{\frac{1}{L_2+1}\sum_{\ell=0}^{L_2} p(\theta_\ell|y)\prod_{k\neq\ell} q(\theta_k|y)}\right] \tag{30}$$

where the expectation is over $p(y|\xi)p(\theta_0|y,\xi)\prod_{\ell=1}^{L_2} q(\theta_\ell|y)$ and

$$Q(\theta_{0:L_2}|y) = \frac{1}{L_1+1}\sum_{\ell=0}^{L_1} p(\theta_\ell|y)\prod_{k\neq\ell}^{L_2} q(\theta_k|y). \tag{31}$$

As in 1., the integrand is unchanged if we permute the labels $0,...,L_1$. By this symmetry, the expectation is the same when taken over the distribution $p(y|\xi)Q(\theta_{0:L_2}|y)$. We therefore recognise ε as the expectation of a KL divergence. Hence $\varepsilon \geq 0$ as required.

4. follows by Bayes Theorem, i.e.

$$\frac{p(\theta)p(y|\theta,\xi)}{p(\theta|y,\xi)} = p(y|\xi). \tag{32}$$

which completes the proof. □

We also present the proof of Theorem 2.

Theorem 2. *Consider a model $p(\theta)p(y|\theta,\xi)$ and inference network $q_\phi(\theta|y)$. Let $f_\psi(\theta,y) \geq 0$ be an unnormalized likelihood approximation. Then,*

$$I(\xi) \geq \mathbb{E}\left[\log \frac{f_\psi(\theta_0,y)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(\theta_\ell)f_\psi(\theta_\ell,y)}{q_\phi(\theta_\ell|y)}}\right] \tag{14}$$

where the expectation is over $p(\theta_0)p(y|\theta_0,\xi)q_\phi(\theta_{1:L}|y)$.

Proof. Initially, we note that the contrastive samples $\theta_1, ..., \theta_L$ do not carry additional information about θ_0. Formally, we consider the mutual information between θ_0 and the random variable $(y, \theta_1, ..., \theta_L)$. Using the Chain Rule for mutual information we have

$$\begin{aligned} &\mathrm{MI}(\theta_0; (y, \theta_1, ..., \theta_L)) \\ &= \mathrm{MI}(\theta_0; y) + \mathrm{MI}(\theta_0; (\theta_1, ..., \theta_L)|y) \end{aligned} \tag{33}$$

Now $\mathrm{MI}(\theta_0; (\theta_1, ..., \theta_L)|y) = 0$ since θ_ℓ $(\ell > 0)$ are conditionally independent of θ_0 given y. Therefore

$$\mathrm{MI}(\theta_0; (y, \theta_1, ..., \theta_L)) = \mathrm{MI}(\theta_0; y) = I(\xi). \tag{34}$$

We now use the Donsker-Varadhan representation of mutual information (Donsker and Varadhan, 1975). Specifically, for random variables A, B with joint distribution $p(a, b)$ and any measurable function $T(a, b)$ we have

$$\begin{aligned} &\mathrm{MI}(A; B) \\ &\geq \mathbb{E}_{p(a,b)}[T(a, b)] - \log \mathbb{E}_{p(a)p(b)}\left[e^{T(a,b)}\right]. \end{aligned} \tag{35}$$

We now use this representation with $a = \theta_0, b = (y, \theta_1, ..., \theta_L)$ and $T(a, b)$ the integrand

$$T(\theta_0, (y, \theta_{1:L})) = \log \frac{f_\psi(\theta_0, y)}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell) f_\psi(\theta_\ell, y)}{q_\phi(\theta_\ell|y)}}. \tag{36}$$

We compute the second term in (35), $Z = \mathbb{E}_{p(a)p(b)}\left[e^{T(a,b)}\right]$.

$$Z = \mathbb{E}_{p(\theta_0)p(y|\xi)q_\phi(\theta_{1:L}|y)}\left[\frac{f_\psi(\theta_0, y)}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell) f_\psi(\theta_\ell, y)}{q_\phi(\theta_\ell|y)}}\right] \tag{37}$$

$$= \mathbb{E}_{p(y|\xi)q_\phi(\theta_{0:L}|y)}\left[\frac{\frac{p(\theta_0) f_\psi(\theta_0, y)}{q_\phi(\theta_0|y)}}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell) f_\psi(\theta_\ell, y)}{q_\phi(\theta_\ell|y)}}\right] \tag{38}$$

$$= \mathbb{E}_{p(y|\xi)q_\phi(\theta_{0:L}|y)}\left[\frac{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell) f_\psi(\theta_\ell, y)}{q_\phi(\theta_\ell|y)}}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta_\ell) f_\psi(\theta_\ell, y)}{q_\phi(\theta_\ell|y)}}\right] \tag{39}$$

$$= 1 \tag{40}$$

where the second to last line follows by symmetry. This establishes that $\log Z = 0$, and so (14) constitutes a valid lower bound on $I(\xi)$. That is

$$I(\xi) \geq \mathbb{E}\left[\log \frac{f_\psi(y, \theta_0)}{\frac{1}{L+1}\sum_{\ell=0}^{L} \frac{p(\theta) f_\psi(y, \theta_\ell)}{q_\phi(\theta_\ell, y)}}\right] \tag{41}$$

which completes the proof. □

The following theorem establishes a condition under which the maximum of the ACE objective converges to the maximum of the EIG as $L \to \infty$.

Theorem 3. *Consider a model $p(\theta)p(y|\theta, \xi)$ such that*

$$C \triangleq \sup_{\xi\in\Xi} \inf_{\phi\in\Phi} \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\frac{p(\theta|y, \xi)}{q_\phi(\theta|y, \xi)}\right] < \infty. \tag{42}$$

and $I^* \triangleq \sup_{\xi\in\Xi} I(\xi) < \infty$. Let $q_\phi(\theta|y)$ be an inference network and let

$$I_L = \sup_{\xi\in\Xi,\phi\in\Phi} I_{ACE}(\xi, \phi, L). \tag{43}$$

Then,

$$0 \leq I^* - I_L \leq \frac{C - 1}{L + 1} \tag{44}$$

and in particular $I_L \to I^*$ as $L \to \infty$.

Proof. We have $0 \leq I^* - I_L$ since I_{ACE} is a lower bound on $I(\xi)$ by Theorem 1.

Next, we consider $\Delta(\xi, \phi, L) = I(\xi) - I_{ACE}(\xi, \phi, L)$. We have

$$\Delta = \mathbb{E}_{p(\theta_0)p(y|\theta_0,\xi)q_\phi(\theta_{1:L}|y)}\left[\log \frac{Y_L}{p(y|\xi)}\right] \tag{45}$$

where

$$Y_L = \frac{1}{L+1}\sum_{\ell=0}^{L} w_\ell \quad \text{and} \quad w_\ell = \frac{p(\theta_\ell)p(y|\theta_\ell, \xi)}{q_\phi(\theta_\ell|y)}; \tag{46}$$

we write (45) as

$$\Delta = \mathbb{E}\left[\log\left(1 + \frac{Y_L - p(y|\xi)}{p(y|\xi)}\right)\right] \tag{47}$$

and we apply the inequality $\log(1 + x) \leq x$ to give

$$\Delta \leq \mathbb{E}\left[\frac{Y_L - p(y|\xi)}{p(y|\xi)}\right]. \tag{48}$$

We now observe that for $\ell > 0$, $\mathbb{E}_{q_\phi(\theta_\ell|y)}[w_\ell] = p(y|\xi)$ and hence, taking a partial expectation over $\theta_{1:L}$ we have

$$\Delta \leq \mathbb{E}_{p(\theta_0)p(y|\theta_0,\xi)}\left[\frac{w_0 - p(y|\xi)}{(L+1)p(y|\xi)}\right] \tag{49}$$

$$\leq \frac{1}{L+1}\left(\mathbb{E}_{p(\theta_0)p(y|\theta_0,\xi)}\left[\frac{p(\theta_0|y, \xi)}{q_\phi(\theta_0|y)}\right] - 1\right) \tag{50}$$

Hence

$$I^* - I_L = \sup_{\xi\in\Xi} I(\xi) - \sup_{\xi\in\Xi,\phi\in\Phi} I_{ACE}(\xi, \phi, L)] \tag{51}$$

$$\leq \sup_{\xi\in\Xi}[I(\xi) - \sup_{\phi\in\Phi} I_{ACE}(\xi, \phi, L)] \tag{52}$$

$$\leq \sup_{\xi\in\Xi} \inf_{\phi\in\Phi}[\Delta(\xi, \phi, L)] \tag{53}$$

$$\leq \frac{C - 1}{L + 1} \tag{54}$$

as required. □

A.1 Double reparametrization

We have the ϕ-gradient of the ACE objective

$$\frac{\partial I_{ACE}}{\partial \phi} = \mathbb{E}_{p(\theta_0)p(y|\theta_0,\xi)}\left[-\frac{\partial \mathcal{L}}{\partial \phi}\bigg|_{\theta_0,y}\right] \tag{55}$$

where $\mathcal{L}$ is our estimate of the marginal likelihood with gradient

$$\frac{\partial \mathcal{L}}{\partial \phi}\bigg|_{\theta_0,y} = \frac{\partial}{\partial \phi}\mathbb{E}_{q_\phi(\theta_{1:L}|y)}\left[\log\left(\sum_{\ell=0}^{L} w_\ell\right)\middle|\,\theta_0, y\right] \tag{56}$$

where

$$w_\ell = \frac{p(\theta_\ell)p(y|\theta_\ell,\xi)}{q_\phi(\theta_\ell|y)}. \tag{57}$$

If $q_\phi(\theta|y)$ is reparameterizable as a function of ϕ, then we can apply *double* reparameterization to this gradient. Indeed, were it not for the w_0 term, this would be exactly the IWAE of Burda et al. (2015). We exploit the double reparameterization of Tucker et al. (2018) with a minor variation to account for w_0 to obtain a low variance gradient estimator.

The doubly reparametrized gradient for ACE takes the form

$$\frac{\partial I_{ACE}}{\partial \phi} = \mathbb{E}_{p(\theta_0)p(y|\theta_0,\xi)q_\phi(\theta_{1:L}|y)}\left[\sum_{\ell=0}^{L} v_\ell\right] \tag{58}$$

where

$$v_0 = \frac{w_0}{\sum_{m=0}^{L} w_m}\frac{\partial}{\partial \phi}\log q_\phi(\theta_0|y) \tag{59}$$

and for $\ell > 0$

$$v_\ell = -\left(\frac{w_\ell}{\sum_{m=0}^{L} w_m}\right)^2 \frac{\partial \log w_\ell}{\partial \theta_\ell}\frac{\partial \theta_\ell}{\partial \phi}. \tag{60}$$

A.2 Alternative gradient

We begin with an observation: the true integrand when computing the EIG as an expectation over $p(\theta)p(y|\theta,\xi)$ is given by

$$g_*(y,\theta,\xi) = \log\frac{p(y|\theta,\xi)}{p(y|\xi)}. \tag{61}$$

Recall the score function identity

$$\mathbb{E}_{p(x|\xi)}\left[\frac{\partial}{\partial \xi}\log p(x|\xi)\right] = 0. \tag{62}$$

We have

$$\begin{aligned}
&\mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\frac{\partial g_*}{\partial \xi}\right] && (63)\\
&= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\frac{\partial}{\partial \xi}\log\frac{p(y|\theta,\xi)}{p(y|\xi}\right] && (64)\\
&= \mathbb{E}_{p(\theta)}\left(\mathbb{E}_{p(y|\theta,\xi)}\left[\frac{\partial}{\partial \xi}p(y|\theta,\xi)\right]\right) && (65)\\
&\quad - \mathbb{E}_{p(y|\xi)}\left[\frac{\partial}{\partial \xi}\log p(y|\xi)\right] \\
&= 0 && (66)
\end{aligned}$$

by two applications of the score function identity. This suggests that, as g becomes close to g_*, the $\partial g/\partial \xi$ term in (16) has expectation close to zero, and primarily contributes variance to the gradient estimator.

Theorem 2 shows that if we remove the $\partial g/\partial \xi$ term, the resulting algorithm still optimizes a valid lower bound on $I(\xi)$. Specifically, removing this term is equivalent to the following gradient-coordinate algorithm. First, we choose the family $f_\psi(\theta, y)$ to be $p(y|\theta,\psi)$. Then at time step t we do the following

1. Set $\psi_t = \xi_t$
2. Take a gradient step with respect to (ξ, ϕ) to update ξ_t, ϕ_t

Importantly, the new gradient does not include a $\partial g/\partial \xi$ term, but is the gradient of a valid lower bound on EIG. In practice, this alternative gradient did not yield substantially different performance from the standard approach of including the $\partial g/\partial \xi$ term. All our experiments used the standard approach for simplicity.

B EXPERIMENTS

B.1 Implementation

All experiments were implemented in PyTorch 1.4.0 (Paszke et al., 2019) and Pyro 0.3.4 (Bingham et al., 2018). Supporting code can be found at `https://github.com/ae-foster/pyro/tree/sgboed-reproduce`, see 'README.md' for details on how to run the experiments.

B.2 Death process

We place the prior $\theta \sim \text{LogNormal}(0,1)$ on the infection rate and have the likelihood

$$\begin{aligned}
I_1 &\sim \text{Binomial}(N, e^{-\theta\xi_1})\\
I_2 &\sim \text{Binomial}(N - I_1, e^{-\theta\xi_2}).
\end{aligned} \tag{67}$$

We also have the constraint $\xi_1, \xi_2 \geq 0$.

Table 3: Death process. We present the final EIG for each method (computed using NMC with 200000 samples).

Method	EIG mean ± 1 s.e.
ACE	$\mathbf{0.9830 \pm 0.0001}$
PCE	0.9822 ± 0.0001
BA	0.9822 ± 0.0002
ACE without RB	0.9789 ± 0.0006
PCE without RB	0.9710 ± 0.0025
BA without RB	0.9322 ± 0.0045
BO with NMC	0.9732 ± 0.0009

For each method, we fixed a computational budget of 120 seconds, and did 100 independent runs. For gradient methods, we used the Adam optimizer (Kingma and Ba, 2014) with learning rate 10^{-3} and the default momentum parameters. The inference network made a separate Gaussian approximation to the posterior for each of the 66 outcomes. To evaluate $I(\xi)$ for comparison we used NMC with a large number of samples: 20000 for Figure 2 and 200000 for the final values in the caption and in Table 3. For the BO, we used a Matern52 kernel with variance 1 and lengthscale 0.25, and the GP-UCB1 algorithm (Srinivas et al., 2009) for acquisition.

We used the following number of samples for our Rao-Blackwellized estimators

Method	Number of samples
ACE	10 + 660
PCE	10
BA	10
NMC	2000

B.3 Regression

We consider the following prior on $\theta = (\mathbf{w}, \sigma)$

$$w_j \overset{\text{i.i.d.}}{\sim} \text{Laplace}(1) \text{ for } j = 1, ..., p \tag{68}$$
$$\sigma \sim \text{Exponential}(1) \tag{69}$$

with the likelihood

$$y_i \sim N\left(\sum_{j=1}^{p} \xi_{ij} w_j, \sigma\right) \text{ for } i = 1, ...n. \tag{70}$$

This represents a standard regression model, although with non-Gaussian prior distributions we cannot compute the posterior or true EIG analytically. To ensure the EIG has a finite maximum, we impose the following constraint

$$\sum_j |\xi_{ij}| = 1 \text{ for } i = 1, ..., n. \tag{71}$$

In practice, we set $n = p = 20$.

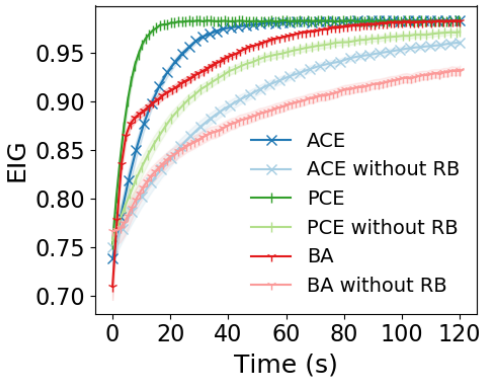


Figure 6: The EIG against time for the death process: comparing Rao-Blackwellization against no Rao-Blackwellization. Each method had a 120 second time budget.

For each of our five methods, we fixed the computational budget to 15 minutes and did 10 independent runs. For gradient methods, we used a learning rate of 10^{-3} and the Adam optimizer with default momentum parameters. The inference network used the following variational family

$$\mathbf{w} \sim N(\boldsymbol{\mu}, s\Sigma_0) \tag{72}$$
$$\sigma \sim \Gamma(\alpha, \beta) \tag{73}$$

and we used a neural network with the following architecture

Operation	Size	Activation
Input $\to$ H1	64	ReLU
H1 $\to$ H2	64	ReLU
H2 $\to \boldsymbol{\mu}$	20	-
H2 $\to (\alpha, \beta)$	2	Softplus
H2 $\to s$	1	Softplus
Σ_0	20×20	-

For BO and random search, point evaluations of $I(\xi)$ were made using VNMC. Each VNMC evaluation took 1000 steps, with the optimization as above (but with ξ fixed). We used a GP with Matern52 kernel with lengthscale 5, variance 10. We used a GP-UCB1 acquisition rule, and terminated once 15 minutes had passed. For random search, we sampled designs using a standard unit Gaussian.

We used the following number of samples

Method	Inner samples L	Outer samples N
ACE	10	10
PCE	10	10
BA	n/a	100
VNMC	10	10

To evaluate designs, we used ACE/VNMC. We first trained ACE using the same procedure as above, for

20000 steps. Then we made the final ACE/VNMC evaluations using the fixed inference network and $L = 2.5 \times 10^3$ inner samples, $N = 10^5$ outer samples.

B.4 Advertising

We introduce a LogNormal likelihood and a D-dimensional latent variable $\boldsymbol{\theta}$ governed by a Normal prior, the joint density of our model is

$$p(\boldsymbol{y}, \boldsymbol{\theta}|\boldsymbol{\xi}) = \mathcal{LN}(\boldsymbol{y}|\boldsymbol{\theta} \odot \boldsymbol{\xi}, \sigma^2 \boldsymbol{\xi}) \mathcal{N}(\boldsymbol{\theta}|\mathbf{0}, \boldsymbol{\Lambda}_0) \tag{74}$$

where σ controls the observation noise, $\boldsymbol{\Lambda}_0$ is a non-diagonal precision matrix and $\odot$ denotes the Hadamard product. Since there are correlations among the D regions, the optimal advertising budget (w.r.t. gaining information about $\boldsymbol{\theta}$) allocates more money to the regions that are tightly correlated.

Throughout we assume that the number of regions D is even. We set the budget to scale with the number of dimensions, $B = \frac{D}{2}$, set $\sigma = 1$ and choose the prior precision matrix to be

$$\boldsymbol{\Lambda}_0 = (1 + \tfrac{1}{D})\mathbb{I}_D - \tfrac{1}{D}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} \qquad \boldsymbol{u}^{\mathrm{T}} \equiv (\alpha, ..., \alpha, 1, ..., 1)$$

where the first $\frac{D}{2}$ components of $\boldsymbol{u}$ equal α and the last $\frac{D}{2}$ components equal 1. We shall see that $\alpha = 0.1$ controls the degree of asymmetry in the optimal design. Discarding an irrelevant constant, we can compute the exact EIG using the formula:

$$I(\boldsymbol{\xi}) = \frac{1}{2} \log\det \boldsymbol{\Lambda}_{\text{post}} \qquad \boldsymbol{\Lambda}_{\text{post}} = \boldsymbol{\Lambda}_0 + \frac{1}{\sigma^2}\text{diag}(\boldsymbol{\xi})$$

Using the matrix determinant lemma for rank-1 matrix updates we can then compute

$$\log\det \boldsymbol{\Lambda}_{\text{post}} = \sum_{i=1}^{D} \log(1 + \tfrac{1}{D} + \xi_i) + \log\left(1 - \sum_{i=1}^{\frac{D}{2}} \left\{\frac{\alpha^2}{1+\frac{1}{D}+\xi_i}\right\} - \sum_{i=1+\frac{D}{2}}^{D} \left\{\frac{1}{1+\frac{1}{D}+\xi_i}\right\}\right).$$

By symmetry the optimum (it is easy to check that it is a maximum) of $\text{EIG}(\boldsymbol{\xi})$ will satisfy $\xi_i = \xi_{i+1}$ for $i = 1, ..., \frac{D}{2} - 1, \frac{D}{2} + 1, ..., D$. In other words $\boldsymbol{\xi}$ is entirely specified by ξ_1 and ξ_D, which must satisfy $\xi_1 + \xi_D = 1$ because of the constraint on the budget $B = \frac{D}{2}$. Thus we have reduced the EIG maximization problem to a univariate optimization problem that can easily be solved to machine precision, for example by gradient methods or brute force bisection. This analytic solution gives us the ground truth EIG, used within BO and for evaluation, and the true optimal design, used for evaluation.

For each of the four methods (ACE, PCE, BA and BO) we fix the computational budget to 120 seconds per design optimization. For the gradient-based methods this corresponds to 1×10^4, 2×10^4, and 1.8×10^4 gradient steps for ACE, PCE, and BA, respectively. For the BO baseline, we run 110 steps of a GP-UCB-like algorithm (Srinivas et al., 2009) in batch-mode, resulting in a total budget of 1650 function evaluations of the EIG oracle. Note that for all four methods the runtime dependence on the dimension D is negligible in the regime in which we are operating; consequently we use the same number of gradient or BO steps for all D.

For the gradient-based methods, we use the Adam optimizer with default momentum hyperparameters and an initial learning rate of $\ell_0 = 0.1$ that is exponentially decayed towards a final learning rate ℓ_f that depends on the particular method. In particular we set $\ell_f = 1 \times 10^{-4}$, $\ell_f = 1 \times 10^{-5}$, and $\ell_f = 3 \times 10^{-4}$ for the ACE, PCE, and BA methods, respectively. For the BO baseline, we used a Matérn kernel with a fixed length scale $\ell = 0.2$. These hyperparameters were chosen by running a grid search with $D = 16$ and choosing hyperparameters that minimized the mean absolute EIG error.

Finally we note that in Fig. 3 at each dimension D we normalize the EIG by the factor

$$Z = \text{EIG}(\boldsymbol{\xi}^*) - \text{EIG}(\boldsymbol{\xi}_{\text{uniform}}) \tag{75}$$

where $\boldsymbol{\xi}^*$ and $\boldsymbol{\xi}_{\text{uniform}}$ are the optimal and uniform budget designs, respectively. Consequently after normalization the absolute error for the uniform budget design $\boldsymbol{\xi}_{\text{uniform}}$ is equal to 1.

B.5 Biomolecular docking

For the docking model, we used the following independent priors

$$\text{top} \sim \text{Beta}(25, 75) \tag{76}$$
$$\text{bottom} \sim \text{Beta}(4, 96) \tag{77}$$
$$\text{ee50} \sim N(-50, 15^2) \tag{78}$$
$$\text{slope} \sim N(-0.15, 0.1^2). \tag{79}$$

For the design $\xi = (\xi_1, ..., \xi_{100})$ we had 100 binary responses

$$y_i \sim \text{Bern}\left(\text{bottom} + \frac{\text{top} - \text{bottom}}{1 + e^{-(\xi_i - \text{ee50}) \times \text{slope}}}\right). \tag{80}$$

For gradient methods, we used the Adam optimizer with learning rate 10^{-3} and default momentum parameters. For each method, we took 5×10^5 gradient steps (each method converged within this number of

steps). The inference network was mean-field with the same distributional families as the prior. We used the following neural architecture

Operation	Size	Activation
Input → H1	64	ReLU
H1 → H2	64	ReLU
H2 → top	2	Softplus
H2 → bottom	2	Softplus
H2 → ee50 mean	1	-
H2 → ee50 s.d.	1	Softplus
H2 → slope mean	1	-
H2 → slope s.d.	1	Softplus

We used the following number of samples

Method	Inner samples L	Outer samples N
ACE	10	10
PCE	10	10
BA	n/a	100

For the expert method, the design of Lyu et al. (2019), which comprised 580 compounds, was subsampled to comprise 100 compounds for a fair comparison.

For evaluation, we used ACE/VNMC, first training ACE for 25000 steps using the same learning rate as above. With the fixed inference network, we made ACE and VNMC evaluations using $L = 2 \times 10^3$ inner samples, $N = 4 \times 10^6$ outer samples.

B.6 Constant elasticity of substitution

We used the exact set-up of Foster et al. (2019). Specifically, we take $U(\mathbf{x}) = \left(\sum_i x_i^\rho \alpha_i\right)^{1/\rho}$ and place the following priors on $\rho, \boldsymbol{\alpha}, u$

$$\rho \sim \text{Beta}(1, 1) \tag{81}$$
$$\boldsymbol{\alpha} \sim \text{Dirichlet}([1, 1, 1]) \tag{82}$$
$$\log u \sim N(1, 3) \tag{83}$$
$$\mu_\eta = u \cdot (U(\mathbf{x}) - U(\mathbf{x}')) \tag{84}$$
$$\sigma_\eta = \tau u \cdot (1 + \|\mathbf{x} - \mathbf{x}'\|) \tag{85}$$
$$\eta \sim N(\mu_\eta, \sigma_\eta^2) \tag{86}$$
$$y = f(\eta) \tag{87}$$

where f is the censored sigmoid function and $\tau = 0.005$. All designs $\xi = (\mathbf{x}, \mathbf{x}')$ were constrained to $[0, 100]^6$.

For gradient methods, we used the Adam optimizer with learning rate 10^{-3} and default momentum parameters. To make the design process 120 seconds per step, we used the following number of gradient steps

Method	Number of steps
ACE	1500
PCE	2500
BA	5000

We found that there was insufficient time to effectively train a neural network guide. Instead we used a mean-field variational family with the same distributional families as the prior, and a linear model using the following features: $\text{logit}(y), \log|\text{logit}(y)|, \mathbf{1}(y > 0.5)$.

We used the following number of samples

Method	Inner samples L	Outer samples N
ACE	10	10
PCE	10	10
BA	n/a	100

For the baseline, we used the marginal upper bound of Foster et al. (2019) with the same variational family used in that paper—an f-transformed Normal with additional point masses at the end-points. We used a GP with a Matérn52 kernel, lengthscale 20, variance set from data, and a GP-UCB1 algorithm to make acquisitions which were done in batches of 8.

At each stage of the sequential experiment, the posterior was fitted using mean-field variational inference using the same distributional families as the prior.

C FUTURE WORK

In this paper, we have focused on continuous design spaces in which gradient methods are applicable. One possible extension of our work would be to facilitate a unified one-stage approach to experimental design over *discrete* design spaces. In this case, the lower bounds I_{BA}, I_{ACE} and I_{PCE} remains valid, and performing a joint maximization over (ξ, ϕ) on any of these objectives may be an attractive choice, although gradient optimization would no longer be appropriate for ξ. We envisage that one could apply existing methods for discrete optimization to the joint optimization problem over design and variational parameters. For instance, a continuous relaxation of the discrete variables, or MCMC-style updates on the discrete variables might be used. Future work might further explore this direction.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	A Unified Stochastic Gradient Approach to Designing Bayesian-Optimal Experiments
Publication Status	Published
Publication Details	Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, Tom Rainforth (2020). A Unified Stochastic Gradient Approach to Designing Bayesian-Optimal Experiments. 23rd International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

Student Confirmation

Student Name:	Adam Foster		
Contribution to the Paper	First author. Led development of the methodology, theory and conducted all the numerical experiments presented in the main paper with the exception of Sec 4.4. Led the writing of the manuscript.		
Signature		Date	25/10/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Dr Tom Rainforth		
Supervisor comments I verify Adam's above account		
Signature 	Date	26/10/21

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 4

Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs

This paper has been accepted for publication in the SIAM Journal on Scientific Computing.

UNBIASED MLMC STOCHASTIC GRADIENT-BASED OPTIMIZATION OF BAYESIAN EXPERIMENTAL DESIGNS*

TAKASHI GODA†, TOMOHIKO HIRONAKA†, WATARU KITADE†, AND ADAM FOSTER‡

Abstract. In this paper we propose an efficient stochastic optimization algorithm to search for Bayesian experimental designs such that the expected information gain is maximized. The gradient of the expected information gain with respect to experimental design parameters is given by a nested expectation, for which the standard Monte Carlo method using a fixed number of inner samples yields a biased estimator. In this paper, applying the idea of randomized multilevel Monte Carlo (MLMC) methods, we introduce an unbiased Monte Carlo estimator for the gradient of the expected information gain with finite expected squared ℓ_2-norm and finite expected computational cost per sample. Our unbiased estimator can be combined well with stochastic gradient descent algorithms, which results in our proposal of an optimization algorithm to search for an optimal Bayesian experimental design. Numerical experiments confirm that our proposed algorithm works well not only for a simple test problem but also for a more realistic pharmacokinetic problem.

Key words. Bayesian experimental design, expected information gain, multilevel Monte Carlo, nested expectation, stochastic gradient descent

AMS subject classifications. 62K05, 62L20, 65C05, 92C45, 94A17

1. Introduction.

In this paper we study optimization of Bayesian experimental designs which aim to maximize the expected amount of information experimental outcomes convey about unobservable, or hidden/latent, random variables of interest by carefully designing an experimental setup. Here we measure the expected amount of information by the Shannon's expected information gain whose definition is given below. Our motivation comes from applications to a number of disciplines, such as mechanical engineering [34], neuroscience [40], bioinformatics [36], psychology [23], and pharmacokinetics [33, 32] among many others.

Let $\theta = (\theta_1, \dots, \theta_s) \in \Theta \subseteq \mathbb{R}^s$ be a vector of continuous unobservable random variables, and we denote the prior probability density of θ by $\pi_0(\theta)$. The information entropy, or the differential entropy, of θ is defined by

$$\mathbb{E}_\theta \left[-\log \pi_0(\theta) \right] = \int_\Theta -\pi_0(\theta) \log \pi_0(\theta) \, \mathrm{d}\theta.$$

Let us consider a situation where, by conducting some experiments under an experimental design ξ, an observation $Y = (Y_1, \dots, Y_t) \in \mathcal{Y} \subseteq \mathbb{R}^t$ is obtained according to the forward model

$$Y = f_\xi(\theta, \epsilon), \tag{1.1}$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_{s'}) \in \mathcal{E} \subseteq \mathbb{R}^{s'}$, representing the observation noise, is another vector of continuous random variables with its density $\varphi(\epsilon)$, and f_ξ is a deterministic bi-variate function parametrized by the design ξ, possibly with multiple outputs. Here we assume that the experimental design ξ is controllable and can be chosen as an element in an open set $\mathcal{X} \subset \mathbb{R}^d$. Throughout this paper, we assume that the

*Submitted to the editors DATE.

Funding: The work of T.G. is supported by JSPS KAKENHI Grant Number 20K0374. The work of A.F. is kindly supported by EPSRC grant no. EP/N509711/1.

†School of Engineering, University of Tokyo, Tokyo, Japan (goda@frcer.t.u-tokyo.ac.jp, hironaka-tomohiko@g.ecc.u-tokyo.ac.jp, kitade-wataru114@g.ecc.u-tokyo.ac.jp).

‡Department of Statistics, University of Oxford, Oxford, UK (adam.foster@stats.ox.ac.uk).

domain $\mathcal{Y}$ is independent of ξ, that ϵ is independent both of θ and ξ, and also that the likelihood function $\rho(Y \mid \theta, \xi)$ is strictly positive and can be computed explicitly with unit cost for any pair of θ, ξ and Y. As is well known, Bayes' theorem states that the posterior probability density of θ given Y, denoted by $\pi^{Y|\xi}$, is given by

$$\pi^{Y|\xi}(\theta) = \frac{\rho(Y \mid \theta, \xi)\pi_0(\theta)}{\rho(Y \mid \xi)}, \tag{1.2}$$

with $\rho(Y \mid \xi)$ being the marginal likelihood of Y, i.e.,

$$\rho(Y \mid \xi) = \mathbb{E}_\theta \left[\rho(Y \mid \theta, \xi)\right] = \int_\Theta \rho(Y \mid \theta, \xi)\pi_0(\theta)\,\mathrm{d}\theta,$$

see for instance [38]. Then, the posterior information entropy of θ after observing Y is given by

$$\mathbb{E}_{\theta|Y,\xi}\left[-\log \pi^{Y|\xi}(\theta)\right] = \int_\Theta -\pi^{Y|\xi}(\theta)\log \pi^{Y|\xi}(\theta)\,\mathrm{d}\theta,$$

and hence, the expected posterior information entropy of θ by conducting an experiment under an experimental design ξ is given by integrating the posterior information entropy of θ over Y using the marginal likelihood $\rho(Y \mid \xi)$, i.e.,

$$\mathbb{E}_{Y|\xi}\mathbb{E}_{\theta|Y,\xi}\left[-\log \pi^{Y|\xi}(\theta)\right] = \int_\mathcal{Y}\int_\Theta -\pi^{Y|\xi}(\theta)\log \pi^{Y|\xi}(\theta)\,\mathrm{d}\theta\,\rho(Y \mid \xi)\,\mathrm{d}Y.$$

Now the difference

$$U(\xi) := \mathbb{E}_\theta\left[-\log \pi_0(\theta)\right] - \mathbb{E}_{Y|\xi}\mathbb{E}_{\theta|Y,\xi}\left[-\log \pi^{Y|\xi}(\theta)\right]$$

is called the *expected information gain*, the quantity originally introduced in [21] as a measure of experimental designs. By using Bayes' theorem (1.2), we see that $U(\xi)$ is equivalently given by

$$\begin{aligned} U(\xi) &= \mathbb{E}_\theta\mathbb{E}_{Y|\theta,\xi}\left[\log \rho(Y \mid \theta, \xi)\right] - \mathbb{E}_{Y|\xi}\left[\log \rho(Y \mid \xi)\right] \\ &= \mathbb{E}_\theta\mathbb{E}_{Y|\theta,\xi}\left[\log \rho(Y \mid \theta, \xi)\right] - \mathbb{E}_{Y|\xi}\left[\log \mathbb{E}_\theta\left[\rho(Y \mid \theta, \xi)\right]\right]. \end{aligned} \tag{1.3}$$

The aim of Bayesian experimental designs is to construct an optimal experimental design $\xi = \xi^*$ which maximizes the expected information gain U [6]. As can be seen from the second term of (1.3), however, estimating $U(\xi)$ is inherently a nested expectation problem with an outer expectation with respect to Y and an inner expectation with respect to θ, which has been considered computationally challenging. The standard, nested Monte Carlo method generates N outer random samples for Y first and then, for each sample of Y, generates M inner random samples for θ. To estimate $U(\xi)$ with root-mean-square accuracy ε,[1] we typically need $N = O(\varepsilon^{-2})$ and $M = O(\varepsilon^{-1})$, resulting in a total computational complexity of $O(\varepsilon^{-3})$ [34, 2, 27]. Recently there have been some attempts in [16, 3] to reduce this cost to $O(\varepsilon^{-2})$ or $O(\varepsilon^{-2}(\log \varepsilon^{-1})^2)$ by applying a multilevel Monte Carlo (MLMC) method [12, 13] in conjunction with Laplace approximation-based importance sampling [22]. Here the difference between the orders of complexity for the MLMC method is a direct consequence from the basic MLMC theorem, see for instance [13, Theorem 2.1], which

[1] Here and in what follows, the difference between the noise ϵ and the accuracy ε should not be confused.

itself depends on the properties of the constructed MLMC estimators. Nevertheless, these results are an intermediate step towards an efficient construction of optimal experimental designs since design optimization has been left behind.

In this paper we deal with this optimization problem more directly. More precisely, under the assumption that the experimental setup, or the set of design parameters, ξ lives in a continuous space such that U is differentiable with respect to ξ, we consider applying stochastic gradient descent optimizations to search for an optimal ξ. As we shall see, the gradient $\nabla_\xi U$ is again given by a nested expectation, for which the standard, nested Monte Carlo method using a fixed number of inner samples yields a biased estimator. By applying an unbiased MLMC method from [29], a randomized version of the original MLMC method, we can construct an unbiased estimator of $\nabla_\xi U$. This way, in this paper, we arrive at a stochastic gradient-based optimization algorithm in which unbiased random samples to estimate $\nabla_\xi U$ are generated at each iteration step.

Here we have to mention that the idea of using stochastic gradient-based methods in Bayesian experimental designs already exists in the literature [18, 9, 10, 5, 20]. In particular, a work by Carlon et al. [5] takes a similar standpoint in that an analytical expression of the gradient $\nabla_\xi U$ is derived and then stochastic gradient-based method is applied in conjunction with Monte Carlo estimation of $\nabla_\xi U$. However, the expression of $\nabla_\xi U$ given in [5, Proposition 1] is proven only for the additive Gaussian noise ϵ, that is, the case where the forward model is given by the form $Y = f_\xi(\theta) + \epsilon$ with $\epsilon \sim N(0, \Sigma)$, and the standard (biased) Monte Carlo estimator is used at each iteration step within stochastic gradient-based methods. In this paper we consider a more general form of the forward model as shown in (1.1), which is useful in some applications [33, 32]. Moreover, given that stochastic gradient-based methods are usually established under the assumption that each sample is drawn from the underlying true distribution, using an unbiased estimator of $\nabla_\xi U$ should be favorable, and by doing so, we do not need to take care of the bias-variance tradeoff. Although application of MLMC methods to stochastic approximation algorithms have been investigated recently in [11, 7], neither of them considers using a randomized MLMC method to generate unbiased random samples at each iteration step.

The rest of this paper is organized as follows. In Section 2, we provide an analytical expression of the gradient $\nabla_\xi U$ and also briefly review some of stochastic gradient-based optimization methods. Although there are a number of stochastic optimization algorithms, one can use any of them in our proposal to optimize Bayesian experimental designs (Algorithm 3.1), and we do not give any recommendation on which method should be used in our algorithm, since it is not the objective of this paper. Again we emphasize that the main contribution of this paper is to provide an unbiased estimator for the gradient $\nabla_\xi U$, which is non-trivial but a key assumption in stochastic gradient-based optimization. In Section 3, after introducing a standard, nested Monte Carlo estimator of $\nabla_\xi U$, which is obviously biased, we provide an unbiased, multilevel Monte Carlo estimator of $\nabla_\xi U$ and prove under some conditions that our estimator has a finite expected squared ℓ_2-norm with finite computational cost per sample. Our proposal for optimizing Bayesian experimental designs is given in Algorithm 3.1. To demonstrate the effectiveness of our proposed algorithm, we conduct numerical experiments not only for a simple test problem but also for a more realistic pharmacokinetic (PK) problem in Section 4. We conclude this paper with some remarks in Section 5.

2. Stochastic gradient-based optimization.

2.1. Gradient of expected information gains. In what follows, we give an explicit form of the gradient $\nabla_\xi U$. As a preparation, let us rewrite the expected information gain $U(\xi)$ according to (1.1) in the following way. First, by noting that generating Y randomly conditional on θ and ξ is equivalent to computing $f_\xi(\theta, \epsilon)$ for a randomly generated ϵ with both θ and ξ given, the independence between ϵ and the pair (θ, ξ) ensures that the first term of (1.3) is equal to

$$\mathbb{E}_\theta \mathbb{E}_\epsilon \left[\log \rho(f_\xi(\theta, \epsilon) \mid \theta, \xi)\right] = \mathbb{E}_{\theta,\epsilon} \left[\log \rho(f_\xi(\theta, \epsilon) \mid \theta, \xi)\right].$$

Similarly, generating Y randomly conditional only on ξ is equivalent to computing $f_\xi(\theta, \epsilon)$ for randomly generated θ and ϵ with a fixed ξ. Therefore, by denoting an i.i.d. copy of θ by θ', the second term of (1.3) is equal to

$$\mathbb{E}_{Y|\xi} \left[\log \mathbb{E}_{\theta'} \left[\rho(Y \mid \theta', \xi)\right]\right] = \mathbb{E}_{\theta,\epsilon} \left[\log \mathbb{E}_{\theta'} \left[\rho(f_\xi(\theta, \epsilon) \mid \theta', \xi)\right]\right].$$

Thus we end up with the following expression of $U(\xi)$:

$$U(\xi) = \mathbb{E}_{\theta,\epsilon} \left[\log \rho(f_\xi(\theta, \epsilon) \mid \theta, \xi)\right] - \mathbb{E}_{\theta,\epsilon} \left[\log \mathbb{E}_{\theta'} \left[\rho(f_\xi(\theta, \epsilon) \mid \theta', \xi)\right]\right]. \tag{2.1}$$

As we have stated in the previous section, we assume throughout this paper that the likelihood function can be computed explicitly with unit cost for any pair of inputs. Here we give some examples for which such an explicit computation of the likelihood function is possible.

Example 2.1 (Additive noise). Let us consider a forward model given by

$$f_\xi(\theta, \epsilon) = g_\xi(\theta) + \epsilon,$$

for a uni-vatiate function $g_\xi : \Theta \to \mathcal{Y} (= \mathbb{R}^t)$ and $\epsilon \sim N(0, \Sigma)$ with a covariance matrix Σ and $s' = t$. Then, denoting the density of ϵ by φ, we have

$$\rho(f_\xi(\theta, \epsilon) \mid \theta', \xi) = \varphi(\epsilon + g_\xi(\theta) - g_\xi(\theta')),$$

with a special case $\rho(f_\xi(\theta, \epsilon) \mid \theta, \xi) = \varphi(\epsilon)$.

Example 2.2 (Multiplicative noise). Let $s' = t = 1$ for simplicity, and consider a forward model given by

$$f_\xi(\theta, \epsilon) = g_\xi(\theta) \times (1 + \epsilon),$$

with $g_\xi : \mathbb{R}^s \to \mathbb{R}_{>0}$ and $\epsilon \sim N(0, \sigma^2)$. Denoting the density of ϵ by φ, we have

$$\rho(f_\xi(\theta, \epsilon) \mid \theta', \xi) = \varphi\left(\frac{g_\xi(\theta)}{g_\xi(\theta')}(1 + \epsilon) - 1\right),$$

with a special case $\rho(f_\xi(\theta, \epsilon) \mid \theta, \xi) = \varphi(\epsilon)$.

Example 2.3 (Mixture of additive and multiplicative noises). Finally, for $t = 1$ and $s' = 2$, i.e., $\epsilon = (\epsilon_1, \epsilon_2) \in \mathbb{R}^2$, let us consider a forward model described by

$$f_\xi(\theta, \epsilon) = g_\xi(\theta) \times (1 + \epsilon_1) + \epsilon_2,$$

with $g_\xi : \mathbb{R}^s \to \mathbb{R}_{>0}$, $\epsilon_1 \sim N(0, \sigma_1^2)$ and $\epsilon_2 \sim N(0, \sigma_2^2)$. Denoting the density of the standard normal random variable by φ, we have

$$\rho(f_\xi(\theta, \epsilon) \mid \theta', \xi) = \frac{1}{\sqrt{|g_\xi(\theta')|^2\sigma_1^2 + \sigma_2^2}} \varphi\left(\frac{g_\xi(\theta) \times (1 + \epsilon_1) + \epsilon_2 - g_\xi(\theta')}{\sqrt{|g_\xi(\theta')|^2\sigma_1^2 + \sigma_2^2}}\right),$$

with a special case

$$\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi) = \frac{1}{\sqrt{|g_\xi(\theta)|^2\sigma_1^2+\sigma_2^2}}\varphi\left(\frac{\epsilon_1 g_\xi(\theta)+\epsilon_2}{\sqrt{|g_\xi(\theta)|^2\sigma_1^2+\sigma_2^2}}\right).$$

Now we are ready to derive the gradient $\nabla_\xi U$. Note that our claim does not assume that the noise ϵ is additive and a Gaussian random variable, as discussed in the last two examples.

PROPOSITION 2.4. *Let θ' be an i.i.d. copy of θ. Assume that the likelihood functions $\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)$ and $\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)$ and their gradients $\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)$ and $\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)$ are all continuous with respect to θ,θ',ϵ and ξ. Then we have*

$$\nabla_\xi U(\xi) = \mathbb{E}_{\theta,\epsilon}\left[\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)} - \frac{\mathbb{E}_{\theta'}\left[\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}{\mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}\right].$$

Proof. Under the continuity assumption on the likelihood function, the Leibniz integral rule applies and we have

$$\begin{aligned}
\nabla_\xi U(\xi) &= \mathbb{E}_{\theta,\epsilon}\left[\nabla_\xi \log \rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)\right] - \mathbb{E}_{\theta,\epsilon}\left[\nabla_\xi \log \mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]\right] \\
&= \mathbb{E}_{\theta,\epsilon}\left[\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)}\right] - \mathbb{E}_{\theta,\epsilon}\left[\frac{\nabla_\xi\mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}{\mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}\right] \\
&= \mathbb{E}_{\theta,\epsilon}\left[\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)}\right] - \mathbb{E}_{\theta,\epsilon}\left[\frac{\mathbb{E}_{\theta'}\left[\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}{\mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}\right].
\end{aligned}$$

∎

As is clear from this proposition, because of the ratio of inner expectations, the gradient $\nabla_\xi U$ is inherently given by a nested expectation with an inner expectation with respect to θ' and an outer expectation with respect to θ and ϵ.

2.2. Basics of stochastic gradient-based optimization. We recall that the aim of Bayesian experimental designs is to find an optimal experimental setup $\xi = \xi^*$ which satisfies

$$\xi^* = \arg\max_{\xi\in\mathcal{X}} U(\xi),$$

where we recall that an open set $\mathcal{X} \subset \mathbb{R}^d$ denotes the feasible domain of ξ. To achieve this goal, one of the reasonable approaches is to use some gradient-based optimization methods in which we set an initial experimental setup $\xi_0 \in \mathcal{X}$ and recursively update itself as

$$\xi_{t+1} = g_t(\xi_t, \nabla_\xi U(\xi_t)) \quad \text{for } t = 0,1,\ldots,$$

until a certain stopping criterion is met. However, computing $\nabla_\xi U$ is already challenging since it is given by a nested expectation. As inferred from the results shown in the next section, it is possible to construct an antithetic MLMC estimator which efficiently estimates $\nabla_\xi U$, but we avoid such a "pointwise" accurate gradient estimation by using *stochastic* gradient-based optimization methods. What we need here is an unbiased estimator of $\nabla_\xi U$ with finite variance and computational cost.

To simplify the presentation, let us define a vector of random variables

$$\psi_\xi := \frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta,\xi)} - \frac{\mathbb{E}_{\theta'}\left[\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}{\mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right]}, \tag{2.2}$$

with $\theta \sim \pi_0$ and $\epsilon \sim \varphi$ being the underlying stochastic variables. It follows from Proposition 2.4 that $\mathbb{E}[\psi_\xi] = \nabla_\xi U(\xi)$. Suppose at this moment that we are able to

generate i.i.d. random samples of ψ_ξ. We emphasize that random sampling of ψ_ξ is far from trivial but we shall show in the next section that this is indeed possible.

In stochastic gradient-based optimization methods, after setting an initial experimental setup $\xi_0 \in \mathcal{X}$, we recursively update itself as

$$\xi_{t+1} = g_t(\xi_t, \psi_{\xi_t}) \quad \text{for } t = 0, 1, \ldots,$$

or more generally,

$$\xi_{t+1} = g_t\left(\xi_t, \frac{1}{N}\sum_{n=1}^{N} \psi_{\xi_t}^{(n)}\right) \quad \text{for } t = 0, 1, \ldots,$$

where $\psi_{\xi_t}^{(1)}, \ldots, \psi_{\xi_t}^{(N)}$ are i.i.d. realizations of ψ_{ξ_t} for a sample size $N \in \mathbb{Z}_{>0}$. This means that, at each iteration, we only need (rough) unbiased Monte Carlo estimate of $\mathbb{E}[\psi_\xi]$ instead of the true value. There have been many examples for this recursion g_t proposed in the literature.

For instance, one of the most classical methods due to Robbins and Monro [30] is simply given by

$$\xi_{t+1} = \Pi_{\mathcal{X}}\left(\xi_t + a_t \cdot \frac{1}{N}\sum_{n=1}^{N} \psi_{\xi_t}^{(n)}\right),$$

with a sequence of non-negative reals called *learning rates* $a_0, a_1, \ldots$ such that

$$\sum_{t=0}^{\infty} a_t = \infty \quad \text{and} \quad \sum_{t=0}^{\infty} a_t^2 < \infty,$$

where $\Pi_{\mathcal{X}}$ denotes the projection operator which maps the input to a closest point in $\mathcal{X}$, i.e., $\Pi_{\mathcal{X}}(\xi') = \arg\min_{\xi \in \mathcal{X}} \|\xi - \xi'\|$ with $\|\cdot\|$ being the Euclidean norm of vector.[2] As described in [37, Chapter 5.9], for instance, if $\mathcal{X}$ is convex, U is strongly concave and differentiable with respect to ξ, and $\mathbb{E}\left[\|\psi_\xi\|_2^2\right] < \infty$ for any $\xi \in \mathcal{X}$, then the estimate ξ_t converges to the optimal ξ^* with the mean squared error of $O(1/t)$.

There have been many variants of the classical Robbins-Monro algorithm proposed in the literature, notably such as Polyak-Ruppert averaging [26, 31] and stochastic counterpart of Nesterov's acceleration [24]. More recently, the idea of using not only the first moment of the gradient estimate but also its second moment to set the learning rates for individual design parameters in ξ adaptively has been explored insensitively, especially in the machine learning community, see [8, 39, 19, 28].

3. Monte Carlo gradient estimation. Here we introduce two Monte Carlo estimators of the gradient $\nabla_\xi U(\xi) = \mathbb{E}[\psi_\xi]$. Subsequently we propose an algorithm to efficiently search for optimal Bayesian experimental designs.

[2]Note that most of the textbooks on stochastic algorithms such as [1, 37] consider minimization problems for which the update rule should be replaced by

$$\xi_{t+1} = \Pi_{\mathcal{X}}\left(\xi_t - a_t \cdot \frac{1}{N}\sum_{n=1}^{N} \psi_{\xi_t}^{(n)}\right),$$

and the objective function is often assumed to be convex instead of concave.

3.1. Standard Monte Carlo. The standard Monte Carlo method is one of the easiest and the most straightforward methods to approximate ψ_ξ. Let us estimate two expectations with respect to θ' by the Monte Carlo averages using common random samples of θ', respectively. Namely, for randomly chosen θ and ϵ, let

$$\psi_{\xi,M} := \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)} - \frac{\nabla \varrho_{\xi,M}(\theta,\epsilon)}{\varrho_{\xi,M}(\theta,\epsilon)},$$

with

$$\varrho_{\xi,M}(\theta,\epsilon) = \frac{1}{M}\sum_{m=1}^{M} \rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi),$$
$$\nabla \varrho_{\xi,M}(\theta,\epsilon) = \frac{1}{M}\sum_{m=1}^{M} \nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi),$$

where $\theta'^{(1)}, \ldots, \theta'^{(M)}$ are independent samples from the prior distribution π_0. More generally, for an importance distribution q which may depend on the value of $f_\xi(\theta,\epsilon)$ or the outer random variables θ and ϵ, we can consider

$$\psi_{\xi,M,q} := \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)} - \frac{\nabla \varrho_{\xi,M,q}(\theta,\epsilon)}{\varrho_{\xi,M,q}(\theta,\epsilon)}, \tag{3.1}$$

with

$$\varrho_{\xi,M,q}(\theta,\epsilon) = \frac{1}{M}\sum_{m=1}^{M} \frac{\rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})},$$
$$\nabla \varrho_{\xi,M,q}(\theta,\epsilon) = \frac{1}{M}\sum_{m=1}^{M} \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})},$$

where $\theta'^{(1)}, \ldots, \theta'^{(M)}$ are independent samples from the distribution q.

Although it holds from the linearity of expectation that

$$\mathbb{E}\left[\varrho_{\xi,M,q}(\theta,\epsilon) \mid \theta, \epsilon\right] = \mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta', \xi)\right],$$
$$\mathbb{E}\left[\nabla \varrho_{\xi,M,q}(\theta,\epsilon) \mid \theta, \epsilon\right] = \mathbb{E}_{\theta'}\left[\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta', \xi)\right],$$

for any M, i.e., both the denominator and the numerator themselves are estimated without any bias, respectively, taking the ratio between these two yields

$$\mathbb{E}[\psi_{\xi,M}], \mathbb{E}[\psi_{\xi,M,q}] \neq \mathbb{E}[\psi_\xi] = \nabla_\xi U(\xi)$$

unless $q = \pi^{f_\xi(\theta,\epsilon)|\xi}$. This means that neither $\psi_{\xi,M}$ nor $\psi_{\xi,M,q}$ is an unbiased estimator of the gradient $\nabla_\xi U(\xi)$.

3.2. Unbiased multilevel Monte Carlo. Here we introduce an unbiased multilevel Monte Carlo estimator by using the debiasing technique from [29] which itself is an extension of the multilevel Monte Carlo method due to Giles [12, 13]. Let us consider an increasing sequence $0 < M_0 < M_1 < \ldots$ such that $M_\ell \to \infty$ as $\ell \to \infty$. Then the strong law of large numbers ensures that

$$\mathbb{P}\left[\lim_{\ell\to\infty} \psi_{\xi,M_\ell,q} = \psi_\xi\right] = 1,$$

see for instance [25, Theorem 9.2], and so the following telescoping sum holds:

$$\nabla_\xi U(\xi) = \mathbb{E}[\psi_\xi] = \lim_{\ell\to\infty} \mathbb{E}[\psi_{\xi,M_\ell,q}] = \mathbb{E}[\psi_{\xi,M_0,q}] + \sum_{\ell=1}^{\infty} \mathbb{E}[\psi_{\xi,M_\ell,q} - \psi_{\xi,M_{\ell-1},q}].$$

More generally, suppose at this moment that we have a sequence of *correction* random variables $\Delta\psi_{\xi,0}, \Delta\psi_{\xi,1}, \dots$ such that $\mathbb{E}[\Delta\psi_{\xi,0}] = \mathbb{E}[\psi_{\xi,M_0,q}]$ and

$$\mathbb{E}[\Delta\psi_{\xi,\ell}] = \mathbb{E}[\psi_{\xi,M_\ell,q} - \psi_{\xi,M_{\ell-1},q}] \quad \text{for } \ell > 0.$$

Then it holds that

$$\nabla_\xi U(\xi) = \mathbb{E}[\psi_\xi] = \sum_{\ell=0}^{\infty} \mathbb{E}[\Delta\psi_{\xi,\ell}]. \tag{3.2}$$

For any sequence of positive reals $w_0, w_1, \dots$ such that $w_0 + w_1 + \cdots = 1$, the expectation of the random variable

$$\frac{\Delta\psi_{\xi,\ell}}{w_\ell}$$

with the index $\ell \geq 0$ being selected randomly with probability w_ℓ, is equal to the gradient $\nabla_\xi U(\xi)$. In fact, it is easy to see that

$$\mathbb{E}\left[\frac{\Delta\psi_{\xi,\ell}}{w_\ell}\right] = \sum_{\ell=0}^{\infty} \frac{\mathbb{E}[\Delta\psi_{\xi,\ell}]}{w_\ell} w_\ell = \sum_{\ell=0}^{\infty} \mathbb{E}[\Delta\psi_{\xi,\ell}] = \nabla_\xi U(\xi).$$

Therefore, for any number of outer samples $N \in \mathbb{Z}_{>0}$,

$$\frac{1}{N} \sum_{n=1}^{N} \frac{\Delta\psi_{\xi,\ell^{(n)}}}{w_{\ell^{(n)}}}$$

with $\ell^{(1)}, \dots, \ell^{(N)}$ being independent and randomly chosen with probability w_ℓ is an unbiased Monte Carlo estimator of $\nabla_\xi U(\xi)$.

Let C_ℓ denote the expected cost of computing $\Delta\psi_{\xi,\ell}$, which is proportional to M_ℓ. In order for the random variable $\Delta\psi_{\xi,\ell}/w_\ell$ to have finite expected squared ℓ_2-norm and finite expected computational cost, we must have

$$\sum_{\ell=0}^{\infty} \frac{\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2]}{w_\ell} < \infty \quad \text{and} \quad \sum_{\ell=0}^{\infty} C_\ell w_\ell < \infty. \tag{3.3}$$

Thus construction of such correction variables $\Delta\psi_{\xi,\ell}$ in conjunction with an associated sequence $w_0, w_1, \dots$, which has not been discussed yet, becomes a central issue.

3.2.1. Naive construction. Throughout this paper let us consider a geometric progression $M_\ell = M_0 2^\ell$ for some $M_0 \in \mathbb{Z}_{\geq 0}$. Although it is possible to change the base of the progression to a general integer $b \geq 2$, we restrict ourselves to the case $b = 2$ for simplicity of exposition.

Probably the most straightforward form of the correction variables $\Delta\psi_{\xi,0}, \Delta\psi_{\xi,1}, \dots$ is $\Delta\psi_{\xi,0} = \psi_{\xi,M_0,q}$ and

$$\Delta\psi_{\xi,\ell} = \psi_{\xi,M_0 2^\ell,q} - \psi_{\xi,M_0 2^{\ell-1},q},$$

for $\ell > 0$, where both $\psi_{\xi,M_0 2^{\ell-1},q}$ and $\psi_{\xi,M_0 2^{\ell},q}$ are given as in (3.1) with $M = M_0 2^{\ell-1}$ and $M = M_0 2^{\ell}$, respectively. Here, instead of using mutually independent $M_0 2^{\ell-1}$ and $M_0 2^{\ell}$ samples on θ' to compute $\psi_{\xi,M_0 2^{\ell-1},q}$ and $\psi_{\xi,M_0 2^{\ell},q}$, respectively, a subset with size $M_0 2^{\ell-1}$ of the $M_0 2^{\ell}$ samples on θ' used to compute $\psi_{\xi,M_0 2^{\ell},q}$, can be reused to compute $\psi_{\xi,M_0 2^{\ell-1},q}$ by the linearity of expectation. By doing so, it is expected that $\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2]$ is much smaller in magnitude than $\mathbb{E}[\|\psi_{\xi,M_0 2^{\ell},q}\|_2^2]$ (or $\mathbb{E}[\|\psi_{\xi,M_0 2^{\ell-1},q}\|_2^2]$).

However, it seems not possible that the order of $\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2]$ is better than $O(2^{-\ell})$. Recalling that $C_\ell \propto M_\ell \propto 2^\ell$, a faster decay of $\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2]$ is required to find a sequence of positive reals $w_0, w_1, \ldots$ which satisfies the condition (3.3). We conjecture that a lower bound on $\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2]$ of order $2^{-\ell}$ exists for this naive construction.

3.2.2. Antithetic construction. Motivated by the MLMC literature [15, 4, 14, 16, 17], we address this issue by considering the following *antithetic coupling* in this paper. A key ingredient here is that we can take two disjoint subsets with equal size $M_0 2^{\ell-1}$ of the $M_0 2^{\ell}$ samples on θ' used to compute $\psi_{\xi,M_0 2^{\ell},q}$, which results in two independent realizations of $\psi_{\xi,M_0 2^{\ell-1},q}$, denoted by $\psi^{(a)}_{\xi,M_0 2^{\ell-1},q}$ and $\psi^{(b)}_{\xi,M_0 2^{\ell-1},q}$, respectively. To be more precise, for the independent samples $\theta'^{(1)}, \ldots, \theta'^{(M_0 2^{\ell})}$ generated from the distribution q, we write

$$\begin{aligned}
\psi_{\xi,M_0 2^{\ell},q} &= \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)} - \frac{\nabla \varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon)}{\varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon)}, \\
\psi^{(a)}_{\xi,M_0 2^{\ell-1},q} &= \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)} - \frac{\nabla \varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}, \quad \text{and} \\
\psi^{(b)}_{\xi,M_0 2^{\ell-1},q} &= \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)}{\rho(f_\xi(\theta,\epsilon) \mid \theta, \xi)} - \frac{\nabla \varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)},
\end{aligned}$$

where, for the second term of each, we have defined

$$\begin{aligned}
\varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon) &= \frac{1}{M_0 2^{\ell}} \sum_{m=1}^{M_0 2^{\ell}} \frac{\rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})}, \\
\nabla \varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon) &= \frac{1}{M_0 2^{\ell}} \sum_{m=1}^{M_0 2^{\ell}} \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})}, \\
\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) &= \frac{1}{M_0 2^{\ell-1}} \sum_{m=1}^{M_0 2^{\ell-1}} \frac{\rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})}, \\
\nabla \varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) &= \frac{1}{M_0 2^{\ell-1}} \sum_{m=1}^{M_0 2^{\ell-1}} \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})}, \\
\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) &= \frac{1}{M_0 2^{\ell-1}} \sum_{m=M_0 2^{\ell-1}+1}^{M_0 2^{\ell}} \frac{\rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})}, \\
\nabla \varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) &= \frac{1}{M_0 2^{\ell-1}} \sum_{m=M_0 2^{\ell-1}+1}^{M_0 2^{\ell}} \frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta'^{(m)}, \xi)\pi_0(\theta'^{(m)})}{q(\theta'^{(m)})},
\end{aligned}$$

respectively.

Now a sequence of the correction random variables $\Delta\psi_{\xi,0}, \Delta\psi_{\xi,1}, \ldots$ is defined by $\Delta\psi_{\xi,0} = \psi_{\xi,M_0,q}$ and

$$\begin{aligned}(3.4)\qquad \Delta\psi_{\xi,\ell} &= \psi_{\xi,M_0 2^\ell,q} - \frac{\psi^{(a)}_{\xi,M_0 2^{\ell-1},q} + \psi^{(b)}_{\xi,M_0 2^{\ell-1},q}}{2} \\ &= \frac{1}{2}\left(\frac{\nabla\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)} + \frac{\nabla\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}\right) - \frac{\nabla\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)},\end{aligned}$$

for $\ell > 0$. The difference between this antithetic construction and the naive construction is that $\psi_{\xi,M_0 2^{\ell-1},q}$ has been replaced by the mean of $\psi^{(a)}_{\xi,M_0 2^{\ell-1},q}$ and $\psi^{(b)}_{\xi,M_0 2^{\ell-1},q}$. This means that each of the $M_0 2^\ell$ samples is used exactly twice in antithetic construction: once in $\psi_{\xi,M_0 2^\ell,q}$ and once in either $\psi^{(a)}_{\xi,M_0 2^{\ell-1},q}$ or $\psi^{(b)}_{\xi,M_0 2^{\ell-1},q}$. For this novel version of $\Delta\psi_{\xi,\ell}$, the linearity of expectation ensures

$$\begin{aligned}\mathbb{E}\left[\Delta\psi_{\xi,\ell}\right] &= \mathbb{E}\left[\psi_{\xi,M_0 2^\ell,q}\right] - \frac{1}{2}\left(\mathbb{E}\left[\psi^{(a)}_{\xi,M_0 2^{\ell-1},q}\right] + \mathbb{E}\left[\psi^{(b)}_{\xi,M_0 2^{\ell-1},q}\right]\right) \\ &= \mathbb{E}\left[\psi_{\xi,M_0 2^\ell,q}\right] - \frac{1}{2}\left(\mathbb{E}\left[\psi_{\xi,M_0 2^{\ell-1},q}\right] + \mathbb{E}\left[\psi_{\xi,M_0 2^{\ell-1},q}\right]\right) \\ &= \mathbb{E}[\psi_{\xi,M_0 2^\ell,q} - \psi_{\xi,M_0 2^{\ell-1},q}],\end{aligned}$$

so that it fits with the telescoping sum representation (3.2) of the gradient $\nabla_\xi U(\xi)$. Despite the distinction being subtle, we will show that the antithetic construction has better properties than the naive construction. Hereafter, $\Delta\psi_{\xi,\ell}$ refers to the antithetic construction given in (3.4).

It is clear that the cost C_ℓ to compute $\Delta\psi_{\xi,\ell}$ is proportional to 2^ℓ, and also that the following *antithetic* properties hold for $\Delta\psi_{\xi,\ell}$:

$$\begin{aligned}(3.5)\qquad \varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon) &= \frac{1}{2}\left(\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) + \varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)\right), \quad \text{and} \\ \nabla\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon) &= \frac{1}{2}\left(\nabla\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) + \nabla\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)\right),\end{aligned}$$

which play a crucial role in showing that this antithetic construction achieves a faster decay rate of $\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2]$ than the naive construction, making it possible to find a sequence of positive reals $w_0, w_1, \ldots$ which satisfies the condition (3.3). The following claim is the main theoretical result of this paper.

THEOREM 3.1. *Assume that*

$$\sup_{\theta,\theta',\epsilon} \left\|\nabla_\xi \log \rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\right\|_\infty < \infty,$$

and that there exists $u > 2$ such that

$$\mathbb{E}_{\theta\sim\pi_0,\theta'\sim q,\epsilon}\left[\left|\frac{\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\pi_0(\theta')}{\rho(f_\xi(\theta,\epsilon) \mid \xi)q(\theta')}\right|^u\right] < \infty.$$

Then the following holds true:

1. *For a fixed ℓ, we have*

$$\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2] = O(2^{-\beta\ell}) \quad \textit{with} \quad \beta = \frac{\min(u,4)}{2}.$$

2. *In order to have* (3.3)*, it suffices to choose* $w_\ell \propto 2^{-\tau\ell}$ *with* $1 < \tau < \beta$.

We postpone the proof of the theorem to Appendix A.

Remark 3.2. It follows from the first item of Theorem 3.1 that

$$\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2] = O(2^{-\ell}),$$

for a fixed ℓ. Using this property, the bias of the standard Monte Carlo estimator $\psi_{\xi,M_0 2^L,q}$ with $M = M_0 2^L$ inner samples is bounded as

$$\begin{aligned}\left\|\nabla_\xi U(\xi) - \mathbb{E}[\psi_{\xi,M_0 2^L,q}]\right\|_2 &= \left\|\sum_{\ell=L+1}^{\infty} \mathbb{E}[\Delta\psi_{\xi,\ell}]\right\|_2 \leq \sum_{\ell=L+1}^{\infty} \mathbb{E}\left[\|\Delta\psi_{\xi,\ell}\|_2\right] \\ &= O(2^{-L}) = O(M^{-1}).\end{aligned}$$

This means that, for small M, the standard Monte Carlo estimator may lead to a wrong trajectory of an experimental design in stochastic gradient-biased optimization and the resulting design will not be close to optimal.

3.3. Unbiased MLMC stochastic optimization. Finally we arrive at our proposal of a stochastic algorithm to search for an optimal Bayesian experimental design $\xi^* \in \mathcal{X}$ as summarized in Algorithm 3.1. Here we note that Algorithm 3.1 assumes that the conditions appearing in Theorem 3.1 hold for any $\xi \in \mathcal{X}$ with a common value of u. Given additional assumptions that the domain $\mathcal{X}$ is convex and that U is strongly concave and differentiable with respect to ξ, most of the stochastic gradient-based optimization algorithms have a theoretical guarantee that ξ_t converges to the optimal $\xi^* \in \mathcal{X}$ with some decay rate, typically with the mean square error of $O(1/t)$ as mentioned in Section 2.2.

Algorithm 3.1 Unbiased MLMC stochastic optimization

For a given $1 < \tau < \beta$, set $w_0, w_1, \ldots > 0$ such that $w_0 + w_1 + \cdots = 1$ and $w_\ell \propto 2^{-\tau\ell}$. For the feasible set $\mathcal{X}$, initialize $\xi_0 \in \mathcal{X}$ and $t = 0$. For $N \in \mathbb{Z}_{>0}$, do the following:

1. Choose $\ell^{(1)}, \ldots, \ell^{(N)} \in \mathbb{Z}_{\geq 0}$ independently and randomly with probability w_ℓ.
2. Compute an unbiased MLMC estimate of the gradient $\nabla_\xi U$ at $\xi = \xi_t$:

$$\frac{1}{N}\sum_{n=1}^{N} \frac{\Delta\psi_{\xi_t,\ell^{(n)}}}{w_{\ell^{(n)}}}.$$

3. Apply a stochastic gradient-based algorithm to get ξ_{t+1}:

$$\xi_{t+1} = g_t\left(\xi_t, \frac{1}{N}\sum_{n=1}^{N} \frac{\Delta\psi_{\xi_t,\ell^{(n)}}}{w_{\ell^{(n)}}}\right).$$

4. Check whether a certain stopping criterion is satisfied. If yes, stop the iteration. Otherwise, go to Step 1 with $t \leftarrow t + 1$.

As in [2, 16, 5], using Laplace approximation-based importance distribution for q helps not only reduce the expected squared ℓ_2-norm of the Monte Carlo gradient estimator but also avoid numerical instability coming from concentrated posterior measures of θ' given $f_\xi(\theta, \epsilon)$. We also refer to [35] for some theoretical analyses on the Laplace approximation.

4. Numerical experiments. Here, we conduct numerical experiments on two example problems in Bayesian experimental design. The first example is aimed at verifying our proposed algorithm by using a simple test problem. Then, in order to see practical performance of our algorithm, we consider a PK model used in [33] for our second example. The Python code used in our experiments is available from https://github.com/Goda-Research-Group/MLMC_stochastic_gradient.

4.1. Simple test case. Let $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2_{>0}$ with $\theta_1, \theta_2 \overset{\text{iid}}{\sim} \text{lognormal}(\mu, \sigma_0^2)$. For an experimental design $\xi \in \mathbb{R}_{>0}$, let an observation $Y = (Y_1, Y_2) \in \mathbb{R}^2_{>0}$ follow

$$\begin{aligned} Y_1 \mid \theta, \xi &\sim \text{lognormal}(g(\xi) \log \theta_1, \sigma_\epsilon^2), \\ Y_2 \mid \theta, \xi &\sim \text{lognormal}(h(\xi) \log \theta_2, \sigma_\epsilon^2), \end{aligned}$$

for some functions g and h. This is equivalent to consider the following forward model:

$$\begin{aligned} Y_1 &= e^{g(\xi) \log \theta_1 + \sigma_\epsilon \epsilon_1}, \\ Y_2 &= e^{h(\xi) \log \theta_2 + \sigma_\epsilon \epsilon_2}, \end{aligned}$$

for $\epsilon_1, \epsilon_2 \overset{\text{iid}}{\sim} N(0,1)$ independently of θ and ξ, which is obviously a special case of (1.1). The expected information gain for a given ξ is analytically calculated as

$$U(\xi) = \frac{1}{2} \log \left((g(\xi))^2 \frac{\sigma_0^2}{\sigma_\epsilon^2} + 1 \right) \left((h(\xi))^2 \frac{\sigma_0^2}{\sigma_\epsilon^2} + 1 \right).$$

Also, applying Jensen's inequality to (2.1), we see that $U(\xi)$ is bounded above by

$$\begin{aligned} U(\xi) \le \tilde{U}(\xi) &:= \mathbb{E}_{\theta,\epsilon} \left[\log \rho(f_\xi(\theta, \epsilon) \mid \theta, \xi) \right] - \mathbb{E}_{\theta,\theta',\epsilon} \left[\log \rho(f_\xi(\theta, \epsilon) \mid \theta', \xi) \right] \\ &= (g(\xi))^2 \frac{\sigma_0^2}{\sigma_\epsilon^2} + (h(\xi))^2 \frac{\sigma_0^2}{\sigma_\epsilon^2}. \end{aligned}$$

Here we note that the standard Monte Carlo gradient estimator with $M = 1$ inner sample from the prior distribution, i.e., $\psi_{\xi,1}$, is nothing but an unbiased estimator of $\nabla_\xi \tilde{U}(\xi)$. Therefore, as long as

$$\xi^* = \arg \max_{\xi \in \mathbb{R}_{>0}} U(\xi) \neq \arg \max_{\xi \in \mathbb{R}_{>0}} \tilde{U}(\xi)$$

holds, stochastic gradient-based optimization based on $\psi_{\xi,1}$ will not converge to the optimal design ξ^*. In our experiments below, let $\mu = 0$, $\sigma_0 = \sigma_\epsilon = 1$,

$$g(\xi) = e^{-\xi^2/2} \quad \text{and} \quad h(\xi) = \sqrt{\frac{3}{2} \left(1 - e^{-\xi^2}\right)}.$$

Fig. 1 compares U and $\tilde{U}$ as functions of ξ for this setting. The optimal design which maximizes U is given by $\xi^* = \sqrt{\log 3} \approx 1.048\ldots$ and we can observe that U is concave around ξ^*. On the other hand, its upper bound $\tilde{U}$ is a strictly monotone increasing function and its supremum attains for $\xi \to \infty$.

Throughout this subsection, we do not use any importance sampling for the unbiased MLMC estimator of $\nabla_\xi U$ and set M_0, the number of level 0 inner samples, to 1. The left panel of Fig. 2 shows the convergence behavior of the MLMC correction variables $\Delta\psi_{\xi,\ell}$ at $\xi = 1.5$. Here the mean squares (expected squared ℓ_2-norms) of

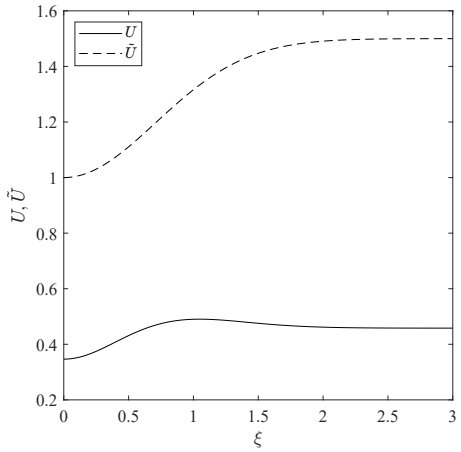


FIG. 1. *The expected information gain U and its upper bound $\tilde{U}$ for the test case.*

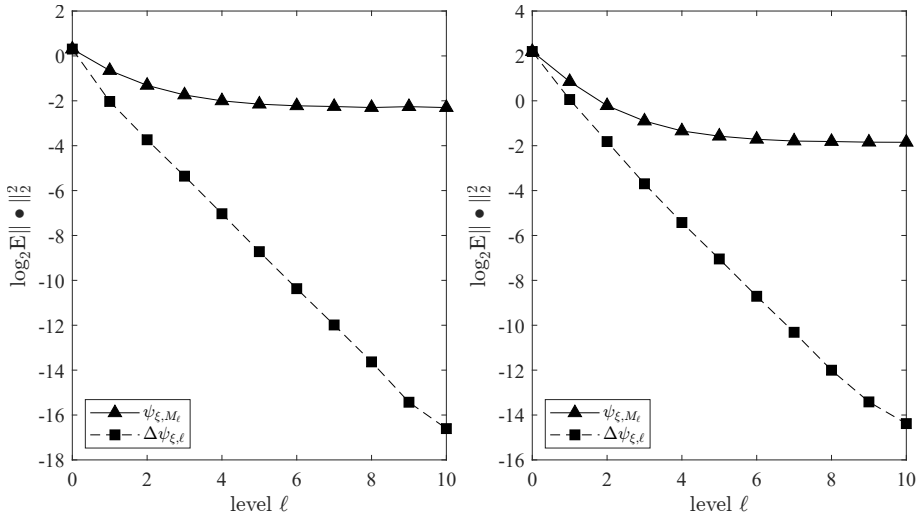


FIG. 2. *The mean squares of the variables ψ_{ξ,M_ℓ} and $\Delta\psi_{\xi,\ell}$ for the test case at $\xi = 1.5$ (left) and at $\xi = \xi^* = \sqrt{\log 3}$ (right).*

ψ_{ξ,M_ℓ} and $\Delta\psi_{\xi,\ell}$ are plotted on a $\log_2$ scale as functions of the level ℓ, where the means are estimated empirically by using 10^5 i.i.d. samples at each level. While the mean square of ψ_{ξ,M_ℓ} takes an almost constant value for $\ell > 4$, that of $\Delta\psi_{\xi,\ell}$ decreases geometrically as the level increases. The linear regression of the data for the range $1 \le \ell \le 10$ provides an estimation of β as 1.64, which agrees well with the theoretical result in Theorem 3.1. As shown in the right panel of Fig. 2, a similar convergence behavior of the MLMC correction variables $\Delta\psi_{\xi,\ell}$ can be observed at the optimal design $\xi = \xi^* = \sqrt{\log 3}$, where β is estimated as 1.63.

Such a fast geometric decay of the correction variables $\Delta\psi_{\xi,\ell}$ justifies us to apply Algorithm 3.1 to search for the optimal design ξ^*. In order to randomly choose the level ℓ, we set $\tau = 1.5$ and $w_\ell = 2^{-3\ell/2}(1 - 2^{-3/2})$. This implies that the expected number of inner samples used in the MLMC estimator is given by

$$\sum_{\ell=0}^{\infty} 2^\ell w_\ell = (1 - 2^{-3/2}) \sum_{\ell=0}^{\infty} 2^{-\ell/2} = \frac{1 - 2^{-3/2}}{1 - 2^{-1/2}} \approx 2.21.$$

For comparison, we also consider the standard Monte Carlo estimators $\psi_{\xi,M}$ for the

gradient of the expected information gain with various values of $M = 1, 2, 4, \ldots, 64$ within stochastic gradient descent. We fix the number of outer samples N to 2000 throughout all the iteration steps for all the estimators. We use the Robbins-Monro algorithm with Polyak-Ruppert averaging and the learning rates $\alpha_t = 5/(t+1)$ as a stochastic descent algorithm, and as the computational cost is proportional to the number of inner samples, we set the maximum iteration steps T to $\lfloor 10^7/M \rfloor$, the largest integer less than or equal to $10^7/M$. Although the number of inner samples is a random variable for the MLMC estimator, we simply set T to $\lfloor 10^7/2.21 \rfloor$. The initial design candidate at $t = 0$ is given by $\xi_0 = 1.5$ and the feasible set $\mathcal{X}$ is set to $\mathbb{R}_{>0}$. Hence the maximum increment of the expected information gain is $U(\sqrt{\log 3}) - U(1.5) \approx 0.0148$. For each gradient estimator, we conduct 10 independent runs and compute the average of the distance $\|\xi_t - \xi^*\|_2^2$ and its standard error for all the iteration steps, which correspond to the line and the shaded area of Fig. 3, respectively.

Fig. 3 shows the convergence behaviors of the estimated experimental design ξ_t for the considered estimators of the gradient $\nabla_\xi U$. Note here that the horizontal axis is given by $M \times t$ as a measure of the total computational cost (here again, we simply let $M = 2.21$ for the MLMC estimator) and both axes use the logarithmic scales. As expected, the standard Monte Carlo estimator with $M = 1$ leads to larger values of ξ_t which make $\tilde{U}$ large, so that the search goes in wrong direction. Even for $M = 2$, the situation is not improved so much and the experimental design ξ_t remains almost the same throughout the iterations. For larger values of M, the standard Monte Carlo estimator works in the early stages, making the distance $\|\xi_t - \xi^*\|_2^2$ small. However, after some iteration steps, the estimate ξ_t converges to some point away from the optimal ξ^*. Although it is natural that such bias can be reduced simply by increasing M, a proper choice of M in practical applications is far from trivial since larger M means a larger computational cost and the bias seems extremely hard to estimate in advance. This is exactly the point where the unbiased MLMC estimator can help. As the black line shows, the distance $\|\xi_t - \xi^*\|_2^2$ decreases consistently from the early stage and overtakes the standard Monte Carlo estimators with fixed M, leading to a better estimate of the optimal experimental design. The linear regression of the data for the whole range $0 < \log_{10}(Mt) \leq 7$ shows that the estimate ξ_t converges to ξ^* with the mean squared error of order $t^{-1.12}$ approximately, which is almost consistent with the standard stochastic optimization theory [37, Chapter 5.9]. A slightly faster decay of the standard Monte Carlo estimators with $M \geq 8$ in the early stages could be because that they estimate the gradients of biased objective functions which are steeper than the gradient of $U(\xi)$ around the initial estimate $\xi_0 = 1.5$ in this case.

4.2. Pharmacokinetic model. Let us consider a PK design problem introduced in [33]. Suppose that a drug with a fixed dose $D = 400$ is administrated to subjects at time $\mathcal{T} = 0$. In order to reduce the uncertainty about a set of PK parameters, which affect the absorption, distribution and the elimination of the drug in the subjects' body, it would be helpful to take blood samples of the subjects at several different times and to measure the concentration of the drug in the samples. Blood samples are assumed to be taken 15 times at $\mathcal{T} = \xi^{(1)}, \ldots, \xi^{(15)}$ hours after the drug administration. Given the set of 15 drug concentration measurements, it is expected that the uncertainty of PK parameters of interest θ can be reduced. Our objective here is to optimize sampling times $\xi = (\xi^{(1)}, \ldots, \xi^{(15)}) \in \mathbb{R}^{15}_{\geq 0}$ such that the expected information gain brought from blood sampling is maximized.

Let $\theta = (\log k_a, \log k_e, \log V) \in \mathbb{R}^3$ where k_a represents the first-order absorption rate constant, k_e does the first-order elimination rate constant and V does the volume

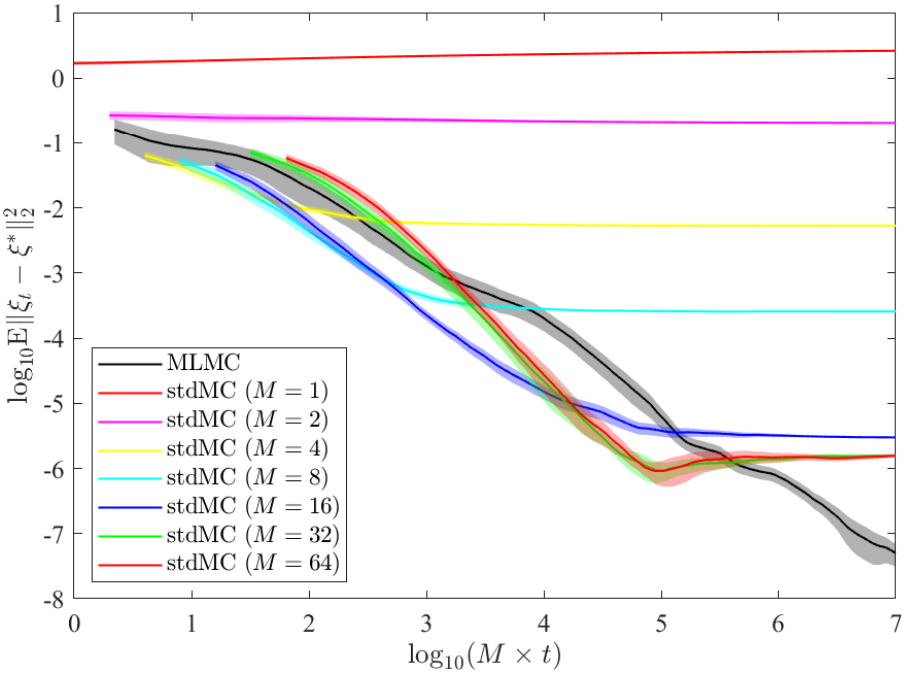


FIG. 3. *The convergence of the estimated experimental design ξ_t to the optimal ξ^* for various Monte Carlo estimators of the gradient $\nabla_\xi U$. For each estimator, the line and the shaded area represent the average and its standard error estimated from 10 independent runs, respectively.*

of distribution. Following [33], assume that the drug concentration of blood sample taken at time $\mathcal{T} \geq 0$ is described as

$$Y_{\mathcal{T}} = \frac{Dk_a}{V(k_a - k_e)} \left(e^{-k_e \mathcal{T}} - e^{-k_a \mathcal{T}}\right)(1 + \epsilon_1) + \epsilon_2 =: g_{\mathcal{T}}(\theta, \epsilon),$$

with $\epsilon = (\epsilon_1, \epsilon_2)$, where ϵ_1 and ϵ_2 represent the multiplicative and additive Gaussian noises, respectively. Then our forward model is given by

$$Y = (Y_{\xi^{(1)}}, \ldots, Y_{\xi^{(15)}}) = \left(g_{\xi^{(1)}}(\theta, \epsilon_{\xi^{(1)}}), \ldots, g_{\xi^{(15)}}(\theta, \epsilon_{\xi^{(15)}})\right) \in \mathbb{R}^{15},$$

where $\epsilon_{\xi^{(1)}}, \ldots, \epsilon_{\xi^{(15)}}$ are assumed mutually independent and follow the same bi-variate normal distribution

$$\epsilon_{\xi^{(j)}} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.01 & 0 \\ 0 & 0.1 \end{pmatrix}\right).$$

The input random variables in θ are assumed independent and the corresponding probability distributions are given by $\log k_a \sim N(0, 0.05), \log k_e \sim N(\log(0.1), 0.05)$ and $\log V \sim N(\log(20), 0.05)$, respectively. This means that the prior information entropy of θ is equal to $3\log(\sqrt{2\pi e \times 0.05}) \approx -0.2368$. Moreover, the likelihood function is given by the product of $\rho(g_{\xi^{(j)}}(\theta, \epsilon_{\xi^{(j)}}) \mid \theta', \xi^{(j)})$ that can be computed explicitly by following Example 2.3.

In this setting the posterior distribution of θ given Y cannot be computed analytically. In order to reduce the expected squared ℓ_2-norm of the unbiased MLMC estimator of the gradient $\nabla_\xi U$, we use Laplace approximation-based importance sampling. Since not only the additive noise but also the multiplicative noise are included in the forward model, we consider a simple modification of the original method in [22]

as follows. Let us write

$$\overline{g_{\mathcal{T}}}(\theta) = \frac{Dk_a}{V(k_a - k_e)} \left(e^{-k_e \mathcal{T}} - e^{-k_a \mathcal{T}} \right) \quad \text{and} \quad \overline{g_\xi}(\theta) = \left(\overline{g_{\xi^{(1)}}}(\theta), \ldots, \overline{g_{\xi^{(15)}}}(\theta) \right).$$

Then, for the data Y generated conditionally on the known value of $\theta = \theta^*$ from the forward model, we approximate the posterior distribution $\pi^{Y|\xi}(\theta)$ by a Gaussian distribution $N(\hat{\theta}, \hat{\Sigma})$ with

$$\hat{\theta} = \theta^* - \left(J(\theta^*)^\top \Sigma_\epsilon^{-1} J(\theta^*) + H(\theta^*)^\top \Sigma_\epsilon^{-1} E - \nabla_\theta \nabla_\theta \log \pi_0(\theta^*) \right)^{-1} J(\theta^*)^\top \Sigma_\epsilon^{-1} E,$$
$$\hat{\Sigma} = \left(J(\hat{\theta})^\top \Sigma_\epsilon^{-1} J(\hat{\theta}) - \nabla_\theta \nabla_\theta \log \pi_0(\hat{\theta}) \right)^{-1}.$$

Here J and H denote the Jacobian and Hessian of $-\overline{g_\xi}$, respectively, that is, $J(\theta) = -\nabla_\theta \overline{g_\xi}(\theta)$ and $H(\theta) = -\nabla_\theta \nabla_\theta \overline{g_\xi}(\theta)$. Also we write $E := Y^\top - \overline{g_\xi}(\theta^*)^\top$ and

$$\Sigma_\epsilon = \operatorname{diag}\left(0.01 \left(\overline{g_{\xi^{(1)}}}(\theta) \right)^2 + 0.1, \ldots, 0.01 \left(\overline{g_{\xi^{(15)}}}(\theta) \right)^2 + 0.1 \right).$$

We use this $N(\hat{\theta}, \hat{\Sigma})$ as an importance distribution q. The only difference from the one in [22] is that the matrix Σ_ϵ depends on the mean response $\overline{g_\xi}(\theta)$ due to the multiplicative noise in our setting. Although a first-order approximation argument similar to [22] might be possible and lead to different forms of $\hat{\theta}$ and $\hat{\Sigma}$, such a detailed analysis on the Laplace approximation is beyond the scope of this paper.

In order to search for optimal design parameters $\xi = (\xi^{(1)}, \ldots, \xi^{(15)})$, we do not represent them by a smaller number of parameters as considered in [33], but instead we optimize them directly. We set a design at the initial iteration step $t = 0$ to equi-spaced times $\xi_0 = (1, 2, \ldots, 15)$. In Algorithm 3.1, we fix $M_0 = 1$, set $w_0 = 0.9$ and $w_\ell \propto 2^{-3\ell/2}$ for $\ell \geq 1$ such that they are summed up to 1, and set the number of outer samples to $N = 2000$ at each iteration step. This implies that the expected number of inner samples used in the MLMC estimator is given by

$$\sum_{\ell=0}^{\infty} 2^\ell w_\ell = \frac{9}{10} + \frac{2^{3/2} - 1}{10} \sum_{\ell=1}^{\infty} 2^{-\ell/2} \approx 1.34.$$

We use the AMSGrad optimizer with constant learning rate $\alpha_t = 0.004$ and exponential moving average parameters $\beta_1 = 0.9, \beta_2 = 0.999$ as a stochastic descent algorithm, and set the maximum iteration steps T to 10000 as a stopping criterion. The feasible domain $\mathcal{X}$ is restricted to $[0, 24]^{15}$. For comparison, we also consider the standard (biased) Monte Carlo estimator for the gradient $\nabla_\xi U$ with a fixed number of inner samples $M = 1$ and the Laplace approximation-based importance sampling within stochastic gradient descent. As expected from the numerical results shown in [5], the Laplace approximation-based importance sampling helps reduce the bias of the Monte Carlo estimator significantly even for $M = 1$.

Fig. 4 shows the set of design parameters $\xi = (\xi^{(1)}, \ldots, \xi^{(15)})$ obtained at the iteration steps $t = 0, 100, 500, 1000, 5000, 10000$ for a single run. The overall convergence behaviors both for the standard Monte Carlo estimator and the MLMC estimator look quite similar to each other. That is, the allocations of 15 sampling times become irregular at the earlier steps compared to the initially equi-spaced design, but then some of sampling times gradually get quite close to each other, ending up with three well-separated clusters. It is interesting to see that stochastic gradient-based optimization

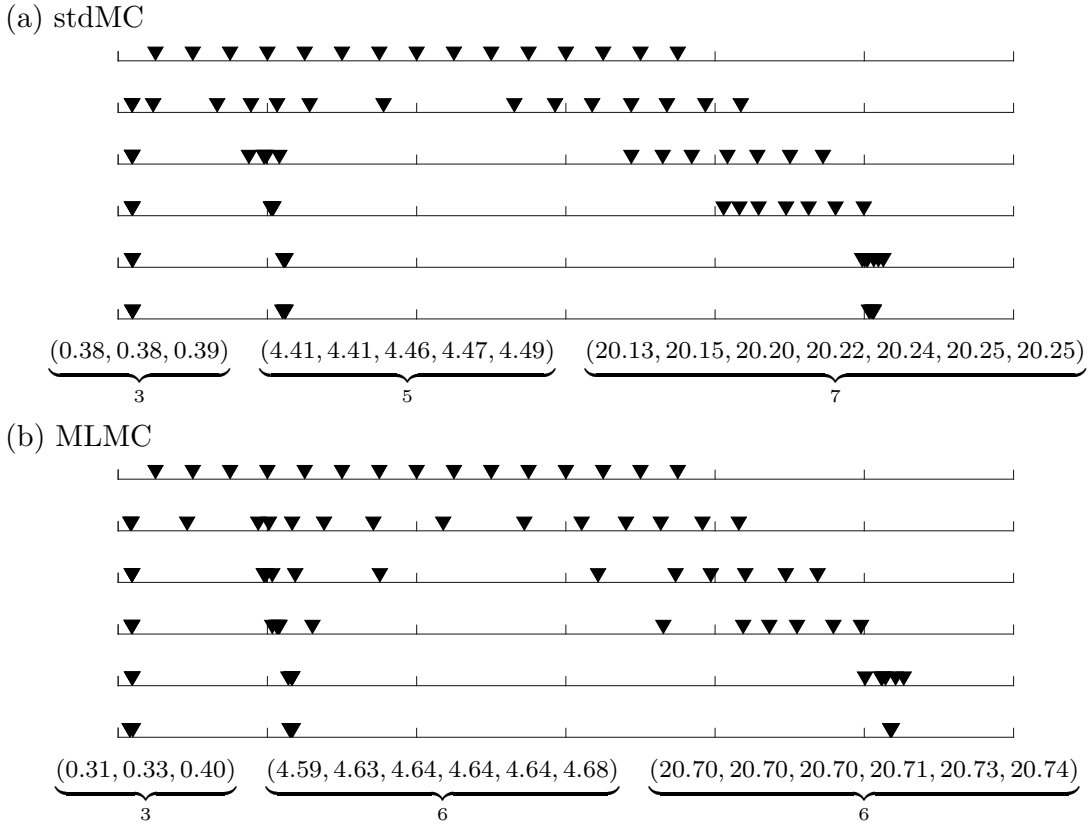


FIG. 4. *Design parameters* $(\xi_1, \ldots, \xi_{15})$ *within the interval* $[0, 24]$ *during the optimization process for a single run at the iteration steps* $t = 0, 100, 500, 1000, 5000, 10000$ *(in descending order): (a) the result for stdMC and (b) the result for MLMC. The resulting design is shown in detail respectively at the bottom.*

naturally finds such so-called *replicate* design that is often considered in the PK applications [32, 33]. Looking into the details, there is a difference between the resulting designs obtained by the standard Monte Carlo estimator and the MLMC estimator. For the standard Monte Carlo estimator, the number of sampling times allocated to each cluster is $3, 5, 7$ (from earlier one to later one), respectively, whereas the corresponding number is $3, 6, 6$, respectively, for the MLMC estimator. These allocations of sampling times are consistent among 10 independent runs for both the estimators. The average sampling time (with its standard deviation) within each cluster, estimated from 10 independent runs, is $0.385\ (0.003), 4.442\ (0.008), 20.202\ (0.006)$ for the standard Monte Carlo estimator, and is $0.367\ (0.010), 4.652\ (0.018), 20.699\ (0.017)$ for the MLMC estimator. The two-sample Wilcoxon test yields the p-value about 10^{-5} for all of the three clusters, which supports that the differences between the centers of the clusters obtained by the two estimators are statistically significant.

Fig. 5 shows the convergence behaviors of the MLMC correction variables $\Delta\psi_{\xi,\ell}$ at the iteration steps $t = 0, T/2, T$ for a single run. Similarly to Fig. 2. the mean squares (expected squared ℓ_2-norms) of ψ_{ξ,M_ℓ} and $\Delta\psi_{\xi,\ell}$ are plotted on a $\log_2$ scale as functions of the level ℓ, where the means are estimated empirically by using 10^5 i.i.d. samples at each level. While the mean square of ψ_{ξ,M_ℓ} takes an almost constant value for $\ell > 4$, that of $\Delta\psi_{\xi,\ell}$ decreases geometrically as the level increases. The linear regression of the data for the range $1 \leq \ell \leq 10$ provides estimations of β as $0.80, 1.36, 1.47$, respectively. The result on the case $\beta \leq 1$ is not covered by Theorem 3.1, and in such

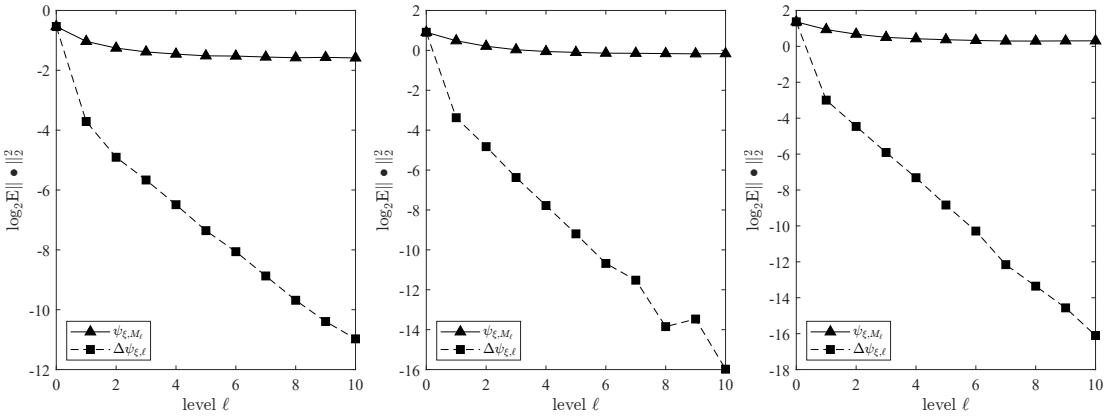


FIG. 5. *The mean squares of the variables ψ_{ξ,M_ℓ} and $\Delta\psi_{\xi,\ell}$ for the PK model at the iteration steps $t = 0, T/2, T$*

a case, we do not have a right choice of w_ℓ which leads to both finite expected cost and finite expected squared ℓ_2-norm. Further theoretical investigation is needed to address this issue. On the other hand, the result $\beta > 1$ for the steps $t = T/2, T$ is as expected from our theoretical result. Nonetheless, our choice $w_\ell \propto 2^{-3\ell/2}$ might be a bit aggressive in the sense that the expected squared ℓ_2-norm of the MLMC estimator possibly does not converge, although we see no evidence of this in our experiments. A practical issue on how to choose w_ℓ properly depending on the problem at hand is also left open for future research.

Finally, Fig. 6 shows the behaviors of the expected information gain U as a function of the number of iteration steps. For this problem, the expected information gain for any design parameter ξ cannot be evaluated exactly, so that we use a randomized variant of the MLMC estimator introduced in [16] with 10^6 outer samples to estimate the expected information gain for every 500 steps. As 10 independent runs are performed, we plot the average of 10 estimated values in mark, while the shaded area represents the linearly interpolated standard error. We can see that the expected information gain increases with some fluctuation as the iteration proceeds, and converges to a constant value. The average converged value for the MLMC estimator is 4.544, which is slightly larger than 4.535 obtained for the standard Monte Carlo estimator. Note that the expected information gain for the initial design is estimated as 3.774, which is well below the maximum values obtained both for the standard Monte Carlo estimator and the MLMC estimator. Just to provide an intuition of this improvement, assume that each individual variable in θ remains independent and follows a normal distribution with an equal variance even after observing Y, which is usually not true. Then it is inferred that the variance of each variable after observing Y with the initial design is reduced on average by the factor $(\exp(3.774/3))^2 \approx 12.379$, whereas that with the resulting design by our proposed optimization algorithm is $(\exp(4.554/3))^2 \approx 20.822$. Although the increment of the maximum expected information gain by using the MLMC estimator seems marginal as compared to the the standard Monte Carlo estimator in this example, it is important to emphasize again that the resulting experimental designs are qualitatively different.

5. Conclusion. In this paper we have developed an efficient stochastic algorithm to optimize Bayesian experimental designs such that the expected information gain is maximized. Since the gradient of the expected information gain with respect to design parameters is expressed as a nested expectation, a straightforward use of stochastic

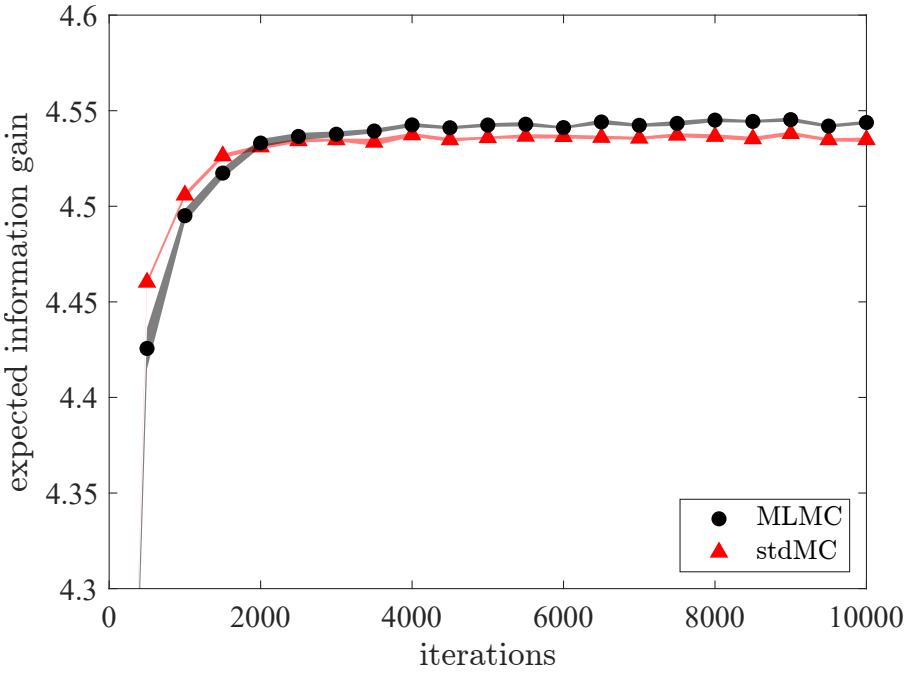


FIG. 6. *The behavior of the expected information gain as a function of the number of iteration steps for the PK model*

gradient-based optimization algorithms in which the number of inner Monte Carlo samples is kept fixed only gives a biased solution of Bayesian experimental design unless i.i.d. sampling from the exact posterior distribution is possible. To overcome this issue, we have introduced an unbiased antithetic multilevel Monte Carlo estimator for the gradient of the expected information gain, and have proven under some conditions that our estimator is unbiased and has finite expected squared ℓ_2-norm and finite computational cost per one sample. This way, combining our unbiased multilevel estimator with stochastic gradient-based optimization algorithms leads to a novel stochastic algorithm to search for optimal Bayesian experimental designs without suffering from any bias. Numerical experiments for a simple test case show that our proposed algorithm can find the true optimal Bayesian experimental design with the convergence behavior as expected from the standard stochastic optimization theory which is built upon the underlying assumption that an unbiased gradient estimation is possible. In contrast, using the standard Monte Carlo estimator with a fixed number of inner samples fails to reach the optimal design. Moreover, our proposed algorithm performs well for a more realistic pharmacokinetic test problem and gives a higher expected information gain and qualitatively different sampling times compared to designs obtained by the existing standard Monte Carlo estimator.

Acknowledgements. The authors would like to thank the reviewers for their helpful comments and suggestions which lead to a significant improvement over the original manuscript. The authors are also grateful to Takuro Mori (University of Tokyo) for his help on numerical experiments during a revision process. TG would like to thank Professor Mike Giles (University of Oxford) for useful discussions and comments at an early stage of this research.

Appendix A. Proof of Theorem 3.1. The proof for the first assertion follows an argument similar to that of [17, Lemma 3.9] which considers a nested expectation

involving the ratio of two scalar inner conditional expectations. Since the numerator is vector-valued in our setting, however, we give a proof for the sake of completeness.

First let us recall the following result proven, for instance, in [14, Lemma 1].

LEMMA A.1. *Let X be a real-valued random variable with mean zero, and let $\overline{X}_N$ be an average of N i.i.d. samples of X. If $\mathbb{E}[|X|^u] < \infty$ for $u > 2$, there exists a constant $C_u > 0$ depending only on u such that*

$$\mathbb{E}\left[\left|\overline{X}_N\right|^u\right] \le C_u \frac{\mathbb{E}[|X|^u]}{N^{u/2}} \quad \text{and} \quad \mathbb{P}\left[\left|\overline{X}_N\right| > c\right] \le C_u \frac{\mathbb{E}[|X|^u]}{c^u N^{u/2}},$$

for any $c > 0$.

For any θ, ϵ and ξ, we write $\rho(f_\xi(\theta,\epsilon) \mid \xi) = \mathbb{E}_{\theta'}\left[\rho(f_\xi(\theta,\epsilon) \mid \theta', \xi)\right]$ and also $\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \xi) = \mathbb{E}_{\theta'}\left[\nabla_\xi \rho(f_\xi(\theta,\epsilon) \mid \theta', \xi)\right]$. For randomly chosen θ and ϵ, we define an extreme event A by

$$A := \left\{ \left| \frac{\varrho^{(a)}_{\xi, M_0 2^{\ell-1}, q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1 \right| > \frac{1}{2} \right\} \bigcup \left\{ \left| \frac{\varrho^{(b)}_{\xi, M_0 2^{\ell-1}, q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1 \right| > \frac{1}{2} \right\}.$$

Then we have

$$\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2] = \mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2 \mathbf{1}_A] + \mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2 \mathbf{1}_{A^c}], \tag{A.1}$$

where $\mathbf{1}_\bullet$ denotes the indicator function of an event $\bullet$ and A^c denotes the complement of the event A.

Let us look at the first term on the right-hand side of (A.1). Since we use the same i.i.d. samples of $\theta' \sim q$ in the denominator and numerator for the three terms of $\Delta\psi_{\xi,\ell}$, i.e., $\psi_{\xi, M_0 2^\ell, q}, \psi^{(a)}_{\xi, M_0 2^{\ell-1}, q}, \psi^{(b)}_{\xi, M_0 2^{\ell-1}, q}$, it follows from the assumption

$$\sup_{\theta,\theta',\epsilon} \left\|\nabla_\xi \log \rho(f_\xi(\theta,\epsilon) \mid \theta', \xi)\right\|_\infty =: \varrho_{\max} < \infty$$

that $\left\|\psi_{\xi, M_0 2^\ell, q}\right\|_2^2, \left\|\psi^{(a)}_{\xi, M_0 2^{\ell-1}, q}\right\|_2^2, \left\|\psi^{(b)}_{\xi, M_0 2^{\ell-1}, q}\right\|_2^2 \le 2d\varrho_{\max}^2$ where d denotes the cardinality of ξ. Applying Jensen's inequality leads to a bound

$$\begin{aligned}
\|\Delta\psi_{\xi,\ell}\|_2^2 &\le \left(\left\|\psi_{\xi, M_0 2^\ell, q}\right\|_2 + \frac{\left\|\psi^{(a)}_{\xi, M_0 2^{\ell-1}, q}\right\|_2}{2} + \frac{\left\|\psi^{(b)}_{\xi, M_0 2^{\ell-1}, q}\right\|_2}{2} \right)^2 \\
&\le 2\left\|\psi_{\xi, M_0 2^\ell, q}\right\|_2^2 + \left\|\psi^{(a)}_{\xi, M_0 2^{\ell-1}, q}\right\|_2^2 + \left\|\psi^{(b)}_{\xi, M_0 2^{\ell-1}, q}\right\|_2^2 \le 8d\varrho_{\max}^2.
\end{aligned}$$

Thus we have

$$\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2 \mathbf{1}_A] \le 8d\varrho_{\max}^2 \mathbb{E}[\mathbf{1}_A] = 8d\varrho_{\max}^2 \mathbb{P}[A].$$

Noting that both $\varrho^{(a)}_{\xi, M_0 2^{\ell-1}, q}(\theta,\epsilon)$ and $\varrho^{(b)}_{\xi, M_0 2^{\ell-1}, q}(\theta,\epsilon)$ are unbiased estimates of the target quantity $\rho(f_\xi(\theta,\epsilon) \mid \xi)$ using $M_0 2^{\ell-1}$ random samples of $\theta' \sim q$, it follows from the assumption of the theorem and Lemma A.1 that

$$\mathbb{P}[A] \le \mathbb{P}\left[\left| \frac{\varrho^{(a)}_{\xi, M_0 2^{\ell-1}, q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1 \right| > \frac{1}{2}\right] + \mathbb{P}\left[\left| \frac{\varrho^{(b)}_{\xi, M_0 2^{\ell-1}, q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1 \right| > \frac{1}{2}\right]$$

$$\leq \frac{2^{u+1}C_u}{(M_0 2^{\ell-1})^{u/2}} \mathbb{E}_{\theta,\theta',\epsilon}\left[\left|\frac{\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\pi_0(\theta')}{\rho(f_\xi(\theta,\epsilon) \mid \xi)q(\theta')} - 1\right|^u\right]$$
$$\leq \frac{2^{u+1}C_u}{(M_0 2^{\ell-1})^{u/2}} \left(\mathbb{E}_{\theta,\theta',\epsilon}\left[\left|\frac{\rho(f_\xi(\theta,\epsilon) \mid \theta',\xi)\pi_0(\theta')}{\rho(f_\xi(\theta,\epsilon) \mid \xi)q(\theta')}\right|^u\right] + 1\right).$$

This gives a bound on the term $\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2 \mathbf{1}_A]$ of order $2^{-(u/2)\ell}$.

Next let us look at the second term on the right-hand side of (A.1). By using the antithetic properties (3.5), we have

$$\begin{aligned}
&\Delta\psi_{\xi,\ell} \\
&= \frac{1}{2}\left(\nabla\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) - \nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)\right)\left(\frac{1}{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)} - \frac{1}{\rho(f_\xi(\theta,\epsilon) \mid \xi)}\right) \\
&\quad + \frac{1}{2}\left(\nabla\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) - \nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)\right)\left(\frac{1}{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)} - \frac{1}{\rho(f_\xi(\theta,\epsilon) \mid \xi)}\right) \\
&\quad - \left(\nabla\varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon) - \nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)\right)\left(\frac{1}{\varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon)} - \frac{1}{\rho(f_\xi(\theta,\epsilon) \mid \xi)}\right) \\
&\quad + \frac{1}{2}\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)}{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}\left(\frac{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1\right)^2 \\
&\quad + \frac{1}{2}\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)}{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}\left(\frac{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1\right)^2 \\
&\quad - \frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)}{\varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon)}\left(\frac{\varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1\right)^2.
\end{aligned}$$

Noting that

$$\left|\frac{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1\right|, \left|\frac{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1\right| \leq \frac{1}{2}$$

on A^c and that $\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)$, $\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)$ and $\rho(f_\xi(\theta,\epsilon) \mid \xi)$ are strictly positive by assumption, it holds that

$$\frac{1}{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}, \frac{1}{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)} \leq \frac{2}{\rho(f_\xi(\theta,\epsilon) \mid \xi)}.$$

The same bound exists also for $\varrho_{\xi,M_0 2^{\ell},q}(\theta,\epsilon)$ because of the antithetic property (3.5). By applying Jensen's inequality and then using these bounds, we obtain

$$\begin{aligned}
\|\Delta\psi_{\xi,\ell}\|_2^2 \leq\ & 2\left\|\frac{\nabla\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) - \nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)}{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}\right\|_2^2 \left(\frac{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1\right)^2 \\
&+ 2\left\|\frac{\nabla\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon) - \nabla_\xi\rho(f_\xi(\theta,\epsilon) \mid \xi)}{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}\right\|_2^2 \left(\frac{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon) \mid \xi)} - 1\right)^2
\end{aligned}$$

$$
\begin{aligned}
&+4\left\|\frac{\nabla\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)-\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}\right\|_2^2\left(\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^2\\
&+2\left\|\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}\right\|_2^2\left(\frac{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^4\\
&+2\left\|\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}\right\|_2^2\left(\frac{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^4\\
&+4\left\|\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}\right\|_2^2\left(\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^4\\
\le\ &8\left\|\frac{\nabla\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)-\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^2\\
&+8\left\|\frac{\nabla\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)-\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^2\\
&+16\left\|\frac{\nabla\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)-\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^2\\
&+8\left\|\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho^{(a)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^4\\
&+8\left\|\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho^{(b)}_{\xi,M_0 2^{\ell-1},q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^4\\
&+16\left\|\frac{\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^4.
\end{aligned}
\tag{A.2}
$$

Let us focus on the third term of (A.2). Applying Hölder's inequality gives

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{\nabla\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)-\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^2\mathbf{1}_{A^c}\right]\\
&\le\mathbb{E}\left[\left\|\frac{\nabla\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)-\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\right.\\
&\qquad\left.\times\,2^{\max(4-u,0)}\left|\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right|^{\min(u,4)-2}\right]\\
&\le\left(\mathbb{E}\left[\left\|\frac{\nabla\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)-\nabla_\xi\rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^{\min(u,4)}\right]\right)^{2/\min(u,4)}\\
&\qquad\times\,2^{\max(4-u,0)}\left(\mathbb{E}\left[\left|\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right|^{\min(u,4)}\right]\right)^{1-2/\min(u,4)}.
\end{aligned}
$$

Using Jensen's inequality and Lemma A.1, the first factor above is bounded by

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{\nabla \varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)-\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^{\min(u,4)}\right]\\
&\le d^{\min(u,4)/2-1}\mathbb{E}\left[\left\|\frac{\nabla \varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)-\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_{\min(u,4)}^{\min(u,4)}\right]\\
&\le \frac{d^{\min(u,4)/2-1}C_{\min(u,4)}}{(M_0 2^\ell)^{\min(u,4)/2}}\\
&\quad\times \mathbb{E}_{\theta,\theta',\epsilon}\left[\left\|\frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')/q(\theta')-\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_{\min(u,4)}^{\min(u,4)}\right]\\
&\le \frac{2^{\min(u,4)-1}d^{\min(u,4)/2-1}C_{\min(u,4)}}{(M_0 2^\ell)^{\min(u,4)/2}}\\
&\quad\times \mathbb{E}_{\theta,\theta',\epsilon}\left[\left\|\frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')/q(\theta')}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_{\min(u,4)}^{\min(u,4)}+\left\|\frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_{\min(u,4)}^{\min(u,4)}\right]\\
&\le \frac{2^{\min(u,4)-1}d^{\min(u,4)/2-1}C_{\min(u,4)}}{(M_0 2^\ell)^{\min(u,4)/2}}\\
&\quad\times \mathbb{E}_{\theta,\theta',\epsilon}\left[\left\|\frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')/q(\theta')}{\rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')/q(\theta')}\cdot\frac{\rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')/q(\theta')}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_{\min(u,4)}^{\min(u,4)}\right.\\
&\qquad\left.+\left\|\frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_{\min(u,4)}^{\min(u,4)}\right]\\
&\le \frac{2^{\min(u,4)-1}d^{\min(u,4)/2}\varrho_{\max}^{\min(u,4)}C_{\min(u,4)}}{(M_0 2^\ell)^{\min(u,4)/2}}\\
&\quad\times\left(\mathbb{E}_{\theta,\theta',\epsilon}\left[\left|\frac{\rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')}{\rho(f_\xi(\theta,\epsilon)\mid\xi)q(\theta')}\right|^{\min(u,4)}\right]+1\right),
\end{aligned}
$$

whereas a bound on the second factor directly follows from Lemma A.1, i.e., we have

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right|^{\min(u,4)}\right]\\
&\le \frac{C_{\min(u,4)}}{(M_0 2^\ell)^{\min(u,4)/2}}\mathbb{E}_{\theta,\theta',\epsilon}\left[\left|\frac{\rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')}{\rho(f_\xi(\theta,\epsilon)\mid\xi)q(\theta')}-1\right|^{\min(u,4)}\right]\\
&\le \frac{C_{\min(u,4)}}{(M_0 2^\ell)^{\min(u,4)/2}}\left(\mathbb{E}_{\theta,\theta',\epsilon}\left[\left|\frac{\rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')}{\rho(f_\xi(\theta,\epsilon)\mid\xi)q(\theta')}\right|^{\min(u,4)}\right]+1\right).
\end{aligned}
$$

Substituting these bounds shows that the third term is of order

$$
\left(2^{-\min(u,4)\ell/2}\right)^{2/\min(u,4)}\cdot\left(2^{-\min(u,4)\ell/2}\right)^{1-2/\min(u,4)}=2^{-\min(u,4)\ell/2}
$$

for given $u>2$.

Similarly, the expectation of the sixth term of (A.2) can be bounded above by

$$
\mathbb{E}\left[\left\|\frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left(\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right)^4\mathbf{1}_{A^c}\right]
$$

$$
\begin{aligned}
&\leq 2^{\max(4-u,0)}\mathbb{E}\left[\left\|\frac{\nabla_\xi \rho(f_\xi(\theta,\epsilon)\mid\xi)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}\right\|_2^2\left|\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right|^{\min(u,4)}\right]\\
&\leq 2^{\max(4-u,0)}d\varrho_{\max}^2\mathbb{E}\left[\left|\frac{\varrho_{\xi,M_0 2^\ell,q}(\theta,\epsilon)}{\rho(f_\xi(\theta,\epsilon)\mid\xi)}-1\right|^{\min(u,4)}\right]\\
&\leq \frac{2^{\max(4-u,0)}d\varrho_{\max}^2 C_{\min(u,4)}}{(M_0 2^\ell)^{\min(u,4)/2}}\left(\mathbb{E}_{\theta,\theta',\epsilon}\left[\left|\frac{\rho(f_\xi(\theta,\epsilon)\mid\theta',\xi)\pi_0(\theta')}{\rho(f_\xi(\theta,\epsilon)\mid\xi)q(\theta')}\right|^{\min(u,4)}\right]+1\right).
\end{aligned}
$$

It is obvious that the other terms of (A.2) can be bounded similarly. This way we obtain a bound on the term $\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2\mathbf{1}_{A^c}]$ of order $2^{-\min(u,4)\ell/2}$, which completes the proof of the first assertion of the theorem.

Let us move on to the second assertion. By choosing $w_\ell \propto 2^{-\tau\ell}$, it follows from the first assertion that

$$
\sum_{\ell=0}^{\infty}\frac{\mathbb{E}[\|\Delta\psi_{\xi,\ell}\|_2^2]}{w_\ell}\propto\sum_{\ell=0}^{\infty}2^{-(\beta-\tau)\ell},
$$

and

$$
\sum_{\ell=0}^{\infty}C_\ell w_\ell\propto\sum_{\ell=0}^{\infty}2^{-(\tau-1)\ell}.
$$

Thus, if $1<\tau<\beta$, these two quantities are obviously bounded. It is important to remark that we have these finite bounds on the expected squared ℓ_2-norm and the expected computational cost of the random variable $\Delta\psi_{\xi,\ell}/w_\ell$, since we assume $u>2$, which ensures $\beta>1$.

REFERENCES

[1] S. Asmussen and P. W. Glynn, *Stochastic Simulation*, Springer, New York, 2007.

[2] J. Beck, B. M. Dia, L. F. R. Espath, Q. Long, and R. Tempone, *Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain*, Computer Methods in Applied Mechanics and Engineering, 334 (2018), pp. 523–553, https://doi.org/10.1016/j.cma.2018.01.053.

[3] J. Beck, B. M. Dia, L. F. R. Espath, and R. Tempone, *Multilevel double loop Monte Carlo and stochastic collocation methods with importance sampling for Bayesian optimal experimental design*, International Journal for Numerical Methods in Engineering, 121 (2020), pp. 3482–3503, https://doi.org/10.1002/nme.6367.

[4] K. Bujok, B. Hambly, and C. Reisinger, *Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives*, Methodology and Computing in Applied Probability, 17 (2015), pp. 579–604, https://doi.org/10.1007/s11009-013-9380-5.

[5] A. G. Carlon, B. M. Dia, L. F. R. Espath, R. H. Lopez, and R. Tempone, *Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization*, Computer Methods in Applied Mechanics and Engineering, 363 (2020), 112909, https://doi.org/10.1016/j.cma.2020.112909.

[6] K. Chaloner and I. Verdinelli, *Bayesian experimental design: a review*, Statistical Science, 10 (1995), pp. 273–304, https://doi.org/10.1214/ss/1177009939.

[7] S. Dereich and T. Müller-Gronbach, *General multilevel adaptations for stochastic approximation algorithms of Robbins-Monro and Polyak-Ruppert type*, Numerische Mathematik, 142 (2019), pp. 279–328, https://doi.org/10.1007/s00211-019-01024-y.

[8] J. Duchi, E. Hazan, and Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization*, Journal of Machine Learning Research, 12 (2011), pp. 2121–2159.

[9] A. Foster, M. Jankowiak, E. Bingham, P. Horsfall, Y. W. Teh, T. Rainforth, and N. Goodman, *Variational Bayesian optimal experimental design*, in 33rd Conference on

Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, 2019, https://arxiv.org/abs/1903.05480.

[10] A. Foster, M. Jankowiak, M. O'Meara, Y. W. Teh, and T. Rainforth, *A unified stochastic gradient approach to designing Bayesian-optimal experiments*, in 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), Palermo, Italy, 2020, https://arxiv.org/abs/1911.00294.

[11] N. Frikha, *Multilevel stochastic approximation algorithms*, Annals of Applied Probability, 26 (2016), pp. 933–985, https://doi.org/10.1214/15-AAP1109.

[12] M. B. Giles, *Multilevel Monte Carlo path simulation*, Operations Research, 56 (2008), pp. 607–617, https://doi.org/10.1287/opre.1070.0496.

[13] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numerica, 24 (2015), pp. 259–328, https://doi.org/10.1017/S096249291500001X.

[14] M. B. Giles and T. Goda, *Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI*, Statistics and Computing, 29 (2019), pp. 739–751, https://doi.org/10.1007/s11222-018-9835-1.

[15] M. B. Giles and L. Szpruch, *Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation*, Annals of Applied Probability, 24 (2014), pp. 1585–1620, https://doi.org/10.1214/13-AAP957.

[16] T. Goda, T. Hironaka, and T. Iwamoto, *Multilevel Monte Carlo estimation of expected information gains*, Stochastic Analysis and Applications, 38 (2020), pp. 581–600, https://doi.org/10.1080/07362994.2019.1705168.

[17] T. Hironaka, M. B. Giles, T. Goda, and H. Thom, *Multilevel Monte Carlo estimation of the expected value of sample information*, SIAM/ASA Journal on Uncertainty Quantification, 8 (2020), pp. 1236–1259, https://doi.org/10.1137/19M1284981.

[18] X. Huan and Y. M. Marzouk, *Gradient-based stochastic optimization methods in Bayesian experimental design*, International Journal for Uncertainty Quantification, 4 (2014), pp. 479–510, https://doi.org/10.1615/Int.J.UncertaintyQuantification.2014006730.

[19] D. P. Kingma and J. L. Ba, *Adam: A method for stochastic optimization*, Dec. 2014, https://arxiv.org/abs/1412.6980.

[20] S. Kleinegesse and M. U. Gutmann, *Bayesian experimental design for implicit models by mutual information neural estimation*, in 37th International Conference on Machine Learning (ICML 2020), 2020, https://arxiv.org/abs/2002.08129.

[21] D. V. Lindley, *On a measure of the information provided by an experiment*, The Annals of Mathematical Statistics, 27 (1956), pp. 986–1005, https://doi.org/10.1214/aoms/1177728069.

[22] Q. Long, M. Scavino, R. Tempone, and S. Wang, *Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations*, Computer Methods in Applied Mechanics and Engineering, 259 (2013), pp. 24–39, https://doi.org/10.1016/j.cma.2013.02.017.

[23] J. I. Myung, D. R. Cavagnaro, and M. A. Pitt, *A tutorial on adaptive design optimization*, Journal of Mathematical Psychology, 57 (2013), pp. 53–67, https://doi.org/10.1016/j.jmp.2013.05.005.

[24] Y. E. Nesterov, *A method of solving a convex programming problem with convergence rate* $o(1/k^2)$, Soviet Mathematics Doklady, 27 (1983), pp. 372–376.

[25] A. B. Owen, *Monte carlo theory, methods and examples*, 2019, https://statweb.stanford.edu/~owen/mc/.

[26] B. T. Polyak, *A new method of stochastic approximation type (in Russian)*, Avtomatika i Telemekhanika, 7 (1990), pp. 98–107.

[27] T. Rainforth, R. Cornish, H. Yang, A. Warrington, and F. Wood, *On nesting Monte Carlo estimators*, in 35th International Conference on Machine Learning, Stockholm, Sweden, 2018, http://proceedings.mlr.press/v80/rainforth18a.html.

[28] S. J. Reddi, S. Kale, and S. Kumar, *On the convergence of Adam and beyond*, Apr. 2019, https://arxiv.org/abs/1904.09237.

[29] C. H. Rhee and P. Glynn, *Unbiased estimation with square root convergence for SDE models*, Operations Research, 63 (2015), pp. 1026–1043, https://doi.org/10.1287/opre.2015.1404.

[30] H. Robbins and S. Monro, *A stochastic approximation method*, The Annals of Mathematical Statistics, 22 (1951), pp. 400–407, https://doi.org/10.1214/aoms/1177729586.

[31] D. Ruppert, *Stochastic approximation*, in Handbook of Sequential Analysis, B. K. Ghosh and P. K. Sen, eds., Dekker, New York, 1991, pp. 503–529.

[32] E. G. Ryan, C. D. Drovandi, and A. N. Pettitt, *Fully Bayesian experimental design for pharmacokinetic studies*, Entropy, 17 (2015), pp. 1063–1089, https://doi.org/10.3390/e17031063.

[33] E. G. Ryan, C. D. Drovandi, M. Thompson, and A. N. Pettitt, *Towards Bayesian experimental design for nonlinear models that require a large number of sampling times*, Computational Statistics and Data Analysis, 70 (2014), pp. 45–60, https://doi.org/10.1016/j.csda.2013.08.017.

[34] K. J. Ryan, *Estimating expected information gains for experimental designs with application to the random fatigue-limit model*, Journal of Computational and Graphical Statistics, 12 (2003), pp. 585–603, https://doi.org/10.1198/1061860032012.

[35] C. Schillings, B. Sprungk, and P. Wacker, *On the convergence of the Laplace approximation and noise-level-robustness of Laplace-based Monte Carlo methods for Bayesian inverse problems*, Numerische Mathematik, 145 (2020), pp. 915–971, https://doi.org/10.1007/s00211-020-01131-1.

[36] B. Shababo, B. Paige, A. Pakman, and L. Paninski, *Bayesian inference and online experimental design for mapping neural microcircuits*, in Advances in Neural Information Processing Systems 26 (NIPS 2013), 2013.

[37] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming*, SIAM, Philadelphia, 2009.

[38] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numerica, 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061.

[39] T. Tieleman and G. Hinton, *Lecture 6.5 – RMSProp*, COURSERA: Neural networks for machine learning, 4 (2012), pp. 26–31.

[40] J. Vanlier, C. A. Tiemann, P. A. J. Hilbers, and N. A. W. van Riel, *A Bayesian approach to targeted experiment design*, Bioinformatics, 28 (2012), pp. 1136–1142, https://doi.org/10.1093/bioinformatics/bts092.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs
Publication Status	Accepted for publication
Publication Details	Takashi Goda, Tomohiko Hironaka, Wataru Kitade, Adam Foster (2021). Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs. SIAM Journal on Scientific Computing (to appear).

Student Confirmation

Student Name:	Adam Foster		
Contribution to the Paper	Fourth author. Contributed to the development of the paper, including the use of reparametrization. Editing of the entire manuscript, including proofs.		
Signature		Date	25/10/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Dr Tom Rainforth		
Supervisor comments I have verified Adam's above account with the lead author of the paper.		
Signature 	Date	26/10/21

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 5

Deep Adaptive Design: Amortizing Bayesian Experimental Design

This paper was published as the following

Adam Foster, Desi R Ivanova, Ilyas Malik, and Thomas Rainforth. Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design. In *Proceedings of the 38th International Conference on Machine Learning*, pages 3384–3395. PMLR, 2021.

Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design

Adam Foster [*1] **Desi R. Ivanova** [*1] **Ilyas Malik** [2] **Tom Rainforth** [1]

Abstract

We introduce *Deep Adaptive Design* (DAD), a method for amortizing the cost of adaptive Bayesian experimental design that allows experiments to be run in real-time. Traditional sequential Bayesian optimal experimental design approaches require substantial computation at *each* stage of the experiment. This makes them unsuitable for most real-world applications, where decisions must typically be made quickly. DAD addresses this restriction by learning an amortized *design network* upfront and then using this to rapidly run (multiple) adaptive experiments at deployment time. This network represents a design *policy* which takes as input the data from previous steps, and outputs the next design using a single forward pass; these design decisions can be made in milliseconds during the live experiment. To train the network, we introduce contrastive information bounds that are suitable objectives for the sequential setting, and propose a customized network architecture that exploits key symmetries. We demonstrate that DAD successfully amortizes the process of experimental design, outperforming alternative strategies on a number of problems.

1. Introduction

A key challenge across disciplines as diverse as psychology (Myung et al., 2013), bioinformatics (Vanlier et al., 2012), pharmacology (Lyu et al., 2019) and physics (Dushenko et al., 2020) is to design experiments so that the outcomes will be as informative as possible about the underlying process. Bayesian optimal experimental design (BOED) is a powerful mathematical framework for tackling this problem (Lindley, 1956; Chaloner & Verdinelli, 1995).

In the BOED framework, outcomes y are modeled in a Bayesian manner (Gelman et al., 2013; Kruschke, 2014) using a likelihood $p(y|\theta, \xi)$ and a prior $p(\theta)$, where ξ is our controllable design and θ is the set of parameters we wish to learn about. We then optimize ξ to maximize the *expected information gained* about θ (equivalently the mutual information between y and θ):

$$I(\xi) := \mathbb{E}_{p(\theta)p(y|\theta,\xi)} \left[\log p(y|\theta, \xi) - \log p(y|\xi)\right]. \quad (1)$$

The true power of BOED is realized when it is used to design a sequence of experiments $\xi_1, ..., \xi_T$, wherein it allows us to construct *adaptive* strategies which utilize information gathered from past data to tailor each successive design ξ_t during the progress of the experiment. The conventional, iterative, approach for selecting each ξ_t is to fit the posterior $p(\theta|\xi_{1:t-1}, y_{1:t-1})$ representing the updated beliefs about θ after $t-1$ iterations have been conducted, and then substitute this for the prior in (1) (Ryan et al., 2016; Rainforth, 2017; Kleinegesse et al., 2020). The design ξ_t is then chosen as the one which maximizes the resulting objective.

Unfortunately, this approach necessitates significant computational time to be expended *between each step of the experiment* in order to update the posterior and compute the next optimal design. In particular, $I(\xi)$ is doubly intractable (Rainforth et al., 2018; Zheng et al., 2018) and its optimization constitutes a significant computational bottleneck. This can be prohibitive to the practical application of sequential BOED as design decisions usually need to be made quickly for the approach to be useful (Evans & Mathur, 2005).

To give a concrete example, consider running an adaptive survey to understand political opinions (Pasek & Krosnick, 2010). A question ξ_t is put to a participant who gives their answer y_t and this data is used to update an underlying model with latent variables θ. Here sequential BOED is of immense value because previous answers can be used to guide future questions, ensuring that they are pertinent to the particular participant. However, it is not acceptable to have lengthy delays between questions to compute the next design, precluding existing approaches from being used.

To alleviate this problem, we propose *amortizing* the cost of sequential experimental design, performing upfront training before the start of the experiment to allow very fast design decisions at deployment, when time is at a premium. This amortization is particularly useful in the common scenario where the same adaptive experimental framework will be

[*]Equal contribution [1]Department of Statistics, University of Oxford, UK [2]Work undertaken whilst at the University of Oxford. Correspondence to: Adam Foster <adam.foster@stats.ox.ac.uk>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

deployed numerous times (e.g. having multiple participants in a survey). Here amortization not only removes the computational burden from the live experiment, it also allows for sharing computation across multiple experiments, analogous to inference amortization that allows one to deal with multiple datasets (Stuhlmüller et al., 2013).

Our approach, called **Deep Adaptive Design (DAD)**, constructs a single *design network* which takes as input the designs and observations from previous stages, and outputs the design to use for the next experiment. The network is learned by simulating hypothetical experimental trajectories and then using these to train the network to make near-optimal design decisions automatically. That is, it learns a *design policy* which makes decisions as a function of the past data, and we optimize the parameters of this policy rather than an individual design. Once learned, the network eliminates the computational bottleneck at each iteration of the experiment, enabling it to be run both adaptively and quickly; it can also be used repeatedly for different instantiations of the experiment (e.g. different human participants).

To allow for efficient, effective, and simple training, we show how DAD networks can be learned without any direct posterior or marginal likelihood estimation. This is achieved by reformulating the sequential BOED problem from its conventional iterative form, to a single holistic objective based on the *overall* expected information gained from the entire experiment when using a policy to make each design decision deterministically given previous design outcome pairs. We then derive contrastive bounds on this objective that allow for end-to-end training of the policy parameters with stochastic gradient ascent, thereby sidestepping both the need for inference and the double intractability of the EIG objective. This approach has the further substantial benefit of allowing non-myopic adaptive strategies to be learned, that is strategies which take account of their own future decisions, unlike conventional approaches.

We further demonstrate a key permutation symmetry property of the optimal design policy, and use this to propose a customized architecture for the experimental design network. This is critical to allowing effective amortization across time steps. The overall result of the theoretical formulation, novel contrastive bounds, and neural architecture is a methodology which enables us to bring the power of deep learning to bear on adaptive experimental design.

We apply DAD to a range of problems relevant to applications such as epidemiology, physics and psychology. We find that DAD is able to accurately amortize experiments, opening the door to running adaptive BOED in real time.

2. Background

Because experimentation is a potentially costly endeavour, it is essential to design experiments in manner that maximizes the amount of information garnered. The BOED framework, pioneered by Lindley (1956), provides a powerful means of doing this in a principled manner. Its key idea is to optimize the experimental design ξ to maximize the expected amount of *information* that will be gained about the latent variables of interest, θ, upon observing the experiment outcome y.

To implement this approach, we begin with the standard Bayesian modelling set-up consisting of an explicit likelihood model $p(y|\theta, \xi)$ for the experiment, and a prior $p(\theta)$ representing our initial beliefs about the unknown latent. After running a hypothetical experiment with design ξ and observing y, our updated beliefs are the posterior $p(\theta|\xi, y)$. The amount of information that has been gained about θ can be mathematically described by the reduction in entropy from the prior to the posterior

$$\mathrm{IG}(\xi, y) = \mathrm{H}\left[p(\theta)\right] - \mathrm{H}\left[p(\theta|\xi, y)\right]. \tag{2}$$

The *expected information gain* (EIG) is formed by taking the expectation over possible outcomes y, using the model itself to simulate these. Namely we take an expectation with respect to $y \sim p(y|\xi) = \mathbb{E}_{p(\theta)}[p(y|\theta, \xi)]$, yielding

$$\begin{aligned} I(\xi) &:= \mathbb{E}_{p(y|\xi)}\left[\mathrm{IG}(\xi, y)\right] \\ &= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log p(\theta|\xi, y) - \log p(\theta)\right] \\ &= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log p(y|\theta, \xi) - \log p(y|\xi)\right] \end{aligned}$$

which is the mutual information between y and θ under design ξ. The optimal design is defined as $\xi^* = \arg\max_{\xi\in\Xi} I(\xi)$, where Ξ is the space of feasible designs.

It is common in BOED settings to be able to run multiple experiment iterations with designs $\xi_1, ..., \xi_T$, observing respective outcomes $y_1, ..., y_T$. One simple strategy for this case is *static* design, also called fixed or batch design, which selects all $\xi_1, ..., \xi_T$ before making any observation. The designs are optimized to maximize the EIG, with $y_{1:T}$ in place of y and $\xi_{1:T}$ in place of ξ, effectively treating the whole sequence of experiments as one experiment with enlarged observation and design spaces.

2.1. Conventional adaptive BOED

This static design approach is generally sub-optimal as it ignores the fact that information from previous iterations can substantially aid in the design decisions at future iterations. The power of the BOED framework can thus be significantly increased by using an *adaptive* design strategy that chooses each ξ_t dependent upon $\xi_{1:t-1}, y_{1:t-1}$. This enables us to use what has already been learned in previous experiments to design the next one optimally, resulting in a virtuous cycle of refining beliefs and using our updated beliefs to design good experiments for future iterations.

The conventional approach to computing designs adaptively is to fit the posterior distribution $p(\theta|\xi_{1:t-1}, y_{1:t-1})$ at each

step, and then optimize the EIG objective that uses this posterior in place of the prior (Ryan et al., 2016)

$$I(\xi_t) = \mathbb{E}_{p(\theta|\xi_{1:t-1}, y_{1:t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(y_t|\theta,\xi_t)}{p(y_t|\xi_t)}\right] \quad (3)$$

where $p(y_t|\xi_t) = \mathbb{E}_{p(\theta|\xi_{1:t-1}, y_{1:t-1})}[p(y_t|\theta,\xi_t)]$.

Despite the great potential of the adaptive BOED framework, this conventional approach is very computationally expensive. At each stage t of the experiment we must compute the posterior $p(\theta|\xi_{1:t-1}, y_{1:t-1})$, which is costly and cannot be done in advance as it depends on $y_{1:t-1}$. Furthermore, the posterior is then used to obtain ξ_t by maximizing the objective in (3), which is computationally even more demanding as it involves the optimization of a doubly intractable quantity (Rainforth et al., 2018; Foster et al., 2019). Both of these steps must be done during the experiment, meaning it is infeasible to run adaptive BOED in real time experiment settings unless the model is unusually simple.

2.2. Contrastive information bounds

In Foster et al. (2020), the authors noted that if $\xi \in \Xi$ is continuous, approximate optimization of the EIG at each stage of the experiment can be achieved in a single *unified* stochastic gradient procedure that both estimates and optimizes the EIG simultaneously. A key component of this approach is the derivation of several contrastive lower bounds on the EIG, inspired by work in representation learning (van den Oord et al., 2018; Poole et al., 2019). One such bound is the Prior Contrastive Estimation (PCE) bound, given by

$$I(\xi) \geq \mathbb{E}\left[\log \frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(y|\theta_\ell,\xi)}\right] \quad (4)$$

where $\theta_0 \sim p(\theta)$ is the sample used to generate $y \sim p(y|\theta,\xi)$ and $\theta_{1:L}$ are L contrastive samples drawn independently from $p(\theta)$; as $L \to \infty$ the bound becomes tight. The PCE bound can be maximized by stochastic gradient ascent (SGA) (Robbins & Monro, 1951) to approximate the optimal design ξ. As discussed previously, in a sequential setting this stochastic gradient optimization is repeated T times, with $p(\theta)$ replaced by $p(\theta|\xi_{1,t-1}, y_{1:t-1})$ at step t.

3. Rethinking Sequential BOED

To enable adaptive BOED to be deployed in settings where design decisions must be taken quickly, we first need to rethink the traditional iterative approach to produce a formulation which considers the entire design process holistically. To this end, we introduce the concept of a *design function*, or *policy*, π that maps from the set of all previous design–observation pairs to the next chosen design.

Let h_t denote the experimental *history* $(\xi_1, y_1), ..., (\xi_t, y_t)$. We can simulate histories for a given policy π, by sampling a $\theta \sim p(\theta)$, then, for each $t = 1, ..., T$, fixing $\xi_t = \pi(h_{t-1})$ (where $h_0 = \varnothing$) and sampling $y_t \sim p(y|\theta,\xi_t)$. The density of this generative process can be written as

$$p(\theta)p(h_T|\theta,\pi) = p(\theta)\prod_{t=1}^{T} p(y_t|\theta,\xi_t). \quad (5)$$

The standard sequential BOED approach described in § 2.1 now corresponds to a costly implicit policy π_s, that performs posterior estimation followed by EIG optimization to choose each design. By contrast, in DAD, we will learn a deterministic π that chooses designs directly.

Another way to think about π_s is that it is the policy which piecewise optimizes the following objective for $\xi_t|h_{t-1}$

$$I_{h_{t-1}}(\xi_t) := \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(y_t|\theta,\xi_t)}{p(y_t|h_{t-1},\xi_t)}\right] \quad (6)$$

where $p(y_t|h_{t-1},\xi_t) = \mathbb{E}_{p(\theta|h_{t-1})}[p(y_t|\theta,\xi_t)]$. It is thus the optimal *myopic* policy—that is a policy which fails to reason about its own future actions—for an objective given by the sum of EIGs from each experiment iteration. Note that this is not the optimal *overall* policy as it fails to account for future decision making: some designs may allow better future design decisions than others than others (González et al., 2016; Jiang et al., 2020).[1]

Trying to learn an efficient policy that directly mimics π_s would be very computationally challenging because of the difficulties of dealing with both inference and EIG estimation at each iteration of the training. Indeed, the natural way to do this involves running a full, very expensive, simulated sequential BOED process to generate each training example.

We instead propose a novel strategy that reformulates the sequential decision problem in a way that completely eliminates the need for calculating either posterior distributions or intermediate EIGs, while also allowing for non-myopic policies to be learned. This is done by exploiting an important property of the EIG: the total EIG of a sequential experiment is the sum of the (conditional) EIGs for each experiment iteration. This is formalized in the following result, which provides a single expression for the expected information gained from the entire sequence of T experiments.

Theorem 1. *The total expected information gain for policy π over a sequence of T experiments is*

$$\mathcal{I}_T(\pi) := \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\right] \quad (7)$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(h_T|\theta,\pi) - \log p(h_T|\pi)\right] \quad (8)$$

where $p(h_T|\pi) = \mathbb{E}_{p(\theta)}[p(h_T|\theta,\pi)]$.

[1] To give an intuitive example, consider the problem of placing two breakpoints on the line $[0, 1]$ to produce the most evenly sized segments. The optimal myopic policy places its first design at $1/2$ and its second at either $1/4$ or $3/4$. This is suboptimal since the best strategy is to place the two breakpoints at $1/3$ and $2/3$.

The proof is given in Appendix A. Intuitively, $\mathcal{I}_T(\pi)$ is the expected reduction in entropy from the prior $p(\theta)$ to the *final* posterior $p(\theta|h_T)$, without considering the intermediate posteriors at all. Note here a critical change from previous BOED formulations: $\mathcal{I}_T(\pi)$ is a function of the policy, not the designs themselves, with the latter now being random variables (due to their dependence on previous outcomes) that we take an expectation over. This is actually a strict generalization of conventional BOED frameworks: static design corresponds to policy that consists of T fixed designs with no adaptivity, for which (8) coincides with $I(\xi_{1:T})$, while conventional adaptive BOED approximates π_s.

By reformulating our objective in terms of a policy, we have constructed a single end-to-end objective for adaptive, non-myopic design and which requires negligible computation at deployment time: once π is learned, it can just be directly evaluated during the experiment itself.

4. Deep Adaptive Design

Theorem 1 showed that the optimal design function $\pi^* = \arg\max_\pi \mathcal{I}_T(\pi)$ is the one which maximizes the mutual information between the unknown latent θ and the full rollout of histories produced using that policy, h_T. DAD looks to approximate π^* explicitly using a neural network, which we now refer to as the *design network* π_ϕ, with trainable parameters ϕ. This policy-based approach marks a major break from existing methods, which do not represent design decisions explicitly as a function, but instead optimize designs on the fly during the experiment.

DAD amortizes the cost of experimental design—by training the network parameters ϕ, the design network is taught to make correct design decisions across a wide range of possible experimental outcomes. This removes the cost of adaptation for the live experiment itself: during deployment the design network will select the next design nearly instantaneously with a single forward pass of the network. Further, it offers a simplification and streamlining of the sequential BOED process: it only requires the upfront end-to-end training of a single neural network and thus negates the need to set up complex *automated* inference and optimization schemes that would otherwise have to run in the background during a live experiment. A high-level summary of the DAD approach is given in Algorithm 1.

Two key technical challenges still stand in the way of realizing the potential of adaptive BOED in real time. First, whilst the unified objective $\mathcal{I}_T(\pi)$ does not require the computation of intermediate posterior distributions, it remains an intractable objective due to the presence of $p(h_T|\pi)$. To deal with this, we derive a family of lower bounds that are appropriate for the policy-based setting and use them to construct stochastic gradient training schemes for ϕ. Second, to ensure that this network can efficiently learn a mapping from histories to designs, we require an effective architecture. As we show later, the optimal policy is invariant to the order of the history, and we use this key symmetry to architect an effective design network.

Algorithm 1 Deep Adaptive Design (DAD)

Input: Prior $p(\theta)$, likelihood $p(y|\theta,\xi)$, number of steps T
Output: Design network π_ϕ
while training compute budget not exceeded **do**
  Sample $\theta_0 \sim p(\theta)$ and set $h_0 = \varnothing$
  for $t = 1, ..., T$ **do**
    Compute $\xi_t = \pi_\phi(h_{t-1})$
    Sample $y_t \sim p(y|\theta_0, \xi_t)$
    Set $h_t = \{(\xi_1, y_1), ..., (\xi_t, y_t)\}$
  end
  Compute estimate for $d\mathcal{L}_T/d\phi$ as per § 4.2
  Update ϕ using stochastic gradient ascent scheme
end
At deployment, π_ϕ is fixed, we take $\xi_t = \pi_\phi(h_{t-1})$, and each y_t is obtained by running an experiment with ξ_t.

4.1. Contrastive bounds for sequential experiments

Our high-level aim is to train π_ϕ to maximize the mutual information $\mathcal{I}_T(\pi_\phi)$. In contrast to most machine learning tasks, this objective is *doubly* intractable and cannot be directly evaluated or even estimated with a conventional Monte Carlo estimator, except in very special cases (Rainforth et al., 2018). In fact, it is extremely challenging and costly to derive *any unbiased* estimate for it or its gradients. To train π_ϕ with stochastic gradient methods, we will therefore introduce and optimize *lower bounds* on $\mathcal{I}_T(\pi_\phi)$, building on the ideas of § 2.2.

Equation (8) shows that the objective function is the expected logarithm of a ratio of two terms. The first is the likelihood of the history, $p(h_T|\theta,\pi)$, and can be directly evaluated using (5). The second term is an intractable marginal $p(h_T|\pi)$ that is different for each sample of the outer expectation and must thus be estimated separately each time.

Given a sample $\theta_0, h_T \sim p(\theta, h_T|\pi)$, we can perform this estimation by introducing L independent *contrastive* samples $\theta_{1:L} \sim p(\theta)$. We can then approximate the log-ratio in two different ways, depending on whether or not we include θ_0 in our estimate for $p(h_T|\pi)$:

$$g_L(\theta_{0:L}, h_T) = \log \frac{p(h_T|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)} \tag{9}$$

$$f_L(\theta_{0:L}, h_T) = \log \frac{p(h_T|\theta_0,\pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell,\pi)}. \tag{10}$$

These functions can both be evaluated by recomputing the likelihood of the history under each of the contrastive samples $\theta_{1:L}$. We note that g cannot exceed $\log(L+1)$, whereas f is potentially unbounded (see Appendix A for a proof).

We now show that using g to approximate the integrand leads to a *lower* bound on the overall objective $\mathcal{I}_T(\pi)$, whilst using f leads to an *upper* bound. During training, we focus on the lower bound, because it does not lead to unbounded ratio estimates and is therefore more numerically stable. We refer to this new lower bound as *sequential PCE* (sPCE).

Theorem 2 (Sequential PCE)**.** *For a design function π and a number of contrastive samples $L \geq 0$, let*

$$\mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0, h_T|\pi)p(\theta_{1:L})}\left[g_L(\theta_{0:L}, h_T)\right] \tag{11}$$

where $g_L(\theta_{0:L}, h_T)$ is as per (9), and $\theta_0, h_T \sim p(\theta, h_T|\pi)$, and $\theta_{1:L} \sim p(\theta)$ independently. Given minor technical assumptions discussed in the proof, we have[2]

$$\mathcal{L}_T(\pi, L) \uparrow \mathcal{I}_T(\pi) \text{ as } L \to \infty \tag{12}$$

at a rate $\mathcal{O}\left(L^{-1}\right)$.

The proof is presented in Appendix A. For evaluation purposes, it is helpful to pair sPCE with an upper bound, which we obtain by using f as our estimate of the integrand

$$\mathcal{U}_T(\pi, L) = \mathbb{E}_{p(\theta_0, h_T|\pi)p(\theta_{1:L})}\left[f_L(\theta_{0:L}, h_T)\right]. \tag{13}$$

We refer to this bound as sequential Nested Monte Carlo (sNMC). Theorem 4 in Appendix A shows that $\mathcal{U}_T(\pi, L)$ satisfies complementary properties to $\mathcal{L}_T(\pi, L)$. In particular, $\mathcal{L}_T(\pi, L) \leq \mathcal{I}_T(\pi) \leq \mathcal{U}_T(\pi, L)$ and both bounds become monotonically tighter as L increases, becoming exact as $L \to \infty$ at a rate $\mathcal{O}\left(1/L\right)$. We can thus directly control the trade-off between bias in our objective and the computational cost of training. Note that increasing L has no impact on the cost at deployment time. Critically, as we will see in our experiments, we tend to only need relatively modest values of L for $\mathcal{L}_T(\pi, L)$ to be an effective objective.

If using a sufficiently large L proves problematic (e.g. our available training time is strictly limited), one can further tighten these bounds for a fixed L by introducing an amortized proposal, $q(\theta; h_T)$, for the contrastive samples $\theta_{1:L}$, rather than drawing them from the prior, as in Foster et al. (2020). By appropriately adapting $\mathcal{L}_T(\pi, L)$, the proposal and the design network can then be trained simultaneously with a single unified objective, in a manner similar to a variational autoencoder (Kingma & Welling, 2014), allowing the bound itself to get tighter during training. The resulting more general class of bounds are described in detail in Appendix B and may offer further improvements for the DAD approach. We focus on training with sPCE here in the interest of simplicity of both exposition and implementation.

4.2. Gradient estimation

The design network parameters ϕ can be optimized using a stochastic optimization scheme such as Adam (Kingma & Ba, 2014). Such methods require us to compute unbiased gradient estimates of the sPCE objective (11). Throughout, we assume that the design space Ξ is continuous.

We first consider the case when the observation space $\mathcal{Y}$ is also continuous and the likelihood $p(y|\theta, \xi)$ is reparametrizable. This means that we can introduce random variables $\epsilon_{1:T} \sim p(\epsilon)$, which are independent of $\xi_{1:T}$ and $\theta_{0:L}$, such that $y_t = y(\theta_0, \xi_t, \epsilon_t)$. As we already have that $\xi_t = \pi_\phi(h_{t-1})$, we see that h_t becomes a deterministic function of h_{t-1} given ϵ_t and θ_0. Under these assumptions we can take the gradient operator inside the expectation and apply the law of the unconscious statistician to write[3]

$$\frac{d\mathcal{L}_T}{d\phi} = \mathbb{E}_{p(\theta_{0:L})p(\epsilon_{1:T})}\left[\frac{d}{d\phi} g_L(\theta_{0:L}, h_T)\right]. \tag{14}$$

We can now construct unbiased gradient estimates by sampling from $p(\theta_{0:L})p(\epsilon_{1:T})$ and evaluating, $dg_L(\theta_{0:L}, h_T)/d\phi$. This gradient can be easily computed via an automatic differentiation framework (Baydin et al., 2018; Paszke et al., 2019).

For the case of discrete observations $y \in \mathcal{Y}$, first note that given a policy π_ϕ, the only randomness in the history h_T comes from the observations $y_1, \ldots, y_T$, since the designs are computed deterministically from past histories. One approach to computing the gradient of (11) in this case is to sum over all possible histories h_T, integrating out the variables $y_{1:T}$, and take gradients with respect to ϕ to give

$$\frac{d\mathcal{L}_T}{d\phi} = \mathbb{E}\left[\sum_{h_T} \frac{d}{d\phi}\Big(p(h_T|\theta_0) g_L(\theta_{0:L}, h_T)\Big)\right], \tag{15}$$

where the expectation is over $\theta_{0:L} \sim p(\theta)$. Unbiased gradient estimates can be computed using samples from the prior. Unfortunately, this gradient estimator has a computational cost $\mathcal{O}(|\mathcal{Y}|^T)$ and is therefore only applicable when both the number of experiments T and the number of possible outcomes $|\mathcal{Y}|$ are relatively small.

To deal with the cases when it is either impractical to enumerate all possible histories, or $\mathcal{Y}$ is continuous but the likelihood $p(h_T|\theta, \pi_\phi)$ is non-reparametrizable, we propose using the score function gradient estimator, which is also known as the REINFORCE estimator (Williams, 1992). The score function gradient, is given by

$$\frac{d\mathcal{L}_T}{d\phi} = \mathbb{E}\left[\left(\log \frac{p(h_T|\theta_0, \pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi_\phi)}\right) \frac{d}{d\phi} \log p(h_T|\theta_0, \pi_\phi) - \frac{d}{d\phi} \log \sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi_\phi)\right] \tag{16}$$

[2] $x_L \uparrow x$ means that x_L is a monotonically increasing sequence in L with limit x.

[3] We use $\partial a/\partial b$ and da/db to represent the Jacobian matrices of partial and total derivatives respectively for vectors a and b.

where the expectation is over $\theta_0, h_T \sim p(\theta, h_T|\pi)$ and $\theta_{1:L} \sim p(\theta)$, and unbiased estimates may again be obtained using samples. This gradient is amenable to the wide range of existing variance reduction methods such as control variates (Tucker et al., 2017; Mohamed et al., 2020). In our experiments, however, we found the standard score function gradient to be sufficiently low variance. For complete derivations of the gradients estimators we use, see Appendix C.

4.3. Architecture

Finally, we discuss the deep learning architecture used for π_ϕ. To allow efficient and effective training, we take into account a key permutation invariance of the BOED problem as highlighted by the following result (proved in Appendix A).

Theorem 3 (Permutation invariance)**.** *Consider a permutation $\sigma \in S_k$ acting on a history h_k^1, yielding $h_k^2 = (\xi_{\sigma(1)}, y_{\sigma(1)}), ..., (\xi_{\sigma(k)}, y_{\sigma(k)})$. For all such σ, we have*

$$\mathbb{E}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^1\right] = \mathbb{E}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^2\right]$$

such that the EIG is unchanged under permutation. Further, the optimal policies starting in h_k^1 and h_k^2 are the same.

This permutation invariance is an important and well-studied property of many machine learning problems (Bloem-Reddy & Teh, 2019). The knowledge that a system exhibits permutation invariance can be exploited in neural architecture design to enable significant *weight sharing*. One common approach is pooling (Edwards & Storkey, 2016; Zaheer et al., 2017; Garnelo et al., 2018a;b). This involves summing or otherwise combining representations of multiple inputs into a single representation that is invariant to their order.

Using this idea, we represent the history h_t with a fixed dimensional representation that is formed by pooling representations of the distinct design-outcome pairs of the history

$$R(h_t) := \sum_{k=1}^{t} E_{\phi_1}(\xi_k, y_k), \tag{17}$$

where E_{ϕ_1} is a neural network *encoder* with parameters ϕ_1 to be learned. Note that this pooled representation is the same if we reorder the labels $1, ..., t$. By convention, the sum of an empty sequence is 0.

We then construct our design network to make decisions based on the pooled representation $R(h_t)$ by setting $\pi_\phi(h_t) = F_{\phi_2}(R(h_t))$, where F_{ϕ_2} is a learned *emitter* network. The trainable parameters are $\phi = \{\phi_1, \phi_2\}$. By combining simple networks in a way that is sensitive to the permutation invariance of the problem, we facilitate parameter sharing in which the network E_{ϕ_1} is re-used for each input pair and for each time step t. This results in significantly improved performance compared to networks that are forced to *learn* the relevant symmetries of the problem.

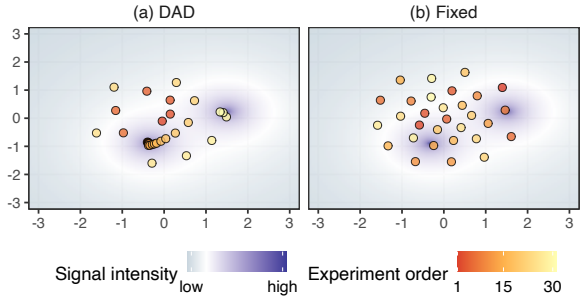


Figure 1. An example of the designs learnt by (a) the DAD network and (b) the fixed baseline for a given θ sampled from the prior.

5. Related Work

Existing approaches to sequential BOED typically follow the path outlined in § 2.1. The posterior inference performed at each stage of the conventional approach has been done using sequential Monte Carlo (Del Moral et al., 2006; Drovandi et al., 2014), population Monte Carlo (Rainforth, 2017), variational inference (Foster et al., 2019; 2020), and Laplace approximation (Lewi et al., 2009; Long et al., 2013).

The estimation of the mutual information objective at each step has been performed by nested Monte Carlo (Myung et al., 2013; Vincent & Rainforth, 2017), variational bounds (Foster et al., 2019; 2020), Laplace approximation (Lewi et al., 2009), ratio estimation (Kleinegesse et al., 2020), and hybrid methods (Senarathne et al., 2020). The optimization over designs has been performed by Bayesian optimization (Foster et al., 2019; Kleinegesse et al., 2020), interacting particle systems (Amzal et al., 2006), simulated annealing (Müller, 2005), utilizing regret bounds (Zheng et al., 2020), or bandit methods (Rainforth, 2017).

There are approaches that simultaneously estimate the mutual information and optimize it, using a single stochastic gradient procedure. Examples include perturbation analysis (Huan & Marzouk, 2014), variational lower bounds (Foster et al., 2020), or multi-level Monte Carlo (Goda et al., 2020).

Some recent work has focused specifically on models with intractable likelihoods (Hainy et al., 2016; Kleinegesse & Gutmann, 2020; Kleinegesse et al., 2020). Other work has sought to learn a non-myopic strategy focusing on specific tractable cases (Huan & Marzouk, 2016; Jiang et al., 2020).

6. Experiments

We now compare **DAD** to a number of baselines across a range of experimental design problems. We implement DAD by extending PyTorch (Paszke et al., 2019) and Pyro (Bingham et al., 2018) to provide an implementation that is abstracted from the specific problem. Code is publicly available at `https://github.com/ae-foster/dad`.

Method	Lower bound, $\mathcal{L}_{30}$	Upper bound, $\mathcal{U}_{30}$
Random	8.303 ± 0.043	8.322 ± 0.045
Fixed	8.838 ± 0.039	8.914 ± 0.038
DAD	$\mathbf{10.926 \pm 0.036}$	$\mathbf{12.382 \pm 0.095}$
Variational	8.776 ± 0.143	9.064 ± 0.187

Table 1. Upper and lower bounds on the total EIG, $\mathcal{I}_{30}(\pi)$, for the location finding experiment. Errors indicate ± 1 s.e. estimated over 256 (variational) or 2048 (others) rollouts.

As our aim is to adapt designs in *real-time*, we primarily compare to strategies that are fast at deployment time. The simplest baseline is **random** design, which selects designs uniformly at random. The **fixed** baseline completely ignores the opportunity for adaptation and uses static design to learn a fixed $\xi_1, ..., \xi_T$ before the experiment. We use the SG-BOED approach of Foster et al. (2020) with the PCE bound to optimize the fixed design $\xi_{1:T}$. We also compare to tailor-made heuristics for particular models as appropriate.

Similarly to the notion of the amortization gap in amortized inference (Cremer et al., 2018), one might initially expect to a drop in performance of DAD compared to conventional (non-amortized) BOED methods that use the traditional iterative approach of § 2.1. To assess this we also consider using the SG-BOED approach of Foster et al. (2020) in a traditional iterative manner to approximate π_s, referring to this as the **variational** baseline, noting this requires significant runtime computation. We also look at several iterative BOED baselines that are specifically tailored to the examples that we choose (Vincent & Rainforth, 2017; Kleinegesse et al., 2020). Perhaps surprisingly, we find that DAD is not only competitive compared to these non-amortized methods, but often outperforms them. This is discussed in § 7.

The first performance metric that we focus on is total EIG, $\mathcal{I}_T(\pi)$. When no direct estimate of $\mathcal{I}_T(\pi)$ is available, we estimate both the sPCE lower bound and sNMC upper bound. We also present the standard error to indicate how the performance varies between different experiment realizations (rollouts). We further consider the deployment time (i.e. the time to run the experiment itself, after pre-training); a critical metric for our aims. Full experiment details are given in Appendix D.

6.1. Location finding in 2D

Inspired by the acoustic energy attenuation model of Sheng & Hu (2005), we consider the problem of finding the locations of multiple hidden sources which each emits a signal whose intensity attenuates according to the inverse-square law. The *total intensity* is a superposition of these signals. The design problem is to choose where to make observations of the total signal to learn the locations of the sources.

We train a DAD network to perform $T = 30$ experiments with $K = 2$ sources. The designs learned by DAD are visualized in Figure 1(a). Here our network learns a complex strategy that initially explores in a spiral pattern. Once it detects a strong signal, multiple experiments are performed close together to refine knowledge of that location (note the high density of evaluations near the sources). The fixed design strategy, displayed in Figure 1(b) must choose all design locations up front, leading to an evenly dispersed strategy that cannot "hone in" on the critical areas, thus gathering less information.

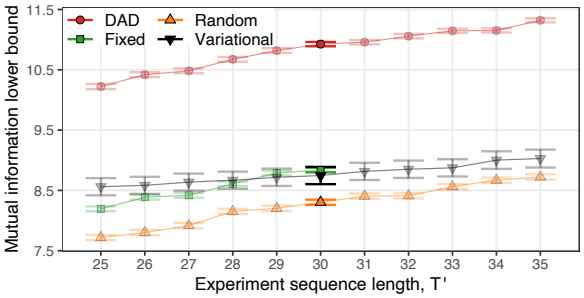


Figure 2. Generalizing sequence length for the location finding experiment. The DAD network and the fixed strategy were trained to perform $T = 30$ experiments, whilst other strategies do not require pre-training. The fixed strategy cannot be generalized to sequences longer than its training regime. We present sPCE estimates with error bars computed as in Table 1.

Table 1 reports upper and lower bounds on $\mathcal{I}_T(\pi)$ for each strategy and confirms that DAD significantly outperforms all the considered baselines. DAD is also orders of magnitude faster to deploy than the variational baseline, the other adaptive method, with DAD taking 0.0474 ± 0.0003 secs to make all 30 design decisions on a lightweight CPU, compared to 8963 secs for the variational method.

Varying the Design Horizon In practical situations the exact number of experiments to perform may be unknown. Figure 2 indicates that our DAD network that is pretrained to perform $T = 30$ experiments can generalize well to perform $T' \neq 30$ experiments at deployment time, still outperforming the baselines, indicating that DAD is robust to the length of training sequences.

Training Stability To assess the stability between different training runs, we trained 16 different DAD networks. Computing the mean and standard error of the lower bound on $\mathcal{I}_T(\pi)$ over these 16 runs gave 10.91 ± 0.014, and the matching upper bounds were 12.47 ± 0.046. We see that the variance across different training seeds is modest, indicating that DAD reaches designs of a similar quality each time. Comparing with Table 1, we see that the natural variability across rollouts (i.e. different θ) with a single DAD network tends to be larger than the variance between the average performance of different DAD networks.

Method	Deployment time (s)
Frye et al. (2016)	0.0902 ± 0.0003
Kirby (2009)	N/A
Fixed	N/A
DAD	0.0901 ± 0.0007
Badapted	25.2679 ± 0.1854

Table 2. Deployment times for Hyperbolic Temporal Discounting methods. We present the total design time for $T = 20$ questions, taking the mean and ± 1 s.e. over 10 realizations. Tests were conducted on a lightweight CPU (see Appendix D).

Method	Lower bound	Upper bound
Frye et al. (2016)	3.500 ± 0.029	3.513 ± 0.029
Kirby (2009)	1.861 ± 0.008	1.864 ± 0.009
Fixed	2.518 ± 0.007	2.524 ± 0.007
DAD	$\mathbf{5.021 \pm 0.013}$	$\mathbf{5.123 \pm 0.015}$
Badapted	4.454 ± 0.016	4.536 ± 0.018

Table 3. Final lower and upper bounds on the total information $\mathcal{I}_T(\pi)$ for the Hyperbolic Temporal Discounting experiment. The bounds are finite sample estimates of $\mathcal{L}_T(\pi, L)$ and $\mathcal{U}_T(\pi, L)$ with $L = 5000$. The errors indicate ± 1 s.e. over the sampled histories.

6.2. Hyperbolic temporal discounting

In psychology, temporal discounting is the phenomenon that the utility people attribute to a reward typically decreases the longer they have to wait to receive it (Critchfield & Kollins, 2001; Green & Myerson, 2004). For example, a participant might be willing to trade £90 today for £100 in a month's time, but not for £100 in a year. A common parametric model for temporal discounting in humans is the hyperbolic model (Mazur, 1987); we study a specific form of this model proposed by Vincent (2016).

We design a sequence of $T = 20$ experiments, each taking the form of a binary question "Would you prefer £R today, or £100 in D days?" with design $\xi = (R, D)$ that must be chosen at each stage. As real applications of this model would involve human participants, the available time to choose designs is strictly limited. We consider DAD, the aforementioned fixed design policy, and strategies that have been used specifically for experiments of this kind: Kirby (2009), a human constructed fixed set of designs; Frye et al. (2016), a problem-specific adaptive strategy; and Vincent & Rainforth (2017), a partially customized sequential BOED method, called Badapted, that uses population Monte Carlo (Cappé et al., 2004) to approximate the posterior distribution at each step and a bandit approach to optimize the EIG over possible designs.

We begin by investigating the time required to deploy each of these methods. As shown in Table 2, the non-amortized Badapted method takes the longest time, while for DAD, the total deployment time is less than 0.1 seconds—totally imperceptible to a participant.

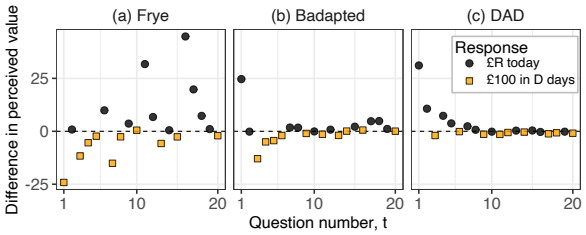


Figure 3. An example of the designs learnt by two of the problem-specific baselines and DAD. We plot the difference in perceived value of the two propositions "£R today" and "£100 in D days" for a certain participant, represented by a specific value of the latent variable θ. A difference of 0 indicates that the participant is indifferent between the two offers.

Table 3 shows the performance of each method. We see that DAD performs best, surpassing bespoke design methods that have been proposed for this problem, including Badapted which has a considerably larger computation budget. Figure 3 demonstrates how the designs learnt by DAD compare qualitatively with the two most competitive problem-specific baselines. As with Badapted, DAD designs rapidly cluster near the indifference point.

This experiment demonstrates that DAD can successfully amortize the process of experimental design in a real application setting. It outperforms some of the most successful non-amortized and highly problem-specific approaches with a fraction of the cost during the real experiment.

6.3. Death process

We conclude with an example from epidemiology (Cook et al., 2008) in which healthy individuals become infected at rate θ. The design problem is to choose observations times $\xi > 0$ at which to observe the number of infected individuals: we select $T = 4$ designs sequentially with an independent stochastic process observed at each iteration. We compare to our fixed and variational baselines, along with the adaptive SeqBED approach of Kleinegesse et al. (2020).

First, we examine the compute time required to deploy each method for a single run of the sequential experiment. The times illustrated in Table 4 show that the adaptive strategy learned by DAD can be deployed in under 0.01 seconds, many orders of magnitude faster than the non-amortized methods, with SeqBED taking hours for *one* rollout.

Next, we estimate the objective $\mathcal{I}_T(\pi)$ by averaging the information gain over simulated rollouts. The results in Table 4 reveal that DAD designs are superior to both fixed design and variational adaptive design, tending to uncover more information about the latent θ across many possible experimental trajectories. For comparison with SeqBED, we were unable to perform sufficient rollouts to obtain a high

Method	Deployment time (s)	$\mathcal{I}_T(\pi)$
Fixed	N/A	2.023 ± 0.007
DAD	$0.0051 \pm 12\%$	$\mathbf{2.113 \pm 0.008}$
Variational	$1935.0 \pm 2\%$	2.076 ± 0.034
SeqBED*	25911.0	1.590

Table 4. Total EIG $\mathcal{I}_T(\pi)$ and deployment times for the Death Process. We present the EIG ± 1 s.e. over 10,000 rollouts (fixed and DAD), 500 rollouts (variational) or *1 rollout (SeqBED). The IG can be efficiently evaluated in this case (see Appendix D). Runtimes computed as per Table 2.

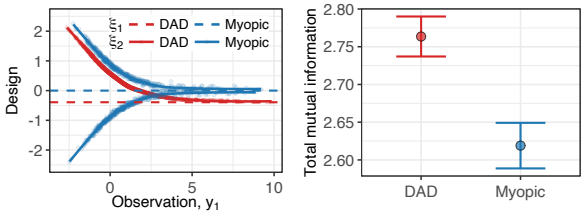


Figure 4. 1D location finding with 1 source, $T = 2$. [Left] the design function, dashed lines correspond to the first design ξ_1, which is independent of y_1. [Right] $\mathcal{I}_2(\pi)$, the total EIG ± 1 s.e.

quality estimate of $\mathcal{I}_T(\pi)$. Instead, we conducted a single rollout of each method with $\theta = 1.5$ fixed. The resulting information gains for this one rollout were: 1.590 (SeqBED), 1.719 (Variational), 1.678 (Fixed), **1.779 (DAD)**.

7. Discussion

In this paper we introduced DAD—a new method utilizing the power of deep learning to amortize the cost of sequential BOED and allow adaptive experiments to be run in real time. In all experiments DAD performed significantly better than baselines with a comparable deployment time. Further, DAD showed competitive performance against conventional BOED approaches that do not use amortization, but make costly computations at each stage of the experiment.

Surprisingly, we found DAD was often able to outperform these non-amortized approaches despite using a tiny fraction of the resources at deployment time. We suggest two reasons for this. Firstly, conventional methods must approximate the posterior $p(\theta|h_t)$ at each stage. If this approximation is poor, the resulting design optimization will yield poor results regardless of the EIG optimization approach chosen. Careful tuning of the posterior approximation could alleviate this, but would increase computational time further and it is difficult to do this in the required automated manner. DAD sidesteps this problem altogether by eliminating the need for directly approximating a posterior distribution.

Secondly, the policy learnt by DAD has the potential to be *non-myopic*: it does not choose a design that is optimal for the current experiment in isolation, but takes into account the fact that there are more experiments to be performed in the future. We can see this in practice in a simple experiment using the location finding example with one source in 1D with prior $\theta \sim N(0, 1)$ and with $T = 2$ steps. This setting is simple enough to compute the *exact* one-step optimal design via numerical integration. Figure 4 [Left] shows the design function learnt by DAD alongside the exact optimal myopic design. The optimal myopic strategy for $t = 1$ is to sample at the prior mean $\xi_1 = 0$. At time $t = 2$ the myopic strategy selects a positive or negative design with equal probability. In contrast, the policy learnt by DAD is to sample at $\xi_1 \approx -0.4$, which does not optimize the EIG for $T = 1$ in isolation, but leads to a better *overall* design strategy that focuses on searching the positive regime $\xi_2 > \xi_1$ in the second experiment. Figure 4 [Right] confirms that the policy learned by DAD achieves higher total EIG from the two step experiment than the *exact* myopic approach.

Limitations and Future Work The present form of DAD still possesses some restrictions that future work might look to address. Firstly, it requires the likelihood model to be *explicit*, i.e. that we can evaluate the density $p(y_t|\theta, \xi_t)$. Secondly, it requires the experiments to be conditionally independent given θ, i.e. $p(y_{1:T}|\theta, \xi_{1:T}) = \prod_{t=1}^{T} p(y_t|\theta, \xi_t)$, which may not be the case for, e.g. time series models. Thirdly, it requires the designs themselves, ξ_t, to be continuous to allow for gradient-based optimization. On another note, DAD's use of a policy to make design decisions establishes a critical link between experimental design and model-based reinforcement learning (Sekar et al., 2020). Though DAD is distinct in several important ways (e.g. the lack of observed rewards), investigating these links further might provide an interesting avenue for future work.

Conclusions DAD represents a new conception of adaptive experimentation that focuses on learning a design *policy* network offline, then deploying it during the live experiment to quickly make adaptive design decisions. This marks a departure from the well-worn path of myopic adaptive BOED (Sec. 2), eliminating the need to estimate intermediate posterior distributions or optimize over designs during the live experiment itself; it represents the first approach to allow adaptive BOED to be run in real-time for general problems. As such, we believe it may be beneficial to practitioners in a number of fields, from online surveys to clinical trials.

Acknowledgements

AF gratefully acknowledges funding from EPSRC grant no. EP/N509711/1. DRI is supported by EPSRC through the Modern Statistics and Statistical Machine Learning (StatML) CDT programme, grant no. EP/S023151/1. AF would like to thank Martin Jankowiak and Adam Golinski for helpful discussions about amortizing BOED.

References

Amzal, B., Bois, F. Y., Parent, E., and Robert, C. P. Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical association*, 101(474):773–785, 2006.

Angelova, J. A. On moments of sample mean and variance. *Int. J. Pure Appl. Math*, 79(1):67–85, 2012.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., and Siskind, J. M. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18, 2018.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D. Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, 2018.

Bloem-Reddy, B. and Teh, Y. W. Probabilistic symmetry and invariant neural networks. *arXiv preprint arXiv:1901.06082*, 2019.

Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. P. Population monte carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004.

Chaloner, K. and Verdinelli, I. Bayesian experimental design: A review. *Statistical Science*, pp. 273–304, 1995.

Cook, A. R., Gibson, G. J., and Gilligan, C. A. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

Cremer, C., Li, X., and Duvenaud, D. Inference suboptimality in variational autoencoders. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1078–1086. PMLR, 2018.

Critchfield, T. S. and Kollins, S. H. Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of applied behavior analysis*, 34(1): 101–122, 2001.

Del Moral, P., Doucet, A., and Jasra, A. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A sequential monte carlo algorithm to incorporate model uncertainty in bayesian sequential design. *Journal of Computational and Graphical Statistics*, 23(1):3–24, 2014.

Dushenko, S., Ambal, K., and McMichael, R. D. Sequential bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers. *Physical Review Applied*, 14(5):054036, 2020.

Edwards, H. and Storkey, A. Towards a neural statistician. *arXiv preprint arXiv:1606.02185*, 2016.

Evans, J. R. and Mathur, A. The value of online surveys. *Internet research*, 2005.

Foster, A., Jankowiak, M., Bingham, E., Horsfall, P., Teh, Y. W., Rainforth, T., and Goodman, N. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems 32*, pp. 14036–14047. Curran Associates, Inc., 2019.

Foster, A., Jankowiak, M., O'Meara, M., Teh, Y. W., and Rainforth, T. A unified stochastic gradient approach to designing bayesian-optimal experiments. volume 108 of *Proceedings of Machine Learning Research*, pp. 2959–2969, Online, 26–28 Aug 2020. PMLR.

Frye, C. C., Galizio, A., Friedel, J. E., DeHart, W. B., and Odum, A. L. Measuring delay discounting in humans using an adjusting amount task. *JoVE (Journal of Visualized Experiments)*, (107):e53584, 2016.

Garnelo, M., Rosenbaum, D., Maddison, C. J., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y. W., Rezende, D. J., and Eslami, S. Conditional neural processes. *arXiv preprint arXiv:1807.01613*, 2018a.

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S., and Teh, Y. W. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018b.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

Goda, T., Hironaka, T., and Kitade, W. Unbiased mlmc stochastic gradient-based optimization of bayesian experimental designs. *arXiv preprint arXiv:2005.08414*, 2020.

González, J., Osborne, M., and Lawrence, N. Glasses: Relieving the myopia of bayesian optimisation. In *Artificial Intelligence and Statistics*, pp. 790–799. PMLR, 2016.

Green, L. and Myerson, J. A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, 130(5):769, 2004.

Hainy, M., Drovandi, C. C., and McGree, J. M. Likelihood-free extensions for bayesian sequentially designed experiments. In Kunert, J., Müller, C. H., and Atkinson, A. C. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, pp. 153–161. Springer International Publishing, 2016.

Huan, X. and Marzouk, Y. Gradient-based stochastic optimization methods in bayesian experimental design. *International Journal for Uncertainty Quantification*, 4(6), 2014.

Huan, X. and Marzouk, Y. M. Sequential bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*, 2016.

Jiang, S., Chai, H., Gonzalez, J., and Garnett, R. Binoculars for efficient, nonmyopic sequential experimental design. In *International Conference on Machine Learning*, pp. 4794–4803. PMLR, 2020.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kingma, D. P. and Welling, M. Auto-encoding variational Bayes. In *ICLR*, 2014.

Kirby, K. N. One-year temporal stability of delay-discount rates. *Psychonomic bulletin & review*, 16(3):457–462, 2009.

Kleinegesse, S. and Gutmann, M. Bayesian experimental design for implicit models by mutual information neural estimation. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pp. 5316–5326. PMLR, 2020. URL http://proceedings.mlr.press/v119/kleinegesse20a.html.

Kleinegesse, S., Drovandi, C., and Gutmann, M. U. Sequential bayesian experimental design for implicit models via mutual information. *arXiv preprint arXiv:2003.09379*, 2020.

Kruschke, J. Doing bayesian data analysis: A tutorial with r, jags, and stan. 2014.

Lewi, J., Butera, R., and Paninski, L. Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21(3):619–687, 2009.

Lindley, D. V. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pp. 986–1005, 1956.

Long, Q., Scavino, M., Tempone, R., and Wang, S. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39, 2013.

Lyu, J., Wang, S., Balius, T. E., Singh, I., Levit, A., Moroz, Y. S., OMeara, M. J., Che, T., Algaa, E., Tolmachova, K., et al. Ultra-large library docking for discovering new chemotypes. *Nature*, 566(7743):224, 2019.

Mazur, J. E. An adjusting procedure for studying delayed reinforcement. *Commons, ML.; Mazur, JE.; Nevin, JA*, pp. 55–73, 1987.

Mohamed, S., Rosca, M., Figurnov, M., and Mnih, A. Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

Müller, P. Simulation based optimal design. *Handbook of Statistics*, 25:509–518, 2005.

Myung, J. I., Cavagnaro, D. R., and Pitt, M. A. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, 2013.

Nowozin, S. Debiasing evidence approximations: On importance-weighted autoencoders and jackknife variational inference. In *International Conference on Learning Representations*, 2018.

Pasek, J. and Krosnick, J. A. Optimizing survey questionnaire design in political science. In *The Oxford handbook of American elections and political behavior*. 2010.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

Poole, B., Ozair, S., van den Oord, A., Alemi, A., and Tucker, G. On variational bounds of mutual information. In *International Conference on Machine Learning*, pp. 5171–5180, 2019.

Rainforth, T. *Automating Inference, Learning, and Design using Probabilistic Programming*. PhD thesis, University of Oxford, 2017.

Rainforth, T., Cornish, R., Yang, H., Warrington, A., and Wood, F. On nesting monte carlo estimators. In *International Conference on Machine Learning*, pp. 4267–4276. PMLR, 2018.

Robbins, H. and Monro, S. A stochastic approximation method. *The annals of mathematical statistics*, pp. 400–407, 1951.

Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016.

Sekar, R., Rybkin, O., Daniilidis, K., Abbeel, P., Hafner, D., and Pathak, D. Planning to explore via self-supervised world models. In *International Conference on Machine Learning*, pp. 8583–8592. PMLR, 2020.

Senarathne, S., Drovandi, C., and McGree, J. A laplace-based algorithm for bayesian adaptive design. *Statistics and Computing*, 30(5):1183–1208, 2020.

Sheng, X. and Hu, Y. H. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Transactions on Signal Processing*, 2005. ISSN 1053587X. doi: 10.1109/TSP.2004.838930.

Stuhlmüller, A., Taylor, J., and Goodman, N. Learning stochastic inverses. In *Advances in neural information processing systems*, pp. 3048–3056, 2013.

Tucker, G., Mnih, A., Maddison, C. J., Lawson, D., and Sohl-Dickstein, J. Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models. *arXiv preprint arXiv:1703.07370*, 2017.

van den Oord, A., Li, Y., and Vinyals, O. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Vanlier, J., Tiemann, C. A., Hilbers, P. A., and van Riel, N. A. A Bayesian approach to targeted experiment design. *Bioinformatics*, 28(8):1136–1142, 2012.

Vincent, B. T. Hierarchical bayesian estimation and hypothesis testing for delay discounting tasks. *Behavior research methods*, 48(4):1608–1620, 2016.

Vincent, B. T. and Rainforth, T. The DARC toolbox: automated, flexible, and efficient delayed and risky choice experiments using bayesian adaptive design. 2017.

Williams, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 1992. ISSN 0885-6125. doi: 10.1007/bf00992696.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R., and Smola, A. Deep sets. *arXiv preprint arXiv:1703.06114*, 2017.

Zheng, S., Pacheco, J., and Fisher, J. A robust approach to sequential information theoretic planning. In *International Conference on Machine Learning*, pp. 5941–5949, 2018.

Zheng, S., Hayden, D., Pacheco, J., and Fisher III, J. W. Sequential bayesian experimental design with variable cost structure. *Advances in Neural Information Processing Systems*, 33, 2020.

A. Proofs

We begin by showing that $g_L(\theta_{0:L}, h_T)$ from equation (9) is bounded by $\log(L+1)$ and that $f_L(\theta_{0:L}, h_T)$ from equation (10) can potentially be unbounded. For the former

$$g_L(\theta_{0:L}, h_T) = \log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi)} \tag{18}$$

$$= \log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\theta_0, \pi) + \sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)} + \log(L+1) \tag{19}$$

$$\leq \log(1) + \log(L+1). \tag{20}$$

For the latter we have

$$f_L(\theta_{0:L}, h_T) = \log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)} \to +\infty \text{ as } \max_{1\leq\ell\leq L} p(h_T|\theta_\ell, \pi) \to 0 \text{ with } p(h_T|\theta_0, \pi) \text{ held constant.}$$

Next we present proofs for all Theorems in the main paper, with each restated for convenience.

Theorem 1. *The total expected information gain for policy π over a sequence of T experiments is*

$$\mathcal{I}_T(\pi) := \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\right] \tag{7}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(h_T|\theta, \pi) - \log p(h_T|\pi)\right] \tag{8}$$

where $p(h_T|\pi) = \mathbb{E}_{p(\theta)}[p(h_T|\theta, \pi)]$.

Proof. We begin by rewriting $I_{h_{t-1}}$ in terms of the information gain. This closely mimics the development that we presented in Section 2. By repeated appplication of Bayes Theorem we have

$$I_{h_{t-1}}(\xi_t) = \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(y_t|\theta, \xi_t)}{p(y_t|h_{t-1}, \xi_t)}\right] \tag{21}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(\theta|h_{t-1})p(y_t|\theta, \xi_t)}{p(\theta|h_{t-1})p(y_t|h_{t-1}, \xi_t)}\right] \tag{22}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log \frac{p(\theta|h_{t-1}, \xi_t, y_t)}{p(\theta|h_{t-1})}\right] \tag{23}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})}\left[-\log p(\theta|h_{t-1})\right] + \mathbb{E}_{p(y_t,\theta|\xi_t,h_{t-1})}\left[\log p(\theta|h_{t-1}, \xi_t, y_t)\right] \tag{24}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})}\left[-\log p(\theta|h_{t-1})\right] + \mathbb{E}_{p(y_t|\xi_t,h_{t-1})p(\theta|h_{t-1},\xi_t,y_t)}\left[\log p(\theta|h_{t-1}, \xi_t, y_t)\right] \tag{25}$$

$$= \mathbb{E}_{p(y_t|\xi_t,h_{t-1})}\left[\, H[\, p(\theta|h_{t-1})\,] - H[\, p(\theta|h_{t-1}, \xi_t, y_t)\,]\,\right]. \tag{26}$$

Now noting that each $I_{h_{t-1}}(\xi_t)$ is completely determined by h_{t-1} and π (in particular noting that ξ_t is deterministic given these, while θ is already marginalized out in each $I_{h_{t-1}}(\xi_t)$), we can write

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(h_T|\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\right] \tag{27}$$

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_{t-1}|\pi)}\left[I_{h_{t-1}}(\xi_t)\right] \tag{28}$$

and substituting in our earlier formulation for $I_{h_{t-1}}(\xi_t)$

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_{t-1}|\pi)}\left[\mathbb{E}_{p(y_t|\xi_t,h_{t-1})}\left[\, H[\, p(\theta|h_{t-1})\,] - H[\, p(\theta|h_{t-1}, \xi_t, y_t)\,]\,\right]\right]. \tag{29}$$

We now observe that we can write $h_t = h_{t-1} \cup \{(\xi_t, y_t)\}$, which allows us to rewrite this as

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_t|\pi)} \left[H\left[p(\theta|h_{t-1}) \right] - H\left[p(\theta|h_t) \right] \right] \tag{30}$$

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_T|\pi)} \left[H\left[p(\theta|h_{t-1}) \right] - H\left[p(\theta|h_t) \right] \right] \tag{31}$$

$$= \mathbb{E}_{p(h_T|\pi)} \left[\sum_{t=1}^{T} H\left[p(\theta|h_{t-1}) \right] - H\left[p(\theta|h_t) \right] \right] \tag{32}$$

$$= \mathbb{E}_{p(h_T|\pi)} \left[H\left[p(\theta) \right] - H\left[p(\theta|h_T) \right] \right], \tag{33}$$

where the last line follows from the fact that we have a telescopic sum. To complete the proof, we rearrange this as

$$= \mathbb{E}_{p(\theta, h_T|\pi)} \left[\log p(\theta|h_T) - \log p(\theta) \right] \tag{34}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)} \left[\log \frac{p(\theta)p(h_T|\theta,\pi)}{p(h_T|\pi)} - \log p(\theta) \right] \tag{35}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)} \left[\log p(h_T|\theta,\pi) - \log p(h_T|\pi) \right] \tag{36}$$

as required. □

Theorem 2 (Sequential PCE)**.** *For a design function π and a number of contrastive samples $L \geq 0$, let*

$$\mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0, h_T|\pi)p(\theta_{1:L})} \left[g_L(\theta_{0:L}, h_T) \right] \tag{11}$$

where $g_L(\theta_{0:L}, h_T)$ is as per (9), and $\theta_0, h_T \sim p(\theta, h_T|\pi)$, and $\theta_{1:L} \sim p(\theta)$ independently. Given minor technical assumptions discussed in the proof, we have[4]

$$\mathcal{L}_T(\pi, L) \uparrow \mathcal{I}_T(\pi) \text{ as } L \to \infty \tag{12}$$

at a rate $\mathcal{O}\left(L^{-1}\right)$.

Proof. We first show that $\mathcal{L}_T(\pi, L)$ is a lower bound on $\mathcal{I}_T(\pi)$:

$$\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0, h_T|\pi)} \left[\log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\pi)} \right] - \mathbb{E}_{p(\theta_0, h_T|\pi)} \mathbb{E}_{p(\theta_{1:L})} \left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1} \sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi)} \right] \tag{37}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \mathbb{E}_{p(\theta_{1:L})} \left[\log \frac{\frac{1}{L+1} \sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi)}{p(h_T|\pi)} \right] \tag{38}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \mathbb{E}_{p(\theta_{1:L})} \left[\log \left(\frac{1}{L+1} \sum_{\ell=0}^{L} \frac{p(\theta_\ell|h_T)}{p(\theta_\ell)} \right) \right] \tag{39}$$

now introducing the shorthand $p(\theta_{0:L}^{-\ell}) := p\left(\theta_{0:L \setminus \{\ell\}}\right) = \prod_{j=0, j\neq\ell}^{L} p(\theta_j)$,

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \mathbb{E}_{p(\theta_{1:L})} \left[\log \frac{\frac{1}{L+1} \sum_{\ell=0}^{L} p(\theta_\ell|h_T) p(\theta_{0:L}^{-\ell})}{p(\theta_{0:L})} \right]. \tag{40}$$

[4] $x_L \uparrow x$ means that x_L is a monotonically increasing sequence in L with limit x.

Now by the systemtry on term in side the log, we see that this expectation would be the same if it were instead taken over $p(\theta_i, h_T|\pi)p(\theta_{0:L}^{-i})$ for any $i \in \{0, \dots, L\}$ (with $i = 0$ giving the original form). Furthermore, the result is unchanged if we take the expectation over the mixture distribution $\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i, h_T|\pi)p(\theta_{0:L}^{-i}) = p(h_T|\pi)\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i|h_T)p(\theta_{0:L}^{-i})$ and thus we have

$$= \mathbb{E}_{p(h_T|\pi)}\mathbb{E}_{\frac{1}{L+1}\sum_{i=0}^{L} p(\theta_i|h_T)p(\theta_{0:L}^{-i})}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T)p(\theta_{0:L}^{-\ell})}{p(\theta_{0:L})}\right] \tag{41}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[\mathrm{KL}\left(\tilde{p}(\theta_{0:L}|h_T)||p(\theta_{0:L})\right)\right] \tag{42}$$

where $\tilde{p}(\theta_{0:L}|h_T) = \frac{1}{L+1}\sum_{\ell=0}^{L} p(\theta_\ell|h_T)p(\theta_{0:L}^{-\ell})$, which is indeed a distribution since

$$\int \tilde{p}(\theta_{0:L}|h_T)d\theta_{0:L} = \frac{1}{L+1}\sum_{\ell=0}^{L}\left(\int p(\theta_\ell|h_T)d\theta_\ell \cdot \int p(\theta_{0:L}^{-\ell})d\theta_{0:L}^{-\ell}\right) = 1. \tag{43}$$

Now by Gibbs' inequality the expected KL in (42) must be non-negative, establishing $\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) \geq 0$ and thus $\mathcal{I}_T(\pi) \geq \mathcal{L}_T(\pi, L)$ as required.

We next show monotonicity in L, i.e. $\mathcal{L}_T(\pi, L_2) \geq \mathcal{L}_T(\pi, L_1)$ for $L_2 \geq L_1 \geq 0$, using similar argument as above

$$\mathcal{L}_T(\pi, L_2) - \mathcal{L}_T(\pi, L_1) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1} p(h_T|\theta_i, \pi)}{\frac{1}{L_2+1}\sum_{j=0}^{L_2} p(h_T|\theta_j, \pi)}\right] \tag{44}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1}\left(p(\theta_i|h_T)/p(\theta_i)\right)}{\frac{1}{L_2+1}\sum_{j=0}^{L_2}\left(p(\theta_j|h_T)/p(\theta_j)\right)}\right] \tag{45}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1}\left(p(\theta_i|h_T)p(\theta_{0:L_1}^{-i})\right)/p(\theta_{0:L_1})}{\frac{1}{L_2+1}\sum_{j=0}^{L_2}\left(p(\theta_j|h_T)p(\theta_{0:L_2}^{-j})\right)/p(\theta_{0:L_2})}\right] \tag{46}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1} p(\theta_i|h_T)p(\theta_{0:L_2}^{-i})}{\frac{1}{L_2+1}\sum_{j=0}^{L_2} p(\theta_j|h_T)p(\theta_{0:L_2}^{-j})}\right] \tag{47}$$

$$= \mathbb{E}_{p(h_T|\pi)}\mathbb{E}_{\frac{1}{L+1}\sum_{\ell=0}^{L_1} p(\theta_\ell|h_T)p(\theta_{0:L_2}^{-\ell})}\left[\log \frac{\frac{1}{L_1+1}\sum_{i=0}^{L_1} p(\theta_i|h_T)p(\theta_{0:L_2}^{-i})}{\frac{1}{L_2+1}\sum_{j=0}^{L_2} p(\theta_j|h_T)p(\theta_{0:L_2}^{-j})}\right] \tag{48}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[\mathrm{KL}(\tilde{p}_1||\tilde{p}_2)\right] \geq 0 \tag{49}$$

where $\tilde{p}_1$ and $\tilde{p}_2$ are, respectively, the distributions in the numerator and denominator in (48). The result then again follows by Gibbs' inequality.

Next we show $\mathcal{L}_T(\pi, L) \to \mathcal{I}_T(\pi)$ as $L \to \infty$. First, note that the denominator in (11), $\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi)$, is a consistent estimator of the marginal $p(h_T|\pi)$, since $\frac{1}{L+1}p(h_T|\theta_0, \pi) \to 0$, and by the Strong Law of Large Numbers

$$\frac{1}{L+1}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi) = \frac{L}{L+1}\cdot\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi) \xrightarrow{\text{a.s.}} \mathbb{E}_{p(\theta)}\left[p(h_T|\theta, \pi)\right] = p(h_T|\pi). \tag{50}$$

Now from (38) we also have that

$$\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi)}{p(h_T|\pi)}\right] \tag{51}$$

and we have $\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi)}{p(h_T|\pi)} \to 0$ almost surely as $L \to \infty$. The minor technical assumption, which is required to

establish convergence is that there exist some $0 < \kappa_1, \kappa_2 < \infty$ such that[5]

$$\kappa_1 \leq \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)} \leq \kappa_2 \quad \forall \theta, h_T. \tag{52}$$

using this assumption, the integrand of (51) is bounded, because

$$\left|\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right| = \left|\log\left(\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right)\right| \tag{53}$$

$$\leq \max\left(\left|\log\left(\max_\ell \frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right)\right|, \left|\log\left(\min_\ell \frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right)\right|\right) \tag{54}$$

$$\leq \max\left(|\log\kappa_2|, |\log\kappa_1|\right) \tag{55}$$

$$< \infty. \tag{56}$$

Thus, the Bounded Convergence Theorem can be applied to conclude that $\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) \to 0$ as $L \to \infty$.

Finally, for the rate of convergence we apply the inequality $\log x \leq x - 1$ to (38) to get

$$\mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log\frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)}\right] \tag{57}$$

$$\leq \mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\frac{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)} - 1\right] \tag{58}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\left[\frac{\frac{1}{L+1}\left(p(h_T|\theta_0\pi) + \sum_{\ell=1}^{L}\mathbb{E}_{p(\theta_{1:L})}[p(h_T|\theta_\ell,\pi)]\right)}{p(h_T|\pi)} - 1\right] \tag{59}$$

$$= \mathbb{E}_{p(\theta_0,h_T|\pi)}\left[\frac{\frac{1}{L+1}\left(p(h_T|\theta_0\pi) + Lp(h_T|\pi)\right)}{p(h_T|\pi)} - 1\right] \tag{60}$$

$$= \frac{1}{L+1}\mathbb{E}_{p(\theta_0,h_T|\pi)}\left[\frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)} - 1\right] \tag{61}$$

$$= \frac{C}{L+1}, \tag{62}$$

where we can conclude $C < \infty$ using (52). Combining this with the our previous result showing that $\mathcal{L}_T(\pi, L)$ is a lower bound on $\mathcal{I}_T(\pi)$, we have shown that

$$0 \leq \mathcal{I}_T(\pi) - \mathcal{L}_T(\pi, L) \leq \frac{C}{L+1}. \tag{63}$$

This establishes the $\mathcal{O}(L^{-1})$ rate of convergence. □

Theorem 3 (Permutation invariance)**.** *Consider a permutation $\sigma \in S_k$ acting on a history h_k^1, yielding $h_k^2 = (\xi_{\sigma(1)}, y_{\sigma(1)}), ..., (\xi_{\sigma(k)}, y_{\sigma(k)})$. For all such σ, we have*

$$\mathbb{E}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^1\right] = \mathbb{E}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^2\right]$$

such that the EIG is unchanged under permutation. Further, the optimal policies starting in h_k^1 and h_k^2 are the same.

Technical note: In this statement, the first expectation is with respect to $p(h_T|\pi)$ for policy π and the second is with respect to $p(h_T|\pi')$, where for $t > k$ we set $\pi'(h_t) = \pi(\sigma^{-1}(h_t))$ where σ^{-1} acts on the first k labels by permutation and as the identity on other labels. This means we remove *explicit* variability under permutation caused by π, and show that no other source of variability can arise.

[5] In practice, we can actually weaken this assumption significantly if necessary by making κ_1 and κ_2 dependent on h_T and θ then assuming that the expectation $\mathbb{E}_{p(\theta_0,h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}[\log|\kappa_i(\theta_j, h_T)|]$ is finite for $i \in \{1,2\}$ and $j \in \{0,1\}$. This then permits $\kappa_1(h_T,\theta) \to 0$ and $\kappa_2(h_T,\theta) \to \infty$ for certain h_T and θ, provided that these events are zero measure under both $p(\theta, h_T|\pi)$ and $p(\theta)p(h_T|\pi)$, thereby avoiding potential issues with tail behavior in the limits of extreme values for θ.

Proof. To begin, we set up some notation. Given the partial history $h_k = h_k^1$, we complete the experiment by sampling (ξ_t, y_t) for $t = k+1, ..., T$. We denote the resulting full history as h_T^1, and define h_T^2 similarly. Next, we use Theorem 1 to rewrite the conditional objective under consideration as

$$\mathbb{E}_{p(h_T^1|\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\middle| h_k = h_k^1\right] = \mathbb{E}_{p(\theta|h_k^1)\prod_{t=k+1}^{T} p(y_t|\theta,\xi_t)}\left[\log p(h_T^1|\theta,\pi) - \log p(h_T^1|\pi)\right] \tag{64}$$

$$= \mathbb{E}_{p(\theta|h_k^1)\prod_{t=k+1}^{T} p(y_t|\theta,\xi_t)}\left[\log p(\theta|h_T^1) - \log p(\theta)\right] \tag{65}$$

$$= \mathbb{E}_{p(\theta|h_k^1)p(h_T^1|h_k^1,\theta,\pi)}\left[\log p(\theta|h_T^1) - \log p(\theta)\right]. \tag{66}$$

A central point of the proof is that the posterior distribution $p(\theta|h_t)$ is invariant to the order of the history. Indeed, we have

$$p(\theta|h_t) \propto p(\theta)\prod_{s=1}^{t} p(y_s|\theta,\xi_s) \tag{67}$$

which shows that $p(\theta|h_k^1) = p(\theta|h_k^2)$. Given a continuation of the history $(\xi_{k+1}, y_{k+1}), ..., (\xi_T, y_T)$, if we use the same continuation starting from h_k^1 and h_k^2 to give h_T^1 and h_T^2 then we have $p(\theta|h_T^1) = p(\theta|h_T^2)$. However, we need to show that the continuations $(\xi_{k+1}, y_{k+1}), ..., (\xi_T, y_T)$ are equal in distribution.

We now show that the sampling distributions of $(\xi_{k+1}, y_{k+1}), ..., (\xi_T, y_T)$ are equal starting from h_k^1 and h_k^2. We have shown that $\theta \sim p(\theta|h_k^1)$ is unchanged in distribution if we instead sample $\theta \sim p(\theta|h_k^2)$. Further, we have

$$\xi_{k+1}^1 = \pi(h_k^1) \qquad \xi_{k+1}^2 = \pi'(h_k^2) \tag{68}$$

which, by the construction of π' implies $\xi_{k+1}^1 = \xi_{k+1}^2$. Together, these results imply that the observations y_{k+1}^1 and y_{k+1}^2 are equal in distribution. Proceeding inductively, since h_{k+1}^1 and h_{k+1}^2 are equal in distribution a similar argument shows that h_{k+2}^1 and h_{k+2}^2 have the same distribution. Continuing in this way, we have that h_T^1 and h_T^2 are equal in distribution. Together, these results imply that

$$\mathbb{E}_{p(\theta|h_k^1)p(h_T^1|h_k^1,\theta,\pi)}\left[\log p(\theta|h_T^1) - \log p(\theta)\right] = \mathbb{E}_{p(\theta|h_k^2)p(h_T^2|h_k^2,\theta,\pi')}\left[\log p(\theta|h_T^2) - \log p(\theta)\right] \tag{69}$$

which conclude the first part of the proof.

To establish the permutation invariance of the optimal policy π^*, we reason by induction starting with $k = T-1$, using a dynamic programming style argument. Given h_{T-1}, the total EIG is a function of $p(\theta|h_{T-1})$ and ξ_T. Since we do not need to account for future asymmetry in the policy, we immediately have that the optimal final design ξ_T only depends on $p(\theta|h_{T-1})$, which implies that is invariant to the order of the history.

We now assume that the optimal policy is permutation invariant starting from $k+2$. Using the previous result (69), we separete out the design ξ_{k+1} and substitute π^* for both π and π' (since it is permutation invariant for the steps after $k+1$ by inductive hypothesis) to give

$$\begin{aligned}&\mathbb{E}_{p(\theta|h_k^1)p(y_{k+1}|\theta,\xi_{k+1})\prod_{t=k+2}^{T} p(y_t|\theta,\pi^*(h_{t-1}))}\left[\log p(\theta|h_T^1) - \log p(\theta)\right] \\ &= \mathbb{E}_{p(\theta|h_k^2)p(y_{k+1}|\theta,\xi_{k+1})\prod_{t=k+2}^{T} p(y_t|\theta,\pi^*(h_{t-1}))}\left[\log p(\theta|h_T^2) - \log p(\theta)\right].\end{aligned} \tag{70}$$

To extend the optimal policy to $k+1$, we consider choosing ξ_{k+1} and then following π^* thereafter. As (70) shows us, the decision problem for ξ_{k+1} is the same starting from h_k^1 and h_k^2 because the posterior distributions $p(\theta|h_k^1)$ and $p(\theta|h_k^2)$ are equal, and the optimal policy after $k+1$ does not depend on history order. This implies that the optimal choice of ξ_{k+1} is the same for h_k^1 and h_k^2. This implies that the optimal policies starting in h_k^1 and h_k^2 are the same. This completes the proof. □

Theorem 4. *For a design function π and a number of contrastive samples $L \geq 1$, let*

$$\mathcal{U}_T(\pi, L) = \mathbb{E}\left[\log \frac{p(h_T|\theta_0,\pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell,\pi)}\right] \tag{71}$$

where the expectation is over $\theta_0, h_T \sim p(\theta, h_T|\pi)$ and $\theta_{1:L} \sim p(\theta)$ independently. Then,

$$\mathcal{U}_T(\pi, L) \downarrow \mathcal{I}_T(\pi) \text{ as } L \to \infty \tag{72}$$

at a rate $\mathcal{O}(L^{-1})$.

Proof. We first show $\mathcal{U}_T(\pi, L)$ is an upper bound to $\mathcal{I}_T(\pi)$

$$\mathcal{U}_T(\pi, L) - \mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)}\right] - \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\pi)}\right] \tag{73}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)\right)\right] \tag{74}$$

now using Jensen's inequality

$$\geq \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{p(\theta_\ell)}\left[p(h_T|\theta_\ell, \pi)\right]\right)\right] \tag{75}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log\left(\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\pi)\right)\right] \tag{76}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log p(h_T|\pi) - \log p(h_T|\pi)\right] \tag{77}$$

$$= 0. \tag{78}$$

To show monotonicity in L, pick $L_2 \geq L_1 \geq 0$ and consider the difference

$$\delta := \mathcal{U}_T(\pi, L_1) - \mathcal{U}_T(\pi, L_2) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{\frac{1}{L_2}\sum_{j=1}^{L_2} p(h_T|\theta_j, \pi)}{\frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)}\right]. \tag{79}$$

Notice that we can write expression in the numerator $\frac{1}{L_2}\sum_{j=1}^{L_2} p(h_T|\theta_j, \pi) = \mathbb{E}_{J_1,\ldots,J_{L_1}}\left[\frac{1}{L_1}\sum_{k=1}^{L_1} p(h_T|\theta_{J_k}, \pi)\right]$, where the indices J_k have been uniformly drawn from $1, \ldots, L_2$. We have

$$\delta = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \mathbb{E}_{J_1,\ldots,J_{L_1}}\left[\frac{1}{L_1}\sum_{k=1}^{L_1} p(h_T|\theta_{J_k}, \pi)\right] - \log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)\right] \tag{80}$$

now applying Jensen's Inequality

$$\geq \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\mathbb{E}_{J_1,\ldots,J_{L_1}}\left[\log \frac{1}{L_1}\sum_{k=1}^{L_1} p(h_T|\theta_{J_k}, \pi)\right] - \log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)\right] \tag{81}$$

then use the fact that any L_1-subset of $\theta_1, ..., \theta_{L_2}$ has the same distribution

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L_2})}\left[\log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi) - \log \frac{1}{L_1}\sum_{i=1}^{L_1} p(h_T|\theta_i, \pi)\right] = 0 \tag{82}$$

which establishes monotonicity.

Finally, convergence is shown analogously to Theorem 2. Again we adopt the assumption (52). The Strong Law of Large Numbers gives us almost sure convergence $\log\left(\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)\right) \to \log p(h_T|\pi)$ as $L \to \infty$. Applying the Bounded Convergence Theorem, as in Theorem 2, we have

$$\lim_{L\to\infty}\left(\mathcal{U}_T(\pi, L) - \mathcal{I}_T(\pi, L)\right) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\lim_{L\to\infty} \log \frac{p(h_T|\pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)}\right] \tag{83}$$

$$= 0. \tag{84}$$

Finally, for the rate of convergence, we have

$$\mathcal{U}_T(\pi, L) - \mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{p(h_T|\pi)}{\frac{1}{L}\sum_{\ell=1}^{L} p(h_T|\theta_\ell, \pi)}\right] \tag{85}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[-\log\left(\frac{1}{L}\sum_{\ell=1}^{L}\frac{p(h_T|\theta_\ell, \pi)}{p(h_T|\pi)}\right)\right] \tag{86}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[-\log\left(1 + \frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{p(h_T|\theta_\ell, \pi)}{p(h_T|\pi)} - 1\right)\right)\right] \tag{87}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\sum_{n=1}^{\infty}(-1)^n\frac{x^n}{n}\right] \tag{88}$$

where $x = \frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{p(h_T|\theta_\ell,\pi)}{p(h_T|\pi)} - 1\right)$ and we have applied the Taylor expansion for $\log(1+x)$. We have

$$\mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[x\right] = 0 \tag{89}$$

$$\mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[x^2\right] = \frac{1}{L}\mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\left(\frac{p(h_T|\theta_\ell, \pi)}{p(h_T|\pi)} - 1\right)^2\right] \tag{90}$$

and higher order terms are $o(L^{-1})$ (Angelova, 2012; Nowozin, 2018). This shows that $\mathcal{U}_T(\pi, L) - \mathcal{I}_T(\pi) \to 0$ at a rate $\mathcal{O}(L^{-1})$. This concludes the proof. □

B. Additional bounds

In this section, we consider a more general lower bound on $\mathcal{I}_T(\pi)$ based on the ACE bound of Foster et al. (2020). We consider a parametrized proposal distribution $q(\theta; h_T)$ which can be used to approximate the posterior $p(\theta|h_T)$. One example of such a proposal would be an amortized variational approximation to the posterior that takes as input h_T and outputs a variational distribution over θ. It would be possible to share the representation $R(h_T)$ from (17) between the design network and the inference network. However, the following theorem is not limited to variational posteriors, and concerns any parametrized proposal distribution.

Theorem 5. *For a design function π, a number of contrastive samples $L \geq 1$, and a parametrized proposal $q(\theta; h_T)$, we have the sequential Adaptive Contrastive Estimation (sACE) lower bound*

$$\mathcal{I}_T(\pi) \geq \mathbb{E}_{p(\theta_0, h_T|\pi)q(\theta_{1:L}; h_T)}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}}\right] \tag{91}$$

and the sequential Variational Nested Monte Carlo (sVNMC) upper bound

$$\mathcal{I}_T(\pi) \leq \mathbb{E}_{p(\theta_0, h_T|\pi)q(\theta_{1:L}; h_T)}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L}\sum_{\ell=1}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}}\right]. \tag{92}$$

Proof. We begin by showing the sACE lower bound. The proof closely follows that of Theorem 2. We have the error term

$$\delta_{sACE} = \mathbb{E}_{p(\theta_0, h_T|\pi)}\left[\log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\pi)}\right] - \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}}\right] \tag{93}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log \frac{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{q(\theta_\ell; h_T)}}{p(h_T|\pi)}\right] \tag{94}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)}\mathbb{E}_{q(\theta_{1:L}; h_T)}\left[\log\left(\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(\theta_\ell|h_T)}{q(\theta_\ell; h_T)}\right)\right] \tag{95}$$

now introducing the shorthand $q(\theta_{0:L}^{-\ell}; h_T) := q\left(\theta_{0:L\setminus\{\ell\}}; h_T\right) = \prod_{j=0, j\neq\ell}^{L} q(\theta_j; h_T)$,

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \mathbb{E}_{q(\theta_{1:L}; h_T)} \left[\log \frac{\frac{1}{L+1} \sum_{\ell=0}^{L} p(\theta_\ell|h_T) q(\theta_{0:L}^{-\ell}; h_T)}{q(\theta_{0:L}; h_T)} \right]. \tag{96}$$

Now by the systemtry on term in side the log, we see that this expectation would be the same if it were instead taken over $p(\theta_i, h_T|\pi) q(\theta_{0:L}^{-i}; h_T)$ for any $i \in \{0, \ldots, L\}$. It is also the same if we take the expectation over $\frac{1}{L+1} \sum_{i=0}^{L} p(\theta_i, h_T|\pi) q(\theta_{0:L}^{-i}; h_T) = p(h_T|\pi) \frac{1}{L+1} \sum_{i=0}^{L} p(\theta_i|h_T) q(\theta_{0:L}^{-i}; h_T)$ and thus we have

$$= \mathbb{E}_{p(h_T|\pi)} \mathbb{E}_{\frac{1}{L+1} \sum_{i=0}^{L} p(\theta_i|h_T) q(\theta_{0:L}^{-i}; h_T)} \left[\log \frac{\frac{1}{L+1} \sum_{\ell=0}^{L} p(\theta_\ell|h_T) q(\theta_{0:L}^{-\ell}; h_T)}{q(\theta_{0:L}; h_T)} \right] \tag{97}$$

$$= \mathbb{E}_{p(h_T|\pi)} \left[\mathrm{KL} \left(\breve{q}(\theta_{0:L}; h_T) || q(\theta_{0:L}; h_T) \right) \right] \tag{98}$$

where $\breve{q}(\theta_{0:L}; h_T) = \frac{1}{L+1} \sum_{\ell=0}^{L} p(\theta_\ell|h_T) q(\theta_{0:L}^{-\ell}; h_T)$, which is indeed a distribution since

$$\int \breve{q}(\theta_{0:L}; h_T) d\theta_{0:L} = \frac{1}{L+1} \sum_{\ell=0}^{L} \left(\int p(\theta_\ell|h_T) d\theta_\ell \cdot \int q(\theta_{0:L}^{-\ell}; h_T) d\theta_{0:L}^{-\ell} \right) = 1. \tag{99}$$

Now by Gibb's inequality the expected KL in (98) must be non-negative, establishing the required lower bound.

Turning to the sVNMC bound, we use a proof that is close in spirit to Theorem 4. We have the error term

$$\delta_{sVNMC} = \mathbb{E}_{p(\theta_0, h_T|\pi)} \mathbb{E}_{q(\theta_{1:L}; h_T)} \left[\log \frac{p(h_T|\theta_0, \pi)}{\frac{1}{L} \sum_{\ell=1}^{L} \frac{p(h_T|\theta_\ell, \pi) p(\theta_\ell)}{q(\theta_\ell; h_T)}} \right] - \mathbb{E}_{p(\theta_0, h_T|\pi)} \left[\log \frac{p(h_T|\theta_0, \pi)}{p(h_T|\pi)} \right] \tag{100}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \mathbb{E}_{q(\theta_{1:L}; h_T)} \left[\log p(h_T|\pi) - \log \left(\frac{1}{L} \sum_{\ell=1}^{L} \frac{p(h_T|\theta_\ell, \pi) p(\theta_\ell)}{q(\theta_\ell; h_T)} \right) \right] \tag{101}$$

now using Jensen's inequality

$$\geq \mathbb{E}_{p(\theta_0, h_T|\pi)} \left[\log p(h_T|\pi) - \log \left(\frac{1}{L} \sum_{\ell=1}^{L} \mathbb{E}_{q(\theta_\ell; h_T)} \left[\frac{p(h_T|\theta_\ell, \pi) p(\theta_\ell)}{q(\theta_\ell; h_T)} \right] \right) \right] \tag{102}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \left[\log p(h_T|\pi) - \log \left(\frac{1}{L} \sum_{\ell=1}^{L} \mathbb{E}_{p(\theta_\ell)} \left[p(h_T|\theta_\ell, \pi) \right] \right) \right] \tag{103}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \left[\log p(h_T|\pi) - \log \left(\frac{1}{L} \sum_{\ell=1}^{L} p(h_T|\pi) \right) \right] \tag{104}$$

$$= \mathbb{E}_{p(\theta_0, h_T|\pi)} \left[\log p(h_T|\pi) - \log p(h_T|\pi) \right] \tag{105}$$

$$= 0. \tag{106}$$

This establishes the upper bound. □

C. Gradient details

C.1. Score function gradient

Recall that our sPCE objective is

$$\mathcal{L}_T(\pi_\phi, L) = \mathbb{E}_{p(\theta_{0:L})p(h_T|\theta_0,\pi_\phi)}\left[g_L(\theta_{0:L}, h_T)\right] \tag{107}$$

$$= \mathbb{E}_{p(\theta_{0:L})p(h_T|\theta_0,\pi_\phi)}\left[\log \frac{p(h_T|\theta_0,\pi_\phi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\right] \tag{108}$$

$$= \mathbb{E}_{p(\theta_{0:L})p(h_T|\theta_0,\pi_\phi)}\left[\log \frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\right] + \log(L+1) \tag{109}$$

Differentiating this gives:

$$\frac{d\mathcal{L}_T}{d\phi} = \mathbb{E}_{p(\theta_{0:L})}\left[\int \frac{d}{d\phi}\left(p(h_T|\theta_0,\pi_\phi)\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\right)dh_T\right] \tag{110}$$

$$= \mathbb{E}_{p(\theta_{0:L})}\left[\int \log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\frac{d}{d\phi}p(h_T|\theta_0,\pi_\phi) + p(h_T|\theta_0,\pi_\phi)\frac{d}{d\phi}\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}dh_T\right] \tag{111}$$

$$= \mathbb{E}_{p(\theta_{0:L})}\Bigg[\int p(h_T|\theta_0,\pi_\phi)\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\left(\frac{d}{d\phi}\log p(h_T|\theta_0,\pi_\phi)\right)dh_T \tag{112}$$

$$+ \int p(h_T|\theta_0,\pi_\phi)\left(\frac{d}{d\phi}\log p(h_T|\theta_0,\pi_\phi)\right)dh_T - \int p(h_T|\theta_0,\pi_\phi)\frac{d}{d\phi}\log\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)dh_T\Bigg] \tag{113}$$

$$= \mathbb{E}_{p(\theta_{0:L})}\mathbb{E}_{p(h_T|\theta_0,\pi_\phi)}\left[\log\frac{p(h_T|\theta_0,\pi_\phi)}{\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\left(\frac{d}{d\phi}\log p(h_T|\theta_0,\pi_\phi)\right) - \frac{d}{d\phi}\log\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)\right]. \tag{114}$$

In line (112) we used the log-trick $\frac{d}{dx}f(x) = f(x)\left(\frac{d}{dx}\log f(x)\right)$ and again in line (114) (in the reverse direction), together with the fact $\int \frac{d}{d\phi}p(h_T|\theta_0,\pi_\phi)dh_T = \frac{d}{d\phi}\int p(h_T|\theta_0,\pi_\phi)dh_T = 0$.

C.2. Expanded reparametrized gradient

For completeness, we provided a fully expanded form of the gradient in (14), computed using the chain rule. In practice, derivatives of this form are calculated automatically in PyTorch (Paszke et al., 2019).

Initially, we set up some additional notation. Suppose ξ the design is of dimension D_1 and y the observation is of dimension D_2. Then $u = (\xi, y)$ is of dimension $D_1 + D_2$. For an arbitrary scalar quantity x, we have

$$\frac{\partial x}{\partial u} = \begin{pmatrix}\frac{\partial x}{\partial \xi^{(1)}} & \cdots & \frac{\partial x}{\partial \xi^{(D_1)}} & \frac{\partial x}{\partial y^{(1)}} & \cdots & \frac{\partial x}{\partial y^{(D_2)}}\end{pmatrix} \tag{115}$$

and

$$\frac{\partial u}{\partial x} = \begin{pmatrix}\frac{\partial \xi^{(1)}}{\partial x} & \cdots & \frac{\partial \xi^{(D_1)}}{\partial x} & \sum_{d=1}^{D_1}\frac{\partial y^{(1)}}{\partial \xi^{(d)}}\frac{\partial \xi^{(d)}}{\partial x} & \cdots & \sum_{d=1}^{D_1}\frac{\partial y^{(D_2)}}{\partial \xi^{(d)}}\frac{\partial \xi^{(d)}}{\partial x}\end{pmatrix}^\top. \tag{116}$$

This notation enables us to concisely and clearly deal with both scalar and vector quantities. In general, the derivatives $\partial a/\partial b$ and da/db represent a matrix of shape $(\dim a, \dim b)$ where one or both of a, b may have dimension 1. This notation is particularly attractive because the Chain Rule for partial derivatives can be concisely expressed as follows. Suppose $a = a(b_1(c), ..., b_n(c), c)$, then the total derivative is given by

$$\frac{da}{dc} = \frac{\partial a}{\partial c} + \sum_{i=1}^{n}\frac{\partial a}{\partial b_i}\frac{db_i}{dc} \tag{117}$$

where the normal rules of matrix multiplication apply. We now apply this in the context of the function $g(\theta_{0:L}, h_T)$ which was defined in Section 4.2.

We have $g = g(\theta_{0:L}, u_1, ..., u_T)$. The Chain Rule implies that

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \frac{du_t}{d\phi}. \tag{118}$$

We also have, for $t = 1, ..., T$, that $u_t = u(\phi, h_{t-1}, \theta_0, \epsilon_t) = u(\phi, u_1, ..., u_{t-1}, \theta_0, \epsilon_t)$. This represents the dependence of ξ_t on h_{t-1} via π_ϕ, and the further dependence of y_t on θ_0 and ϵ_t. Expanding the derivatives again using the Chain Rule gives

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \left(\frac{\partial u_t}{\partial \phi} + \sum_{s=1}^{t-1} \frac{\partial u_t}{\partial u_s} \frac{du_s}{d\phi} \right). \tag{119}$$

Again, we can expand the total derivative to give

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \left(\frac{\partial u_t}{\partial \phi} + \sum_{s=1}^{t-1} \frac{\partial u_t}{\partial u_s} \left(\frac{\partial u_s}{\partial \phi} + \sum_{r=1}^{s-1} \frac{\partial u_s}{\partial u_r} \frac{du_r}{d\phi} \right) \right). \tag{120}$$

Rather than continuing in this manner, we observe that the current expansion (120) can be split up as follows

$$\frac{dg}{d\phi} = \sum_{t=1}^{T} \frac{\partial g}{\partial u_t} \frac{\partial u_t}{\partial \phi} + \sum_{1 \le s < t \le T} \frac{\partial g}{\partial u_t} \frac{\partial u_t}{\partial u_s} \frac{\partial u_s}{\partial \phi} + \sum_{1 \le r < s < t \le T} \frac{\partial g}{\partial u_t} \frac{\partial u_t}{\partial u_s} \frac{\partial u_s}{\partial u_r} \frac{du_r}{d\phi} \tag{121}$$

which shows that we have completely enumerated over all paths of length 1 and 2 through the computational graph, and the final term with a total derivative concerns paths of length 3 or more. This approach can be naturally extended to enumerate over all paths. To write this concisely, we introduce a new variable k which denotes the length of the path, and then a sum over all increasing sequences $1 \le t_1 < ... < t_k \le T$. This gives

$$\frac{dg}{d\phi} = \sum_{k=1}^{T} \left[\sum_{1 \le t_1 < ... < t_k \le T} \frac{\partial g}{\partial u_{t_k}} \frac{\partial u_{t_k}}{\partial u_{t_{k-1}}} ... \frac{\partial u_{t_2}}{\partial u_{t_1}} \frac{\partial u_{t_1}}{\partial \phi} \right]. \tag{122}$$

This can be written concisely as

$$\frac{dg}{d\phi} = \sum_{\substack{k \in \{1,...,T\} \\ 1 \le t_1 < ... < t_k \le T}} \frac{\partial g}{\partial u_{t_k}} \left(\prod_{j=1}^{k-1} \frac{\partial u_{t_{j+1}}}{\partial u_{t_j}} \right) \frac{\partial u_{t_1}}{\partial \phi} \tag{123}$$

where the product is interpretted in the order given in (122) for the matrix multiplication to operate correctly, and an empty product is equal to the identity.

D. Experiment details

Our experiments were implemented using PyTorch (Paszke et al., 2019) and Pyro (Bingham et al., 2018). An open-source implementation of DAD, including code for reproducing each experiment, is available at `https://github.com/ae-foster/dad`. Full details on running the code are given in the `README.md` file.

D.1. Location Finding

In this experiment we have K hidden objects or *sources* in $\mathbb{R}^d$, $d \in \{1, 2, 3\}$ and aim to learn their locations, $\theta = \{\theta_k\}_{k=1}^K$. The number of sources, K, is assumed to be known. Each of the sources emits a signal with intensity obeying the inverse-square law. In other words, if a source is located at θ_k and we perform a measurement at a point ξ, the signal strength will be proportional to $\frac{1}{\|\theta_k - \xi\|^2}$.

Since there are multiple sources, we consider the total intensity at location ξ, which is a superposition of the individual ones

$$\mu(\theta, \xi) = b + \sum_{k=1}^{K} \frac{\alpha_k}{m + \|\theta_k - \xi\|^2}, \tag{124}$$

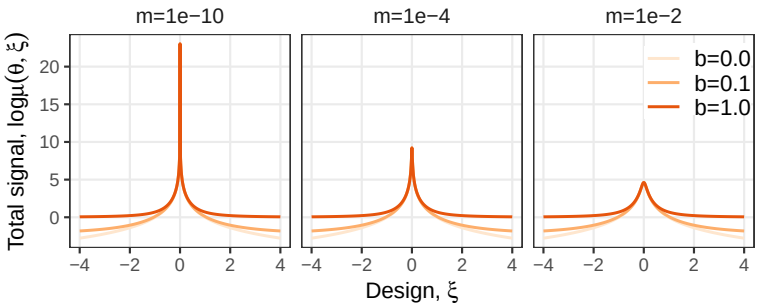


Figure 5. Log-total intensity

where α_k can be known constants or random variables, $b, m > 0$ are constants controlling background and maximum signal, respectively. Figure 5 shows the effect b and m have on log total signal strength.

We place a standard normal prior on each of the location parameters θ_k and we observe the log total intensity with some Gaussian noise. We therefore have the following prior and likelihood:

$$\theta_k \overset{\text{i.i.d.}}{\sim} N(0_d, I_d), \;\; \log y \mid \theta, \xi \sim N(\log \mu(\theta, \xi), \sigma). \tag{125}$$

The model hyperparameters used in our experiments can be found in the table below.

Parameter	Value
Number of sources, K	2
Base signal, b	10^{-1}
Max signal, m	10^{-4}
α_1, α_2	1
Signal noise, σ	0.5

We trained a DAD network to amortize experimental design for this problem, using the neural architecture outlined in Section 4.3. Both the encoder and the decoder are simple feed-forward neural networks with a single hidden layer; details in the following table. For the encoder

Layer	Description	Dimension	Activation
Input	ξ, y	3	-
H1	Fully connected	256	ReLU
Output	Fully connected	16	-

and for the emitter

Layer	Description	Dimension	Activation
Input	$R(h_t)$	16	-
H1	Fully connected	2	-
Output	ξ	2	-

Since the likelihood is reparametrizable, we use (14) to calculate approximate gradients. We optimized the network using Adam (Kingma & Ba, 2014) with exponential learning rate annealing with parameter γ. Full details are given in the following table.

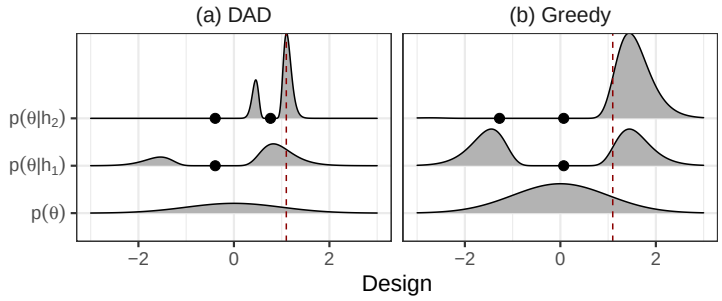


Figure 6. Posterior distributions of the location finding example with $K = 1$ source $\mathbb{R}$.

Parameter	Value
Inner samples, L	2000
Outer samples	2000
Initial learning rate	5×10^{-5}
Betas	(0.8, 0.998)
γ	0.98
Gradient steps	50000
Annealing frequency	1000

We used a greater number of inner and outer samples for a more accurate estimate of $\mathcal{I}_T(\pi)$ for evaluation when computing the presented values in Table 1 and in our Training Stability ablation, specifically $L = 5 \times 10^5$ inner samples, and 256 (variational) or 2048 (other methods) outer samples.

Deployment times Deployment speed tests were performed on a CPU-only machine witht the following specifications:

Memory	16 GB 2133 MHz LPDDR3
Processor	2.8 GHz Quad-Core Intel Core i7
Operating System	MacOS BigSur v.11.2.3

We took the mean and ± 1 s.e. over 10 realizations. Deployment times for all methods are given in the following table

Method	Deployment time (s)
Random	0.0026 ± 0.0001
Fixed	0.0018 ± 0.0001
DAD	0.0474 ± 0.0003
Variational	8963.2 ± 42.2

Discussion details In the discussion, we used a simpler form of the same model with $K = 1$ source and $\theta \in \mathbb{R}, \xi \in \mathbb{R}$. In this simplified setting, we can calculate the true optimal myopic (greedy) baseline using numerical integration. We evaluate equation (1) using line integrals as follows

$$I_t(\xi) = \int p(\theta|h_{t-1}) \mathbb{E}_{p(y|\theta)} \left[\log \frac{p(y|\theta)}{\int p(\theta'|h_{t-1}) p(y|\theta') d\theta'} \right] d\theta \tag{126}$$

$$= \int p(\theta|h_{t-1}) \mathbb{E}_{p(y|\theta)} \left[\log \int p(\theta'|h_{t-1}) p(y|\theta') d\theta' \right] d\theta + C \tag{127}$$

where $C = -H(p(y|\theta))$ is the entropy of a Gaussian, location independent and therefore constant with respect to ξ. We calculate (127) for a range of designs, $\xi \in \Xi_{\text{grid}}$, and select the optimal design $\xi^* = \arg\max_{\Xi_{\text{grid}}} I_t(\xi)$. The integrals themselves are also calculated using numerical integration on a grid, Θ_{grid}, and use sampling to calculate the inner expectation; further details can in the table below.

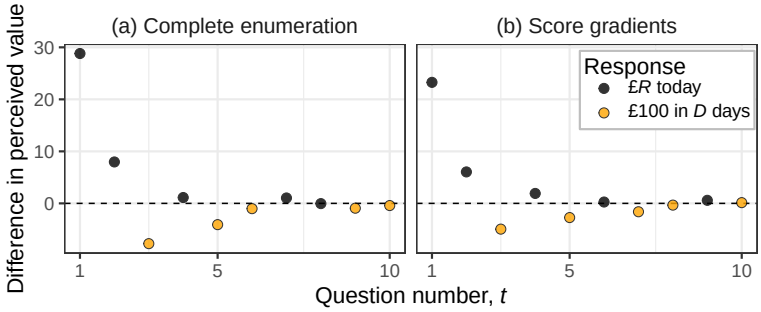


Figure 7. Comparison of two gradient methods for the hyperbolic temporal discounting model with $T = 10$ experiments.

Parameter	Value
Design grid, Ξ_{grid}	300 equally spaced from -3 to 3
θ grid, Θ_{grid}	600 equally spaced from -4 to 4
y samples for inner expectation	400

It is important to emphasize that even in this simple one-dimensional setting evaluating the myopic strategy is extremely costly and may require more sophisticated numerical integration techniques (e.g. quadrature) as posteriors become more peaked. Furthermore, as Figure 6 indicates, the resulting posteriors are complex and multi-modal even in 1D. This multi-modality may also be a reason why the variational method does not work well in this example.

D.2. Hyperbolic temporal discounting

We consider a hyperbolic temporal discounting model (Mazur, 1987; Vincent, 2016; Vincent & Rainforth, 2017) in which a participant's behaviour is characterized by the latent variables $\theta = (k, \alpha)$ with prior distributions

$$\log k \sim N(-4.25, 1.5) \qquad \alpha \sim \text{HalfNormal}(0, 2) \tag{128}$$

where the HalfNormal distribution is a Normal distribution truncated at 0. For given k, α, the value of the two propositions "£R today" and "£100 in D days" with design $\xi = (R, D)$ are given by

$$V_0 = R, \qquad V_1 = \frac{100}{1 + kD}. \tag{129}$$

The probability of the participant selecting the second option, V_1, rather than V_0 is then modelled as

$$p(y = 1|k, \alpha, R, D) = \epsilon + (1 - 2\epsilon)\Phi\left(\frac{V_1 - V_0}{\alpha}\right) \tag{130}$$

where Φ is the c.d.f. of the standard Normal distribution, i.e.

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}z^2\right) \tag{131}$$

and we fix $\epsilon = 0.01$. We considered the iterated version of this experiment, modelling $T = 20$ experiments with each sampled setting for the latents k, α.

We began by training a DAD network to amortize experimental design for this problem. The design parameters R, D have the constraints $D > 0$ and $0 < R < 100$. We represented R, D in an unconstrained space ξ_d, ξ_r and transformed them using the maps

$$D = \exp(\xi_d) \qquad R = 100 \text{ sigmoid}(\xi_r) \tag{132}$$

We used the neural architecture outlined in Section 4.3. For the encoder E_{ϕ_1} we used the following network with two hidden layers

Layer	Description	Dimension	Activation
Design input	ξ_d, ξ_r	2	-
H1	Fully connected	256	Softplus
H2	Fully connected	256	Softplus
H3	Fully connected	16	-
H3'	Fully connected	16	-
Output	$y \odot H3 + (1 - y) \odot H3'$	16	-

The emitter network F_{ϕ_2} similarly used two hidden layers as follows

Layer	Description	Dimension	Activation
Input	$R(h_t)$	16	-
H1	Fully connected	256	Softplus
H2	Fully connected	256	Softplus
Output	ξ_d, ξ_r	2	-

Since the number of experiments we perform is relatively large ($T = 20$), we constructed a score function gradient estimator of (16) (see also § C.1 for details) and optimized this network with Adam (Kingma & Ba, 2014).We used exponential learning rate annealing with parameter γ. Full details are given in the following table.

Parameter	Value
Inner samples, L	500
Outer samples	500
Initial learning rate	10^{-4}
Betas	(0.9, 0.999)
γ	0.96
Gradient steps	100000
Annealing frequency	1000

For the fixed baseline, we used the same optimization settings, except we set the initial learning rate to 10^{-1}. We trained the DAD and fixed methods on a machine with 8 Intel(R) Xeon(R) CPU E5-2637 v4 @ 3.50GHz CPUs, one GeForce GTX 1080 Ti GPU, 126 GiB memory running Fedora 32. Note this is *not* the machine used to conudct speed tests. For the Badapted baseline of Vincent & Rainforth (2017), we used the public code provided at `https://github.com/drbenvincent/badapted`. We used 50 PMC steps with 100 particles. For the baselines of Frye et al. (2016) and Kirby (2009), we used the public code provided at `https://github.com/drbenvincent/darc-experiments-matlab/tree/master/darc-experiments`, which we reimplemented in Python. These methods do not involve a pre-training step, except that we did not include time to compute the first design ξ_1 within the speed test, as this can be computed before the start of the experiment.

To implement the deployment speed tests fairly, we ran each method on a lightweight CPU-only machine, which more closely mimics the computer architecture that we might expect to deploy methods such as DAD on. The specifications of the machine we used are described below

Memory	7.7GiB
Processor	Intel Core M-5Y10c CPU @ 0.80GHz × 4
Operating System	Ubuntu 16.04 LTS

The values in Table 2 show the mean and standard error of the times observed from 10 independent runs on a idle system. To make the final evaluation for each method in Table 3, we computed the sPCE and sNMC bounds using $L = 5000$ inner samples and 10000 outer samples of the outer expectation. We present the mean and standard error from the outer expectation over 10000 rollouts.

D.2.1. Ablation: total enumeration

We compare the two methods for estimating gradients for the case of discrete observations: total enumeration of histories (Equation 15) and score function gradient estimator (Equation 16). To this end we train DAD networks to perform $T = 10$ experiments, which gives rise to a total of $2^{10} = 1024$ possible histories.

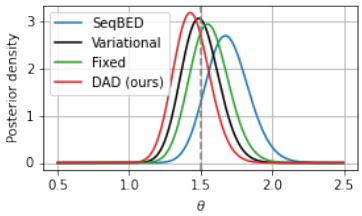


Figure 8. Comparison of posteriors obtained from a single rollout of the Death Process, used to compute the information gains quoted in Section 6.3. The dashed line indicates the true value $\theta = 1.5$ used to simulate responses.

Find that the two methods perform the same, both quantitatively and qualitatively. Table 5 reports the estimated upper and lower bounds on the mutual information objective, indicating statistically equal performance of the two methods (mean estimates are within 2 standard errors of each other). Figure 7 demonstrates the qualitative similarity in the designs learnt by the two networks.

	Lower bound, $\mathcal{L}_{10}$	Upper bound, $\mathcal{U}_{10}$
Complete enumeration	4.068 ± 0.0124	4.090 ± 0.0126
Score function gradient	4.037 ± 0.0126	4.058 ± 0.0128

Table 5. Final lower and upper bounds on the total information $\mathcal{I}_{10}(\pi)$ for the Hyperbolic Temporal Discounting experiment with $T = 10$ experiments and different gradient estimation schemes (see § 4.2 and § C.1 for details). The bounds are finite sample estimates of $\mathcal{L}_{10}(\pi, L)$ and $\mathcal{U}_{10}(\pi, L)$ with $L = 5000$. The errors indicate ± 1 s.e. over the sampled histories.

D.3. Death process

For the Death Process model (Cook et al., 2008), we use the settings that were described by Kleinegesse et al. (2020). Specifically, we use a truncated Normal prior for the infection rate

$$\theta \sim \text{TruncatedNormal}(\mu = 1, \sigma = 1, \text{min} = 0, \text{max} = \infty). \tag{133}$$

The likelihood is then given by

$$\eta = 1 - \exp(-\xi\theta) \qquad y|\theta, \xi \sim \text{Binomial}(N, \eta) \tag{134}$$

where we set $N = 50$. We consider a sequential version of this experiment as in Kleinegesse et al. (2020), with $T = 4$ and in which an independent stochastic process is observed at each step, meaning there are no constraints relating $\xi_1, ..., \xi_4$ other than the natural constraint $\xi_t > 0$.

We began by training a DAD network to perform experimental design for this problem. We used the neural architecture outlined in Section 4.3. For the encoder E_{ϕ_1} we used the following network with two hidden layers

Layer	Description	Dimension	Activation
Input	ξ, y	2	-
H1	Fully connected	128	Softplus
H2	Fully connected	128	Softplus
Output	Fully connected	16	-

The emitter network F_{ϕ_2} similarly used two hidden layers as follows

Layer	Description	Dimension	Activation
Input	$R(h_t)$	16	-
H1	Fully connected	128	Softplus
H2	Fully connected	128	Softplus
Output	ξ	1	Softplus

Although the number of experiments we perform is relatively small ($T = 4$), we could not use complete enumeration due to the prohibitively large size of the outcome space ($|\mathcal{Y}| = 51$). Hence, we constructed a score function gradient estimator of (16) (see also § C.1 for details) and optimized the DAD network with Adam (Kingma & Ba, 2014).We used exponential learning rate annealing with parameter γ. Full details are given in the following table.

Parameter	Value
Inner samples, L	500
Outer samples	500
Initial learning rate	0.001
Betas	$(0.9, 0.999)$
γ	0.96
Gradient steps	100000
Annealing frequency	1000

For the fixed baseline, we used the same optimization settings, except we set the initial learning rate to 10^{-1} and we set $\gamma = 0.85$. We trained the DAD and fixed methods using the same machine as used for training in Section D.2. For the variational baseline, we used a truncated Normal variational family to approximate the posterior at each step. We used SGD with momentum to optimize the design at each step, and to optimize the variational approximation to the posterior at each step. We used exponential learning rate annealing with paramter γ. The settings used were

Parameter	Value
Design inner samples	250
Design outer samples	250
Design initial learning rate	10^{-2}
Design γ	0.9
Design gradient steps	5000
Inference initial learning rate	10^{-3}
Inference γ	0.2
Inference gradient steps	5000
Momentum	0.1
Annealing frequency	1000

For the SeqBED baseline, we used the code publicly available at `https://github.com/stevenkleinegesse/seqbed`. The speed tests except for SeqBED were implemented as in Section D.2. For SeqBED and the variational method, we did not include the time to compute the first design as deployment time, as this can be computed before the start of the experiment. Due to its long-running nature, we implemented the speed test for SeqBED using a more powerful machine with 40 Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz processors and 189GiB memory. Therefore, the timing value for SeqBED given in Table 4 represents a significant *under-estimate* of the expected computational time required to deploy this method. However, we note that SeqBED can be applied to a broader class of implicit likelihood models.

For evaluation of $\mathcal{I}_4(\pi)$ in the Death Process, it is possible to compute the information gain $H[p(\theta)] - H[p(\theta|h_T)]$ to high accuracy using numerical integration. We then took the expectation of the information gain over rollouts, see Table 4 for the exact number of rollouts used. This gives us an estimate

$$\mathcal{I}_4(\pi) = \mathbb{E}_{p(h_T|\pi)}\left[H[p(\theta)] - H[p(\theta|h_T)]\right] \tag{135}$$

which is shown to be a valid form for the total EIG in Section A.

For a comparison with SeqBED which is too slow to use this evaluation, we instead performed one rollout of each of our methods using a fixed value $\theta = 1.5$. This is close in spririt to the evaluation used in Kleinegesse et al. (2020). Figure 8 shows the posterior distributions obtained from this rollout. The information gains were then computed using the aforementioned numerical integration and are quoted in Section 6.3. We observe that, visually, the posterior distributions are similar, and cluster near to the true value of θ.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design
Publication Status	Published
Publication Details	Adam Foster, Desi R Ivanova, Ilyas Malik, Tom Rainforth (2021). Deep Adaptive Design: Amortizing Sequential Bayesian Experimental Design. Proceedings of the 38th International Conference on Machine Learning, PMLR 139.

Student Confirmation

Student Name:	Adam Foster		
Contribution to the Paper	Co-first author. Led development of the methodology, theory, and writing of the paper. Performed experiments in Sec 6.2 and 6.3.		
Signature		Date	25/10/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Dr Tom Rainforth			
Supervisor comments I verify Adam's above account			
Signature		Date	26/10/21

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 6

Implicit Deep Adaptive Design: Policy-Based Experimental Design without Likelihoods

This paper has been accepted for publication at the Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS 2021).

Implicit Deep Adaptive Design: Policy–Based Experimental Design without Likelihoods

Desi R. Ivanova[†] **Adam Foster**[†] **Steven Kleinegesse**[‡] **Michael U. Gutmann**[‡] **Tom Rainforth**[†]
[†]Department of Statistics, University of Oxford
[‡] School of Informatics, University of Edinburgh
desi.ivanova@stats.ox.ac.uk

Abstract

We introduce implicit Deep Adaptive Design (iDAD), a new method for performing adaptive experiments in *real-time* with *implicit* models. iDAD amortizes the cost of Bayesian optimal experimental design (BOED) by learning a design *policy network* upfront, which can then be deployed quickly at the time of the experiment. The iDAD network can be trained on any model which simulates differentiable samples, unlike previous design policy work that requires a closed form likelihood and conditionally independent experiments. At deployment, iDAD allows design decisions to be made in milliseconds, in contrast to traditional BOED approaches that require heavy computation during the experiment itself. We illustrate the applicability of iDAD on a number of experiments, and show that it provides a fast and effective mechanism for performing adaptive design with implicit models.

1 Introduction

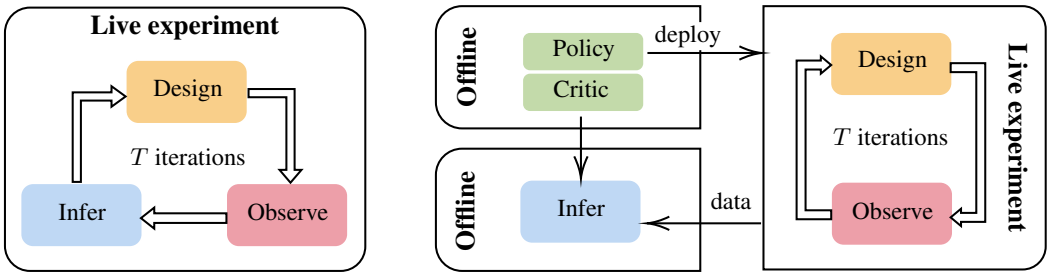
Designing experiments to maximize the information gathered about an underlying process is a key challenge in science and engineering. Most such experiments are naturally *adaptive*—we can design later iterations on the basis of data already collected, refining our understanding of the process with each step [36, 45, 51]. For example, suppose that a chemical contaminant has accidentally been released and is rapidly spreading; we need to quickly discover its unknown source. To this end, we measure the contaminant concentration level at locations $\xi_1, \dots, \xi_T$ (our experimental designs), obtaining observations $y_1, \dots, y_T$. Provided we can perform the necessary computations sufficiently quickly, we can design each ξ_t using data from steps $1, \dots, t-1$ to narrow in on the source.

Bayesian optimal experimental design (BOED) [7, 32] is a principled model-based framework for choosing designs optimally; it has been successfully adopted in a diverse range of scientific fields [52, 58, 60]. In BOED, the unknown quantity of interest (e.g. contaminant location) is encapsulated by a parameter θ, and our initial information about it by a prior $p(\theta)$. A simulator, or likelihood, model $y|\theta, \xi$ describes the relationship between θ, our controllable design ξ, and the experimental outcome y. To select designs *optimally*, the guiding principle is *information maximization*—we select the design that maximizes the expected (Shannon) information gained about θ from the data y, or, equivalently, that maximizes the mutual information between θ and y.

This naturally extends to adaptive settings by considering the *conditional* expected information gain given previously collected data. The traditional approach, depicted in Figure 1a, is to fit a posterior $p(\theta|\xi_{1:t-1}, y_{1:t-1})$ after each iteration, and then select ξ_t in a myopic fashion using the one-step mutual information (see, e.g., [51] for a review). Unfortunately, this approach necessitates significant computation at each t and does not lend itself to selecting optimal designs quickly and adaptively.

Recently, Foster et al. [17] proposed an exciting alternative approach, called Deep Adaptive Design (DAD), that is based on learning design *policies*. DAD provides a way to avoid significant computation

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.



(a) Traditional BOED: costly computations (design optimisation and parameter inference) are required at each iteration.

(b) Policy-based BOED using iDAD: a design policy and critic are learnt before the live experiment. The policy enables quick and adaptive experiments, the critic assists likelihood-free inference.

Figure 1: Overview of adaptive BOED approaches applicable to implicit models.

at deployment-time by, prior to the experiment itself, learning a design policy network that takes past design-outcome pairs and near–instantaneously returns the design for the next stage of the experiment. The required training is done using simulated experimental histories, without the need to estimate any posterior or marginal distributions. DAD further only needs a single policy network to be trained for multiple experiments, further allowing for *amortization* of the adaptive design process. Unfortunately, DAD requires conditionally independent experiments and only works for the restricted class of models that have an explicit likelihood model we can simulate from, evaluate the density of, and calculate derivatives for, substantially reducing its applicability.

To address this shortfall, we instead consider a far more general class of models where we require only the ability to simulate $y|\theta, \xi$ and compute the derivative $\partial y/\partial \xi$, e.g. via automatic differentiation [5]. Such models are ubiquitous in scientific modelling and include differentiable *implicit models* [19], for which the likelihood density $p(y|\theta, \xi)$ is intractable. Examples include mixed effects models [15, 18], various models from chemistry and epidemiology [1], the Lotka Volterra model used in ecology [19], and models specified via stochastic differential equations (such as the SIR model [10]).

To perform rapid adaptive experimentation with this large class of models, we introduce *implicit Deep Adaptive Design* (iDAD), a method for learning adaptive design policy networks using only simulated outcomes (see Figure 1b). To achieve this, we introduce likelihood-free lower bounds on the total information gained from a sequence of experiments, which iDAD utilizes to learn a deep policy network. This policy network amortizes the cost of experimental design for implicit models and can be run in milliseconds at deployment-time. To train it, we show how the InfoNCE [57] and NWJ [37] bounds, popularized in representation learning, can be applied to the policy-based experimental design setting. The optimization of both of these bounds involves simultaneously learning an auxiliary *critic network*, bringing an important added benefit: it can be used to perform likelihood-free posterior inference of the parameters given the data acquired from the experiment.

We also relax DAD's requirement for experiments to be conditionally independent, allowing its application in complex settings like time series data, and, through innovative architecture adaptations, also provide improvements in the conditionally independent setting as well. This further expands the model space for policy-based BOED, and leads to additional performance improvements.

Critically, iDAD forms the first method in the literature that can practically perform real-time adaptive BOED with implicit models: previous approaches are either not fast enough to run in real-time for non-trivial models, or require explicit likelihood models. We illustrate the applicability of iDAD on a range of experimental design problems, highlighting its benefits over existing baselines, even finding that it often outperforms costly non-amortized approaches. Code for iDAD is publicly available at `https://github.com/desi-ivanova/idad`.

2 Background

The BOED framework [32] begins by specifying a Bayesian model of the experimental process, consisting of a prior on the unknown parameters $p(\theta)$, a set of controllable designs ξ, and a data generating process that depends on them $y|\theta, \xi$; as usual in BOED, we assume that $p(\theta)$ does not depend on ξ. In this paper, we consider the situation where $y|\theta, \xi$ is specified *implicitly*. This means

that it is defined by a deterministic transformation, $f(\varepsilon; \theta, \xi)$, of a base (or noise) random variable, ε, that is independent of the parameters and the design; e.g., $\varepsilon \sim \mathcal{N}(\varepsilon; 0, I)$. The function f is itself often not known explicitly in closed form, but is implemented as a stochastic computer program (i.e. simulator) with input (θ, ξ) and ε corresponding to the draws from the underlying random number generator (or equivalently the random seed). Regardless, the resulting induced likelihood density $p(y|\theta, \xi)$ is still generally intractable, but sampling $y|\theta, \xi$ is possible.

Having acquired a design-outcome pair (ξ, y), we can quantify the amount of information we have gained about θ by calculating the reduction in entropy from the prior to the posterior. We can further assess the quality of a design ξ before acquiring y, by computing the expected reduction in entropy with respect to the marginal distribution of the outcome, $p(y|\xi) = \mathbb{E}_{p(\theta)}[p(y|\theta, \xi)]$. The resulting quantity, called the *expected information gain* (EIG), is of central interest in BOED and is defined as

$$I(\xi) \coloneqq \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(\theta|\xi, y)}{p(\theta)}\right] = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{p(y|\theta, \xi)}{p(y|\xi)}\right]. \tag{1}$$

Note that $I(\xi)$ is equivalent to the mutual information (MI) between the parameters θ and data y when performing experiment ξ. The optimal ξ is then the one that maximises the EIG, i.e. $\xi^* = \arg\max_\xi I(\xi)$. Performing this optimization is a major computational challenge since the information objective is doubly intractable [46]. For implicit models, the cost becomes even greater as the likelihood is also not available in closed form, so estimating it, along with the marginal likelihood $p(y|\xi)$, is already itself a major computational problem [11, 20, 33, 54].

Jointly optimizing the design variables for all undertaken experiments at the same time using (1) is called *static* experimental design. In practice, however, we are often more interested in performing multiple experiments *adaptively* in a sequence $\xi_1, \ldots, \xi_T$, so that the choice of each ξ_t can be guided by past experiments, namely the corresponding *history* $h_{t-1} \coloneqq \{(\xi_i, y_i)\}_{i=1:t-1}$. The typical approach in such settings is to sequentially perform (approximate) posterior inference for $\theta|h_{t-1}$, followed by a one-step look ahead (myopic) BOED optimization that conditions on the observed history. In other words, to determine the designs $\xi_1, \ldots, \xi_T$, we sequentially optimize the objectives

$$I_{h_{t-1}}(\xi_t) \coloneqq \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi,h_{t-1})}\left[\log \frac{p(y_t|\theta, \xi, h_{t-1})}{p(y_t|\xi, h_{t-1})}\right], \quad t = 1, \ldots, T. \tag{2}$$

However, such approaches incur significant computational cost during the experiment itself, particularly for implicit models [16, 21, 30]. This has critical consequences: in most cases they cannot be run in real-time, undermining one's ability to use them in practice.

2.1 Policy-based adaptive design with likelihoods

For tractable likelihood models, Foster et al. [17] proposed a new framework, called Deep Adaptive Design (DAD), for adaptive experimental design that avoids expensive computations during the experiment. To achieve this, they introduce a parameterized deterministic design function, or policy, π_ϕ that takes the history h_{t-1} as input and returns the design $\xi_t = \pi_\phi(h_{t-1})$ to be used for the next experiment as output. This set-up allows them to consider the objective

$$\mathcal{I}_T(\pi_\phi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi_\phi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\right], \quad \xi_t = \pi_\phi(h_{t-1}), \tag{3}$$

which crucially depends on the policy π rather than the individual design ξ_t. Learning a policy up-front, rather than designs, is what allows adaptive experiments to be performed in real-time.

Under the assumption that y_t is independent of h_{t-1} conditional on the parameters θ and the design ξ_t, i.e. $p(y_t|\theta, \xi_t, h_{t-1}) = p(y_t|\theta, \xi_t)$, Foster et al. [17] showed that the objective can be simplified to

$$\mathcal{I}_T(\pi_\phi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi_\phi)}\left[\log \frac{p(h_T|\theta, \pi_\phi)}{p(h_T|\pi_\phi)}\right], \quad p(h_T|\theta, \pi_\phi) = \prod_{t=1}^{T} p(y_t|\theta, \xi_t). \tag{4}$$

To deal with the marginal $p(h_T|\pi_\phi)$ in the denominator, they then derived several optimizable lower bounds on $\mathcal{I}_T(\pi_\phi)$, such as the sequential Prior Contrastive Estimation (sPCE) bound

$$\mathcal{L}_T^{\text{sPCE}}(\pi_\phi, L) = \mathbb{E}_{p(\theta_0)p(h_T|\theta,\pi_\phi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0, \pi_\phi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell, \pi_\phi)}\right] \leq \mathcal{I}_T(\pi_\phi) \;\; \forall L \geq 1. \tag{5}$$

The parameters of the policy ϕ, which takes the form of a deep neural network, are now learned prior to the experiment(s) using stochastic gradient ascent on this bound with simulated experimental histories. Design decisions can then be made using a single forward pass of π_ϕ during deployment. Unfortunately, training the DAD network by optimizing (5) requires the likelihood density $p(h_T|\theta, \pi)$ to be analytically available—an assumption that is too restrictive in many practical situations. The architecture for DAD also assumes conditionally independent designs, which is unsuitable in some settings like time-series data. Our method lifts both of these restrictions.

3 Implicit Deep Adaptive Design

We have seen that the traditional step-by-step approach to adaptive design for implicit models [16, 21, 30] is too costly to deploy for most applications, whilst the only existing policy-based approach, DAD [17], makes restrictive assumptions that prevent it being applied to implicit models. We aim to relax the restrictive assumptions of the latter, making policy-based BOED applicable to all models where we can sample from $y|\theta, \xi$ and compute the derivative $\partial y/\partial \xi$, a strict superset of the class of models that can be handled by DAD. This requires new training objectives for the policy network that do not involve an explicit likelihood and are not based on conditionally independent experiments, along with new architectures that work for non-exchangeable models like time series.

3.1 Information lower bounds for policy-based experimental design without likelihoods

To establish a suitable likelihood-free training objective for the implicit setting, our high-level idea is to leverage recent advances in variational MI [see 42, for an overview], which have shown promise for *static* BOED [16, 28, 29]. While using these bounds in the traditional framework of (2) would not permit real-time experiments, one could consider a naive application of them to the policy objective of (3) by replacing each $I_{h_{t-1}}$ with a suitable variational lower bound that uses a 'critic' $U_t : \mathcal{H}^{t-1} \times \Theta \to \mathbb{R}$ to avoid explicit likelihood evaluations, where $\mathcal{H}^{t-1}$ and Θ are the spaces of histories and parameters respectively. An effective critic successfully encapsulates the true likelihood, tightening the bound. Although its form depends on the choice of bound, all critics are parametrized and trained in the same way, namely by a neural network U_{ϕ_t} which is optimized to tighten the bound. Unfortunately, replacing each $I_{h_{t-1}}$ involves learning T such critic networks and requires samples from all posteriors $p(\theta|h_{t-1})$, which will typically be impractically costly.

To avoid this issue, we show that we can obtain a unified information objective similar to (4), *even without conditionally independent experiments*. The following proposition therefore marks the first key milestone in eliminating the restrictive assumptions of [17], by establishing a unified objective without intermediate posteriors that is valid even when the model itself changes between time steps.

Proposition 1 (Generalized total expected information gain)**.** *Consider the data generating distribution* $p(h_T|\theta, \pi) = \prod_{t=1:T} p(y_t|\theta, \xi_t, h_{t-1})$*, where* $\xi_t = \pi(h_{t-1})$ *are the designs generated by the policy and, unlike in* (4)*,* y_t *is allowed to depend on the history* h_{t-1}*. Then we can write* (3) *as*

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(h_T|\theta, \pi)\right] - \mathbb{E}_{p(h_T|\pi)}\left[\log p(h_T|\pi)\right]. \tag{6}$$

Proofs are presented in Appendix A. The advantage of (6) is that we can draw samples from $p(\theta)p(h_T|\theta, \pi)$ simply by sampling our model and taking forward passes through the design network. However, neither of the *densities* $p(h_T|\theta, \pi)$ and $p(h_T|\pi)$ are tractable for implicit models.

To side-step this intractability, we observe that $\mathcal{I}_T(\pi)$ takes an analogous form to a MI between θ and h_T. For measure-theoretic reasons, namely because the $\xi_{1:T}$ are deterministic given $y_{1:T}$ (see Appendix A for a full discussion), it is not the true MI. However, the following two propositions show that we can treat $\mathcal{I}_T(\pi)$ *as if it were this MI*. Specifically, we show that the InfoNCE [57] and NWJ [37] bounds on the MI can be adapted to establish tractable lower bounds on our unified objective $\mathcal{I}_T(\pi)$. These two bounds both utilize a *single* auxiliary critic network U_ψ that is trained simultaneously with the design network.

Proposition 2 (NWJ bound for implicit policy-based BOED)**.** *For a design policy* π *and a critic function* $U : \mathcal{H}^T \times \Theta \to \mathbb{R}$*, let*

$$\mathcal{L}_T^{NWJ}(\pi, U) := \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[U(h_T, \theta)\right] - e^{-1}\mathbb{E}_{p(\theta)p(h_T|\pi)}\left[\exp(U(h_T, \theta))\right], \tag{7}$$

then $\mathcal{I}_T(\pi) \geq \mathcal{L}_T^{NWJ}(\pi, U)$ *holds for any* U*. Further, the inequality is tight for the optimal critic* $U^*_{NWJ}(h_T, \theta) = \log p(h_T|\theta, \pi) - \log p(h_T|\pi) + 1$.

Algorithm 1: Implicit Deep Adaptive Design with (iDAD)

Input: Differentiable simulator f, sampler for prior $p(\theta)$, number of experimental steps T
Output: Design network π_ϕ, critic network U_ψ
while Computational training budget not exceeded **do**
  Sample $\theta \sim p(\theta)$ and set $h_0 = \varnothing$
  for $t = 1, ..., T$ **do**
    Compute $\xi_t = \pi_\phi(h_{t-1})$
    Sample $\varepsilon_t \sim p(\varepsilon)$ and compute $y_t = f(\varepsilon_t; \xi_t, \theta, h_{t-1})$
    Set $h_t = \{(\xi_1, y_1), ..., (\xi_t, y_t)\}$
  end
  Estimate $\nabla_{\phi,\psi}\mathcal{L}_T(\pi_\phi, U_\psi)$ as per (10) where $\mathcal{L}_T$ is $\mathcal{L}_T^{\text{NWJ}}$ (7) or $\mathcal{L}_T^{\text{NCE}}$ (8)
  Update the parameters (ϕ, ψ) using stochastic gradient ascent scheme
end
For deployment, use the deterministic trained design network π_ϕ to obtain a designs ξ_t directly.

Proposition 3 (InfoNCE bound for implicit policy-based BOED)**.** *Let $\theta_{1:L} \sim p(\theta_{1:L}) = \prod_i p(\theta_i)$ be a set of contrastive samples where $L \geq 1$. For design policy π and critic function $U : \mathcal{H}^T \times \Theta \to \mathbb{R}$, let*

$$\mathcal{L}_T^{NCE}(\pi, U; L) := \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{\exp(U(h_T, \theta_0))}{\frac{1}{L+1}\sum_{i=0}^{L}\exp(U(h_T, \theta_i))}\right], \tag{8}$$

then $\mathcal{I}_T(\pi) \geq \mathcal{L}_T^{NCE}(\pi, U; L)$ for any U and $L \geq 1$. Further, the optimal critic, $U^*_{NCE}(h_T, \theta) = \log p(h_T|\theta, \pi) + c(h_T)$ where $c(h_T)$ is any arbitrary function depending only on the history, recovers the sPCE bound in (5); the inequality is tight in the limit as $L \to \infty$ for this optimal critic.

We propose these two alternative bounds due to their complementary properties: the NWJ bound can have large variance, but tends to be less biased. That is, the NWJ bound tends to be tighter for good critics, but is itself more difficult to reliably estimate and thus optimize. While the NWJ critic must learn to self-normalize, the InfoNCE bound avoids this issue but typically will not be tight for finite L even with an optimal critic (note $\mathcal{L}_T^{\text{NCE}} \leq \log(L + 1)$ [42]). Consequently, only the NWJ objective recovers the true optimal policy if our critic has infinite capacity and our optimization scheme is perfect, i.e. $\arg\max_\pi \max_U \mathcal{L}_T^{\text{NWJ}}(\pi, U) = \pi^* \neq \arg\max_\pi \max_U \mathcal{L}_T^{\text{NCE}}(\pi, U; L)$ in general, but it can be more difficult to work with in practice. We present a third bound that provides a potential solution to this, and further discuss the relative merits of the two bounds, in Appendix A.

We note that for both bounds the optimal critic does not depend on the learned policy. The final trained critic can be used to approximate the density ratio $p(h_T|\theta, \pi)/p(h_T|\pi) = p(\theta|h_T)/p(\theta)$, either directly in the case of the NWJ critic, or via self-normalization for the InfoNCE bound. We can use this to help approximate the posterior over θ given the collected real data from the experiment. This means we can perform likelihood-free inference after training the critic, which extends previous results [28, 29] from the static to the adaptive policy-based setting.

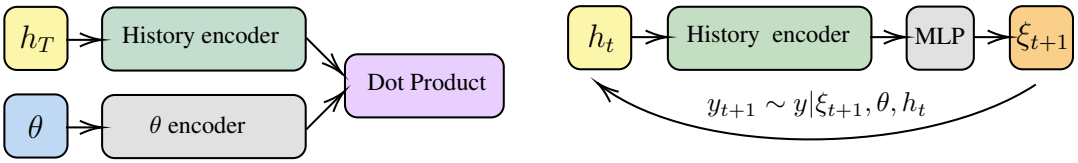
3.2 Parameterization and gradient estimation

In practice, we represent the policy π and the critic U as neural networks, π_ϕ and U_ψ respectively, such that the lower bounds become a function $\mathcal{L}(\pi_\phi, U_\psi)$ of their parameters. By simultaneously optimizing $\mathcal{L}(\pi_\phi, U_\psi)$ with respect to both ϕ and ψ, we both learn a tight bound that accurately represents the true MI and a design policy network that produces high-quality designs under this metric.

We optimize these bounds using stochastic gradient methods [26, 49]. For this, we must account for the fact that the parameter ϕ affects the probability distributions with respect to which expectations are taken. We deal with this problem by utilizing the reparametrization trick [35, 48], for which we assume that design space Ξ and observation space $\mathcal{Y}$ are continuous. To this end, we first formalize the notion of a differentiable implicit model in the adaptive design setting as

$$y_t = f(\varepsilon_t; \xi_t(h_{t-1}), \theta, h_{t-1}), \quad \text{where} \quad \theta \sim p(\theta), \quad \varepsilon_t \sim p(\varepsilon) \;\; \forall t \in \{1, \ldots, T\} \tag{9}$$

and we assume that we can compute the derivatives $\partial f/\partial \xi$ and $\partial f/\partial h$. Interestingly, it is possible to use an implicit prior without access to the density $p(\theta)$, and we do not need access to $\partial f/\partial \theta$.



(a) Critic network, U_ψ (b) Policy network, π_ϕ

Figure 2: Overview of network architectures used in iDAD.

Under these conditions, we can express the bounds in terms of expectations that do not depend on ϕ or ψ, and hence move the gradient operator inside. For $\mathcal{L}_T^{\text{NCE}}(\pi_\phi, U_\psi; L)$, for example, we have

$$\nabla_{\phi,\psi}\mathcal{L}_T^{\text{NCE}} = \mathbb{E}_{p(\theta_{0:L})p(\varepsilon_{1:T})}\left[\nabla_{\phi,\psi}\log\frac{\exp(U_\psi(h_T(\varepsilon_{1:T},\pi_\phi),\theta_0))}{\frac{1}{L+1}\sum_{i=0}^{L}\exp(U_\psi(h_T(\varepsilon_{1:T},\pi_\phi),\theta_i))}\right]. \tag{10}$$

While each element of the history h_T depends on ϕ in a possibly nested manner, we do not need to explicitly keep track of these dependencies thanks to automatic differentiation [5, 41].

Like DAD, our new method—which we call *implicit Deep Adaptive Design* (iDAD)—is trained with simulated histories $h_T = \{(\xi_i, y_i)\}_{i=1:T}$ prior to the actual experiment, allowing design decision to be made using a single forward pass during deployment. Unlike DAD, however, it does not require knowledge of the likelihood function, nor the assumption of conditionally independent designs, which significantly broadens its applicability. A summary of the iDAD approach is given in Algorithm 1.

3.3 Network architectures

The iDAD approach involves the simultaneous training of the *policy* π_ϕ and *critic* U_ψ networks. It is essential to choose the neural architectures of these two components carefully to learn effective policies: poor choices of critic architecture will lead to loose, unrepresentative, bounds, while poor choices of policy architecture will directly lead to ineffective policies. Good choices of architecture need to balance flexibility with ease of training, and will typically require the incorporation of problem-specific inductive biases. A high-level summary of our architectures is shown in Figure 2.

The critic network, U_ψ, takes a *complete* history h_T and the parameter θ as input, and outputs a scalar. Our suggested architecture first encodes the two inputs separately to representations of the same dimension, using a *history encoder*, E_{ψ_h}, and a *parameter encoder*, E_{ψ_θ}, respectively. The output of the critic is then simply taken as their dot product $U_\psi(h_T, \theta) \coloneqq E_{\psi_h}(h_T)^\top E_{\psi_\theta}(\theta)$; after training, the two encodings correspond to approximate sufficient statistics [9]. This setup corresponds to a separable critic architecture, as is commonly used in the representation learning literature [3, 8, 57]. While we use a simple MLP for E_{ψ_θ}, the setup for E_{ψ_h} varies with the context as we discuss below.

The policy network, π_ϕ, takes the available history, h_t, as input, and outputs a design. Our suggested architecture makes use of a history encoder, E_{ϕ_h}, of the same form as E_{ψ_h}, except that it must now take in varying length inputs; its output remains a fixed dimensional embedding. We then pass this embedding through an MLP to produce the design ξ_{t+1}. At the next iteration of the experiment, the same policy network is then called again with the updated history $h_{t+1} = h_t \cup \{(\xi_{t+1}, y_{t+1})\}$.

We use the same architecture for both history encoders, E_{ψ_h} and E_{ϕ_h}, but do not share network parameters between them. This architecture first individually embeds each design–outcome pair (ξ_t, y_t) to a corresponding representation, r_t, using a simple MLP that is shared across all time steps. The produced history encoding is then an aggregation of these representations, with how this is done depending on whether the experiments are *conditionally independent*, i.e. $y_t \perp\!\!\!\perp h_{t-1}|\theta, \xi_t$, or not.

Foster et al. [17] proved that if experiments are conditionally independent, then the optimal policy is invariant to the order of the history. We prove that the same is true for the critic in Proposition 5 in Appendix C. In our setup, we can exploit this result by using a *permutation invariant* aggregation strategy for $\{r_1, \ldots, r_t\}$ when conditional independence holds. The simplest approach to do this would be to use sum-pooling [62], as was done in DAD. However, to improve on this, we instead propose using a more advanced permutation invariant architecture based on self-attention [13, 25, 39, 47, 59], namely that of Parmar et al. [40]; we find this provides notable empirical gains. When conditional independence does not hold, this approach is no longer appropriate and we instead use an LSTM [22] for the aggregation. See Appendix C for further details.

4 Related work

Adaptive policy-based BOED has only recently been introduced [17] and has not yet been extended to implicit models—the gap that this work addresses. Previous approaches to adaptive experiments usually follow the two-step greedy procedure described in Section 2. Methods for MI/EIG estimation without likelihoods include the use of variational bounds [15, 16, 28] and ratio estimation [27, 30]; approximate Bayesian computation together with kernel density estimation [43]; and approximating the intractable likelihood first, for example via polynomial chaos expansion [24], followed by applying likelihood-based estimators, such as nested Monte Carlo [46]. The maximization step in more traditional methods tends to rely on gradient-free optimization, including grid-search, evolutionary algorithms [44], Bayesian optimization [15, 30], or Gaussian process surrogates [38]. More recently, gradient-based approaches have been introduced [15, 28], some of which allow the estimation and optimization simultaneously in a single stochastic-gradient scheme [16, 23, 29]. From a posterior estimation perspective, likelihood-free inference can be performed via approximate Bayesian computation [33, 54], ratio estimation [56], conventional MCMC for methods that make tractable approximation to the likelihood [23, 24], or as a byproduct of MI estimation [16, 27, 29, 30].

5 Experiments

We evaluate the performance of **iDAD** on a number of real-world experimental design problems and a range of baselines. A summary of all the methods that we consider is given in Table 1. Since we aim to perform adaptive experiments in *real-time*, we focus mostly on baselines that do not require significant computational time during the experiment. These include heuristic approaches that require no training, namely **equal** interval designs (when possible) and **random** designs, as well as static BOED approaches, where we, non-adaptively, choose all the designs prior to the experiment by optimising the mutual information objective of Equation (1) with $\xi = \{\xi_1, ..., \xi_T\}$ and $y = \{y_1, \dots, y_T\}$. The static BOED approaches we consider are the **MINEBED** method of [28] and the likelihood-free ACE approach of [16], where we use the prior as a proposal distribution, referring to this baseline as **SG-BOED**. We also implement the expensive traditional non-amortized myopic strategy described in Section 2, for which we use the **variational** approach of [16], with the Barber-Agakov bound [4, 15], at each experiment step (see Appendix D.3 for details). Finally, where possible, we compare our method with DAD [17], in order to assess the performance gap that would arise if we had an analytic likelihood. This comparison is done primarily for evaluation purposes—because it has access to the likelihood density, DAD serves as an upper bound on the performance iDAD can achieve; one should use explicit likelihood methods whenever possible.

Table 1: Key properties of considered methods.

	Adaptive	Real-time	Implicit
Random	✗	N/A	✓
Equal interval	✗	N/A	✓
MINEBED	✗	N/A	✓
SG-BOED	✗	N/A	✓
Variational	✓	✗	✓
DAD	✓	✓	✗
iDAD	✓	✓	✓

The main performance metric that we focus on is the total EIG, $\mathcal{I}_T(\pi)$, as given in (6). In cases where the likelihood is available, we estimate the $\mathcal{I}_T(\pi)$ using the sPCE lower bound in (5) and its sister upper bound, the sequential Nested Monte Carlo bound [sNMC; 17]. To ensure that the bounds are tight, we evaluate them with a large number of contrastive samples, i.e. $L \geq 10^5$. Where the likelihood is truly intractable, we assess the iDAD strategy in a more qualitative manner by looking at the optimal designs and approximate posteriors. For the adaptive experiments, we further consider the deployment time (i.e. the time required to propose a design), which is a critical metric for our aims. All deployment times exclude the time needed to determine the first experiment as it can be computed up-front, during the training phase. Timings for training the policy itself are given in Appendix D.

5.1 Location Finding

We first demonstrate our approach on the location finding experiment from [17]. Inspired by the acoustic energy attenuation model [53], this experiment involves finding the locations of multiple hidden sources, each emitting a signal with intensity that decreases according to the inverse-square law. The *total intensity*—a superposition of these signals—can be measured noisily at any location. The design problem is choosing where to measure the total signal in order to uncover the sources.

Table 2: Lower bounds on the total information, $\mathcal{I}_{10}(\pi)$, for the location finding experiment in Section 5.1. The bounds were estimated using $L = 5 \times 10^5$ contrastive samples. Errors indicate ± 1 standard errors estimated over 4096 histories (128 for variational). Corresponding upper bounds are given in Table 6 in Appendix D.

Method \ θ dim.	4D	6D	10D	20D
Random	4.791 ± 0.040	3.468 ± 0.014	1.889 ± 0.011	0.552 ± 0.006
MINEBED	5.518 ± 0.028	4.221 ± 0.028	2.458 ± 0.029	0.801 ± 0.019
SG-BOED	5.547 ± 0.028	4.215 ± 0.030	2.454 ± 0.029	0.803 ± 0.019
Variational	4.639 ± 0.144	3.625 ± 0.165	2.181 ± 0.151	0.669 ± 0.097
iDAD (NWJ)	$\mathbf{7.694 \pm 0.045}$	5.765 ± 0.036	$\mathbf{3.252 \pm 0.039}$	$\mathbf{0.877 \pm 0.022}$
iDAD (InfoNCE)	$\mathbf{7.750 \pm 0.039}$	$\mathbf{5.986 \pm 0.037}$	$\mathbf{3.251 \pm 0.039}$	$\mathbf{0.871 \pm 0.020}$
DAD	7.967 ± 0.034	6.300 ± 0.030	3.337 ± 0.039	0.937 ± 0.022

Table 3: Lower and upper bounds on MI $\mathcal{I}_{10}(\pi)$ for different network architectures on location finding experiment using the InfoNCE bound. All estimates obtained as in Table 2.

Design	Critic	Lower bound	Upper bound
Attention	**Attention**	$\mathbf{7.750 \pm 0.039}$	$\mathbf{7.863 \pm 0.043}$
Attention	Pooling	7.567 ± 0.037	7.632 ± 0.039
Pooling	Attention	7.398 ± 0.040	7.470 ± 0.042
Pooling	Pooling	7.135 ± 0.034	7.192 ± 0.041

In Table 2 we can see that iDAD substantially outperforms all baselines including, perhaps surprisingly, the traditional (non-amortized) adaptive variational approach, despite its large computational cost shown Table 4. The poor performance of the variational approach is likely driven by the inability of the mean-field variational family to capture the highly non-Gaussian true posterior, highlighting the detrimental effect that wrong posteriors can have on determining optimal designs when using the traditional sequential BOED approach.

Table 4: Deployment time of adaptive methods in 2D, measured on a CPU. Errors were calculated on the basis of 10 runs.

Method	Deployment time (sec.)
Variational	$2256.0 \quad \pm 1\%$
iDAD (NWJ)	$0.0167 \pm 2\%$
iDAD (InfoNCE)	$0.0168 \pm 2\%$
DAD	$0.0070 \pm 6\%$

Table 2 further shows that the performance gap to the likelihood-based DAD method is small, even as the dimension of the design and parameter space grows. Though the information gained by all methods decreases with the dimensionality, this is to be expected: in higher dimensions it is inherently more difficult to infer the relative direction of the sources from observing their intensity. Overall, this experiment demonstrates that iDAD is able to learn near-optimal amortized design policies without likelihoods, while being run in milliseconds at deployment.

Ablation: attention to history. We next assess the benefit of utilizing our proposed more sophisticated permutation invariant architectures, compared to the simple pooling of [62] used in [17]. Our approach incorporates attention layers into both networks that we train. This leads us to four possible combinations of network architectures. Table 3 compares the efficacy of the resulting design policies and strongly suggests that incorporating attention mechanisms in either and/or both networks improves performance, with inclusion in the design network particularly important.

We preform further ablation studies to investigate and demonstrate important properties of our method, such as its scalability with the number of experiments T, stability between different training runs and performance to errors in the design network (introduced by not training the network to convergence). Results and discussion are provided in Appendix D.4.4.

5.2 Pharmacokinetic model

Our next experiment is taken from the pharmacokinetics literature and has been studied in other recent works on BOED for implicit models [28, 63]. Specifically, we consider the compartmental model

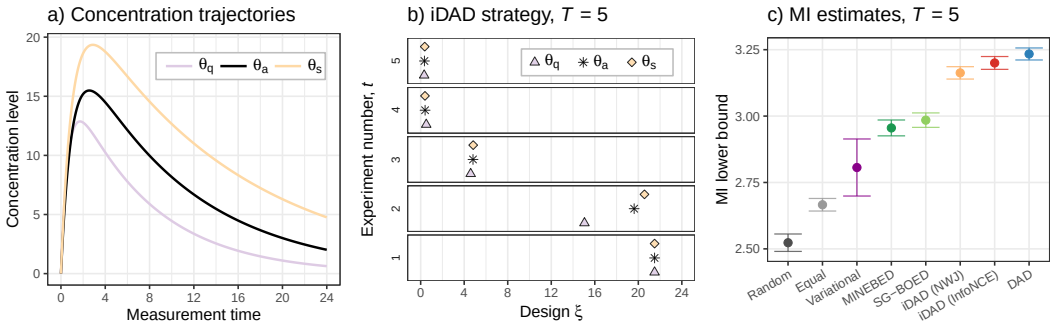


Figure 3: Plots for pharmacokinetics experiment. a) Visualisation of model showing concentration level as a function of measurement time for 3 values of θ, resulting in a quick (θ_q), average (θ_a), or slow (θ_s) trajectory. b) Designs selected by an iDAD policy trained with InfoNCE. c) MI lower bounds achieved by iDAD and baselines. All estimates obtained as in Table 2.

of [50], for which the distribution of an administered drug through the body is governed by three parameters: absorption rate k_α, elimination rate k_e, and volume V, which form the parameters of interest, i.e. $\theta = (k_\alpha, k_e, V)$. Given $T = 5$ patients, the design problem is to adaptively choose blood sampling times, $0 \leq \xi_t \leq 24$ hours, for each, measured from the the point the drug was administered (with patient 2 not being administered until after sampling patient 1 etc). Plausible concentration trajectories are shown in Figure 3a). Full details and further results are given in Appendix D.5.

We first qualitatively consider the design policy of iDAD (trained with the InfoNCE objective) in Figure 3b). As we have not yet observed any data, the optimal design for the first patient (bottom row) is the same for all θ. For the second patient, only guided by ξ_1 and the outcome y_1, iDAD is already able to distinguish between quickly and slowly decaying concentration trajectories: it proposes a significantly earlier measurement time for the quickly decaying trajectory (purple triangle, θ_q) and later time for the slowly decaying one (yellow diamond, θ_s). For the third patient, iDAD always targets the peak of the drug concentration distribution which is quite similar for all θ. Measurements for the last two patients are made soon after the drug has been administered ($\sim 15 - 30$ min), when concentration levels increase rapidly, to capture information about how quickly the drug is absorbed.

To provide more quantitative assessment and compare to our baselines, we again consider the final EIG values as shown in Figure 3c). This reveals that the iDAD strategies perform best among the methods that are applicable to implicit models, confirming that the learnt policies propose superior designs. The performance gap to DAD, which relies on explicit likelihoods, is not statistically significant (at the 5% level) for iDAD trained with InfoNCE, while significant, but still small, for NWJ.

Finally, we consider the convergence of the iDAD networks under the different training objectives and compare to DAD for reference. As shown in Figure 4, although all three converge to approximately the same value, they do so at rather different speeds: while DAD requires about 5000 gradient updates, implicit methods need longer training and tend to exhibit higher variance, particularly NWJ.

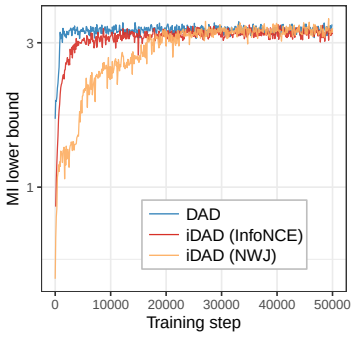


Figure 4: Convergence of MI lower bounds.

5.3 SIR Model

In this experiment, we demonstrate our approach on an implicit model from epidemiology. Namely, we consider a formulation of the stochastic SIR model [10] that is based on stochastic differential equations (SDEs), as done by [29]. Here, individuals in a fixed population belong to one of three categories: susceptible, infected or recovered. Susceptible people can become infected and then recover, with the dynamics of these two events being governed by two model parameters—the infection rate β and the recovery rate γ. Our aim is to determine the optimal times τ at which to measure the number of infected people, $I(\tau)$, in order to estimate the two parameters. This implicit model is challenging because data simulation is expensive, since we need to solve many SDEs, and experimental designs have a time-dependency. See Appendix D.6 for full details.

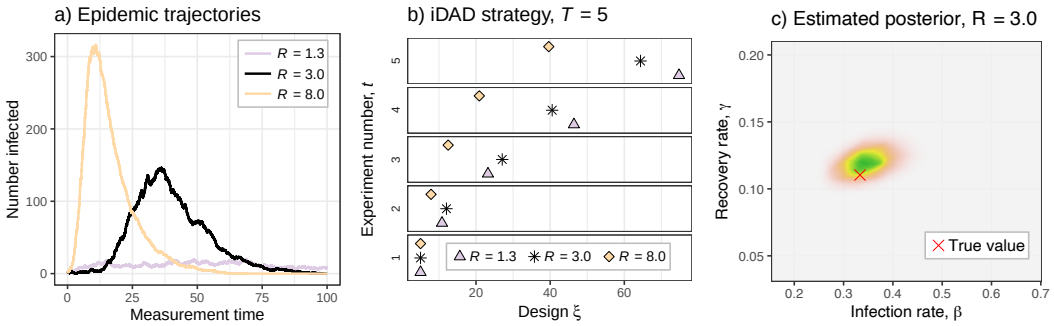


Figure 5: a) Epidemic trajectories for 3 realization of (β, γ) with different reproduction numbers $R = \beta/\gamma$. b) Designs selected by an iDAD policy trained with NWJ. c) Example posterior estimates from the critic network given data generated with the ground-truth parameters shown by the red cross.

We train a iDAD networks to perform $T = 5$ experiments and compare against random, equal interval, and static design baselines; DAD cannot be run because the problem corresponds to a true implicit model. Table 5 shows lower bound estimates on the MI and demonstrates that iDAD outperforms all compared methods. Note that a degree of caution is required when analysing the results, as they are influenced by unavoidable biases in the estimation process. Namely, a critic is still required to estimate the MI lower bound, and there may be variations in the effectiveness of these critics, with less effective ones corresponding to looser bounds and therefore underestimating the true MI. Nonetheless, for other models where such checks are possible, we have found the bounds to be relatively tight, while, even if this turns out not to be the case here, the fact that the critics for the static approaches are easier to train should mean our relative evaluations for iDAD (and Random) are still conservative compared to the other baselines.

Table 5: MI lower bounds (± 1 s.e.).

Method	Lower bound
Random	1.915 ± 0.032
Equal interval	2.669 ± 0.023
MINEBED	3.400 ± 0.001
SG-BOED	3.752 ± 0.020
iDAD (NWJ)	$\mathbf{3.869 \pm 0.001}$
iDAD (InfoNCE)	$\mathbf{3.915 \pm 0.020}$

Figure 5 further demonstrates important qualitative results for this model. Figure 5a) shows different epidemic trajectories, i.e. the number of infected $I(\tau)$ people as a function of measurement time τ, whilst 5b) plots their corresponding designs obtained from the learned iDAD policy. Importantly, diseases with a significantly different profile, e.g. a slow ($R = 1.3$) or a fast ($R = 8.0$) spread result in different sets of optimal designs, highlighting the adaptivity of iDAD. Finally, Figure 5c) shows an example posterior distribution estimate from the learnt iDAD critic network, which we see is consistent with the ground truth parameters.

6 Discussion

Limitations. The benefit that iDAD can be used in live experiments comes at the cost of substantial training that can be computationally expensive. However, this is mitigated by its amortization of the adaptive design process, such that only one network needs training, even if we have multiple experiment instances. The cost–performance trade-off can also be directly controlled by judicious choices of architecture and the amount of training performed. Another natural limitation is that the use of gradients naturally restricts the approach to continuous design settings, something which future work might look to address.

Conclusions. In this paper we introduced iDAD—the first policy-based adaptive BOED method that can be applied to implicit models. By training a design network without likelihoods upfront, iDAD is thus the first method that allows real-time adaptive experiments for simulator-based models. In our experiments, iDAD performed significantly better than all likelihood-free baselines. Further, by using models where the likelihood is available as a test bed, we found that it was able to almost match the analogous likelihood-based adaptive approach, which acts as an upper bound on what might be achieved without access to the likelihood itself. In conclusion, we believe iDAD marks a step change in Bayesian experimental design for *implicit* models, allowing designs to be proposed quickly, adaptively, and non-myopically during the live experiment.

Acknowledgments and Disclosure of Funding

DRI is supported by EPSRC through the Modern Statistics and Statistical Machine Learning (StatML) CDT programme, grant no. EP/S023151/1. AF gratefully acknowledges funding from EPSRC grant no. EP/N509711/1. SK was supported in part by the EPSRC Centre for Doctoral Training in Data Science, funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016427/1) and the University of Edinburgh.

References

[1] Edward J. Allen, Linda J. S. Allen, Armando Arciniega, and Priscilla E. Greenwood. Construction of equivalent stochastic differential equation models. *Stochastic Analysis and Applications*, 26(2):274–297, 2008.

[2] Linda J.S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2):128–142, 2017. ISSN 2468-0427. doi: https://doi.org/10.1016/j.idm.2017.03.001.

[3] Philip Bachman, R Devon Hjelm, and William Buchwalter. Learning representations by maximizing mutual information across views. *arXiv preprint arXiv:1906.00910*, 2019.

[4] David Barber and Felix Agakov. The im algorithm: A variational approach to information maximization. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, NIPS'03, page 201–208, Cambridge, MA, USA, 2003. MIT Press.

[5] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18, 2018.

[6] Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, 2018.

[7] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 13–18 Jul 2020.

[9] Yanzhi Chen, Dinghuai Zhang, Michael U. Gutmann, Aaron Courville, and Zhanxing Zhu. Neural approximate sufficient statistics for implicit models. In *International Conference on Learning Representations*, 2021.

[10] Alex R Cook, Gavin J Gibson, and Christopher A Gilligan. Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868, 2008.

[11] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 2020.

[12] Mahasen B. Dehideniya, Christopher C. Drovandi, and James M. McGree. Optimal bayesian design for discriminating between models with intractable likelihoods in epidemiology. *Computational Statistics & Data Analysis*, 124:277–297, 2018. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2018.03.004.

[13] Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 4171–4186, 2019.

[14] Yann Dubois, Jonathan Gordon, and Andrew YK Foong. Neural Process Family. http://yanndubs.github.io/Neural-Process-Family/, September 2020.

[15] Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems 32*, pages 14036–14047. Curran Associates, Inc., 2019.

[16] Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, and Tom Rainforth. A unified stochastic gradient approach to designing bayesian-optimal experiments. In *International Conference on Artificial Intelligence and Statistics*, pages 2959–2969. PMLR, 2020.

[17] Adam Foster, Desi R Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. *Proceedings of the 38th International Conference on Machine Learning (ICML), PMLR 139*, 2021.

[18] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[19] Matthew Graham and Amos Storkey. Asymptotically exact inference in differentiable generative models. In *Artificial Intelligence and Statistics*, pages 499–508. PMLR, 2017.

[20] PeterJ. Green, Krzysztof Latuszynski, Marcelo Pereyra, and Christian P. Robert. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25(4):835–862, 2015. doi: 10.1007/s11222-015-9574-5.

[21] Markus Hainy, Christopher C. Drovandi, and James M. McGree. Likelihood-free extensions for Bayesian sequentially designed experiments. In Joachim Kunert, Christine H. Müller, and Anthony C. Atkinson, editors, *mODa 11 - Advances in Model-Oriented Design and Analysis*, pages 153–161. Springer International Publishing, 2016.

[22] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

[23] Xun Huan and Youssef Marzouk. Gradient-based stochastic optimization methods in bayesian experimental design. *International Journal for Uncertainty Quantification*, 4(6), 2014.

[24] Xun Huan and Youssef M Marzouk. Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, 2013.

[25] Cheng Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck. Music transformer: Generating music with long-term structure, 2019. ISSN 23318422.

[26] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[27] S. Kleinegesse and M.U. Gutmann. Efficient Bayesian experimental design for implicit models. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89 of *Proceedings of Machine Learning Research*, pages 1584–1592. PMLR, 2019.

[28] Steven Kleinegesse and Michael Gutmann. Bayesian experimental design for implicit models by mutual information neural estimation. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 5316–5326. PMLR, 2020.

[29] Steven Kleinegesse and Michael U. Gutmann. Gradient-based bayesian experimental design for implicit models using mutual information lower bounds. *arXiv preprint arXiv:2105.04379*, 2021.

[30] Steven Kleinegesse, Christopher Drovandi, and Michael U. Gutmann. Sequential Bayesian Experimental Design for Implicit Models via Mutual Information. *Bayesian Analysis*, pages 1 – 30, 2021. doi: 10.1214/20-BA1225.

[31] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*, 2019.

[32] Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.

[33] J. Lintusaari, M.U. Gutmann, R. Dutta, S. Kaski, and J. Corander. Fundamentals and recent developments in approximate Bayesian computation. *Systematic Biology*, 66(1):e66–e82, January 2017.

[34] David McAllester and Karl Stratos. Formal limitations on the measurement of mutual information. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 875–884. PMLR, 2020.

[35] Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

[36] Jay I Myung, Daniel R Cavagnaro, and Mark A Pitt. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, 2013.

[37] Xuanlong Nguyen, Martin J. Wainwright, and Michael I. Jordan. Estimating divergence functionals and the likelihood ratio by convex risk minimization. *IEEE Transactions on Information Theory*, 56(11), 2010. ISSN 00189448. doi: 10.1109/TIT.2010.2068870.

[38] Antony Overstall and James McGree. Bayesian Design of Experiments for Intractable Likelihood Models Using Coupled Auxiliary Models and Multivariate Emulation. *Bayesian Analysis*, 15(1):103 – 131, 2020. doi: 10.1214/19-BA1144.

[39] Emilio Parisotto, Francis Song, Jack Rae, Razvan Pascanu, Caglar Gulcehre, Siddhant Jayakumar, Max Jaderberg, Raphaël Lopez Kaufman, Aidan Clark, Seb Noury, Matthew Botvinick, Nicolas Heess, and Raia Hadsell. Stabilizing transformers for reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7487–7498. PMLR, 2020.

[40] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. Image transformer. In *International Conference on Machine Learning*, pages 4055–4064. PMLR, 2018.

[41] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[42] Ben Poole, Sherjil Ozair, Aäron van den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International Conference on Machine Learning*, pages 5171–5180, 2019.

[43] David J. Price, Nigel G. Bean, Joshua V. Ross, and Jonathan Tuke. On the efficient determination of optimal bayesian experimental designs using abc: A case study in optimal observation of epidemics. *Journal of Statistical Planning and Inference*, 172:1–15, May 2016.

[44] David J Price, Nigel G Bean, Joshua V Ross, and Jonathan Tuke. An induced natural selection heuristic for finding optimal bayesian experimental designs. *Computational Statistics & Data Analysis*, 126:112–124, 2018.

[45] Tom Rainforth. *Automating Inference, Learning, and Design using Probabilistic Programming*. PhD thesis, University of Oxford, 2017.

[46] Tom Rainforth, Rob Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting monte carlo estimators. In *International Conference on Machine Learning*, pages 4267–4276. PMLR, 2018.

[47] Prajit Ramachandran, Niki Parmar, Ashish Vaswani, Irwan Bello, Anselm Levskaya, and Jon Shlens. Stand-alone self-attention in vision models. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[48] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 1278–1286, 2014.

[49] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[50] Elizabeth G. Ryan, Christopher C. Drovandi, M. Helen Thompson, and Anthony N. Pettitt. Towards bayesian experimental design for nonlinear models that require a large number of sampling times. *Computational Statistics & Data Analysis*, 70:45–60, 2014. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2013.08.017.

[51] Elizabeth G Ryan, Christopher C Drovandi, James M McGree, and Anthony N Pettitt. A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016.

[52] Ben Shababo, Brooks Paige, Ari Pakman, and Liam Paninski. Bayesian inference and online experimental design for mapping neural microcircuits. In *Advances in Neural Information Processing Systems*, pages 1304–1312, 2013.

[53] Xiaohong Sheng and Yu Hen Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Transactions on Signal Processing*, 2005. ISSN 1053587X. doi: 10.1109/TSP.2004.838930.

[54] S.A. Sisson, Y. Fan, and M. Beaumont. *Handbook of Approximate Bayesian Computation*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2018. ISBN 9781351643467.

[55] Jiaming Song and Stefano Ermon. Understanding the limitations of variational mutual information estimators. In *International Conference on Learning Representations*, 2020.

[56] Owen Thomas, Ritabrata Dutta, Jukka Corander, Samuel Kaski, and Michael U Gutmann. Likelihood-free inference by ratio estimation. *arXiv preprint arXiv:1611.10242*, 2016.

[57] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[58] Joep Vanlier, Christian A Tiemann, Peter AJ Hilbers, and Natal AW van Riel. A Bayesian approach to targeted experiment design. *Bioinformatics*, 28(8):1136–1142, 2012.

[59] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[60] Benjamin T Vincent and Tom Rainforth. The DARC toolbox: automated, flexible, and efficient delayed and risky choice experiments using bayesian adaptive design. *PsyArXiv preprint*, 2017.

[61] M. Zaharia, Andrew Chen, A. Davidson, A. Ghodsi, S. Hong, A. Konwinski, Siddharth Murching, Tomas Nykodym, Paul Ogilvie, Mani Parkhe, Fen Xie, and Corey Zumar. Accelerating the machine learning lifecycle with MLflow. *IEEE Data Eng. Bull.*, 41:39–45, 2018.

[62] Manzil Zaheer, Satwik Kottur, Siamak Ravanbhakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J Smola. Deep sets. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 2017.

[63] Jiaxin Zhang, Sirui Bi, and Guannan Zhang. A stochastic approximate gradient ascent method for Bayesian experimental design with implicit models. In *The 24nd International Conference on Artificial Intelligence and Statistics*, 2021.

A Proofs

We present proof for all propositions made in the paper, restating each for convenience. We also include additional discussion on technical aspects of the paper.

A.1 Unified objective for non-exchangeable experiments

Proposition 1 (Generalized total expected information gain)**.** *Consider the data generating distribution* $p(h_T|\theta,\pi) = \prod_{t=1:T} p(y_t|\theta,\xi_t,h_{t-1})$*, where* $\xi_t = \pi(h_{t-1})$ *are the designs generated by the policy and, unlike in* (4)*,* y_t *is allowed to depend on the history* h_{t-1}*. Then we can write* (3) *as*

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(h_T|\theta,\pi)\right] - \mathbb{E}_{p(h_T|\pi)}\left[\log p(h_T|\pi)\right]. \tag{6}$$

Proof. Starting with the definition of the total EIG (3) of a policy π:

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\sum_{t=1}^{T} I_{h_{t-1}}(\xi_t)\right] \tag{11}$$

we have by linearity of expectation

$$= \sum_{t=1}^{T} \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[I_{h_{t-1}}(\xi_t)\right] \tag{12}$$

and since $I_{h_{t-1}}$ doesn't depend on data acquired after $t-1$ (the future doesn't influence the past)

$$= \sum_{t=1}^{T} \mathbb{E}_{p(\theta)p(h_{t-1}|\theta,\pi)}\left[I_{h_{t-1}}(\xi_t)\right] \tag{13}$$

which, applying Bayes rule, is equivalent to

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_{t-1}|\pi)p(\theta|h_{t-1})}\left[I_{h_{t-1}}(\xi_t)\right] \tag{14}$$

Next, using Bayes rule we similarly rearrange $I_{h_{t-1}}$:

$$I_{h_{t-1}}(\xi_t) = \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t,h_{t-1})}\left[\log \frac{p(y_t|\theta,\xi_t,h_{t-1})}{p(y_t|\xi_t,h_{t-1})}\right] \tag{15}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t,h_{t-1})}\left[\log \frac{p(\theta|y_t,\xi_t,h_{t-1})}{p(\theta|h_{t-1})}\right] \tag{16}$$

$$= \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t,h_{t-1})}\left[\log p(\theta|y_t,\xi_t,h_{t-1})\right] - \mathbb{E}_{p(\theta|h_{t-1})}\left[\log p(\theta|h_{t-1})\right] \tag{17}$$

$$= \mathbb{E}_{p(\theta|y_t,\xi_t,h_{t-1})p(y_t|\xi_t,h_{t-1})}\left[\log p(\theta|y_t,\xi_t,h_{t-1})\right] - \mathbb{E}_{p(\theta|h_{t-1})}\left[\log p(\theta|h_{t-1})\right] \tag{18}$$

and noting $h_t = h_{t-1} \cup \{(\xi_t, y_t)\}$

$$= \mathbb{E}_{p(\theta|h_t)p(y_t|\xi_t,h_{t-1})}\left[\log p(\theta|h_t)\right] - \mathbb{E}_{p(\theta|h_{t-1})}\left[\log p(\theta|h_{t-1})\right] \tag{19}$$

$$= \mathbb{E}_{p(y_t|\xi_t,h_{t-1})}\Big[\mathbb{E}_{p(\theta|h_t)}\left[\log p(\theta|h_t)\right] - \mathbb{E}_{p(\theta|h_{t-1})}\left[\log p(\theta|h_{t-1})\right]\Big] \tag{20}$$

Substituting this in (14), noting that θ has already been integrated out, yields

$$\mathcal{I}_T(\pi) = \sum_{t=1}^{T} \mathbb{E}_{p(h_{t-1}|\pi)}\mathbb{E}_{p(y_t|\xi_t,h_{t-1})}\Big[\mathbb{E}_{p(\theta|h_t)}\left[\log p(\theta|h_t)\right] - \mathbb{E}_{p(\theta|h_{t-1})}\left[\log p(\theta|h_{t-1})\right]\Big] \tag{21}$$

$$= \sum_{t=1}^{T} \mathbb{E}_{p(h_t|\pi)}\Big[\mathbb{E}_{p(\theta|h_t)}\left[\log p(\theta|h_t)\right] - \mathbb{E}_{p(\theta|h_{t-1})}\left[\log p(\theta|h_{t-1})\right]\Big] \tag{22}$$

$$= \mathbb{E}_{p(h_T|\pi)}\left[\sum_{t=1}^{T} \mathbb{E}_{p(\theta|h_t)}\left[\log p(\theta|h_t)\right] - \mathbb{E}_{p(\theta|h_{t-1})}\left[\log p(\theta|h_{t-1})\right]\right], \tag{23}$$

since we have a telescopic sum this simplifies to

$$= \mathbb{E}_{p(h_T|\pi)}\Big[\mathbb{E}_{p(\theta|h_T)}\left[\log p(\theta|h_T)\right] - \mathbb{E}_{p(\theta)}\left[\log p(\theta)\right]\Big] \tag{24}$$

and finally we apply Bayes rule again to rewrite as

$$= \mathbb{E}_{p(h_T|\pi)p(\theta|h_T)}\Big[\log p(\theta|h_T) - \mathbb{E}_{p(\theta)}\left[\log p(\theta)\right]\Big] \tag{25}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(\theta|h_T) - \log p(\theta)\right] \tag{26}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log p(h_T|\theta,\pi) - p(h_T|\pi)\right] \tag{27}$$

□

A.2 Objective function as a mutual information

We provide some additional discussion on the interpretation of $\mathcal{I}_T(\pi)$ in (6) as a mutual information. First, $\mathcal{I}_T(\pi)$ is not a conventional mutual information between θ and h_T. This is because, for the deterministic policy π considered in this paper, the random variable h_T does not have a density with respect to Lebesgue measure on $\Xi^T \times \mathcal{Y}^T$. Indeed, since the designs $\xi_{1:T}$ are deterministic functions of the observations $y_{1:T}$, to express the sampling distribution of h_T we would have to use Dirac deltas, specifically

$$p(y_{1:T}, \xi_{1:T}|\theta, \pi) = \prod_{t=1}^{T} \delta_{\pi(h_{t-1})}(\xi_t) p(y_t|\theta, \xi_t, h_{t-1}). \tag{28}$$

Due to the presence of Dirac deltas, this is not a conventional probability density, and hence we do not regard $\mathcal{I}_T(\pi)$ as the conventional mutual information between θ and h_T.

We note that we *defined* $p(h_T|\theta, \pi)$ in Proposition 1 differently to $p(y_{1:T}, \xi_{1:T}|\theta, \pi)$ in (28). Specifically, our definition

$$p(h_T|\theta, \pi) = \prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1}) \tag{29}$$

only involves probability densities for $y_{1:T}$, meaning that our $p(h_T|\theta, \pi)$ is a well-defined probability density on $\mathcal{Y}^T$. Formally, we can treat the designs ξ_t, not as additional random variables, but as part of the density for $y_{1:T}$. Indeed, since the policy π is deterministic, it is possible to reconstruct h_{t-1} and ξ_t from $y_{1:t-1}$ and π, so we could write $p(y_t|\theta, y_{1:t-1}, \pi) := p(y_t|\theta, \xi_t, h_{t-1})$. In this formulation, only $y_{1:T}$ are regarded as random variables. This provides a formal justification for the form of $p(h_T|\theta, \pi)$ that we give in Proposition 1. In this setting, we could formally identify $\mathcal{I}_T(\pi)$ as the mutual information between θ and $y_{1:T}$.

However, it is helpful to think of $\mathcal{I}_T(\pi)$ as a mutual information between θ and h_T, because this naturally leads to critics that have access to θ and h_T, rather than θ and $y_{1:T}$. This way of thinking also connects naturally to the case of stochastic policies, which we now discuss.

If we consider additional noise in the design process so that designs are no longer a deterministic function of past data, then $\mathcal{I}_T(\pi)$ is the mutual information between θ and h_T. In this case, we introduce an additional likelihood for designs $p(\xi|\pi, h)$, leading to the overall sampling distribution for the data

$$p(h_T|\theta, \pi) = \prod_{t=1}^{T} p(\xi_t|\pi, h_{t-1}) p(y_t|\theta, \xi_t, h_{t-1}). \tag{30}$$

Unlike in the deterministic case, this is valid probability density on $\Xi^T \times \mathcal{Y}^T$. If we now consider the mutual information between θ and h_T for a fixed policy π we have

$$I(\theta, h_T) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)} \left[\log \frac{\prod_{t=1}^{T} p(\xi_t|\pi, h_{t-1}) p(y_t|\theta, \xi_t, h_{t-1})}{\int_\Theta p(\theta) \prod_{t=1}^{T} p(\xi_t|\pi, h_{t-1}) p(y_t|\theta, \xi_t, h_{t-1}) \, d\theta} \right] \tag{31}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)} \left[\log \frac{\prod_{t=1}^{T} p(\xi_t|\pi, h_{t-1}) \prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1})}{\prod_{t=1}^{T} p(\xi_t|\pi, h_{t-1}) \int_\Theta p(\theta) \prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1}) \, d\theta} \right] \tag{32}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)} \left[\log \frac{\prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1})}{\int_\Theta p(\theta) \prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1}) \, d\theta} \right] \tag{33}$$

noticing that the design likelihood terms cancel out in the integrand, and we reduce to the same integrand given in Proposition 1. Even when the policy is stochastic, the integrand in $I(\theta, h_T)$ only involves terms of the form $p(y_t|\theta, \xi_t, h_{t-1})$, and the likelihood of the design process completely cancels. Thus, the stochasticity of the designs is only present in the sampling distribution $p(h_T|\theta, \pi)$. We therefore see that, as we consider the limiting case of $p(\xi|\pi, h)$ as it approaches a deterministic policy, only the sampling distribution of designs in $I(\theta, h_T)$ changes, with the integrand remaining the same. Under mild assumptions, then, the mutual information between θ and h_T approaches $\mathcal{I}_T(\pi)$ in this limit.

A.3 NWJ and InfoNCE bounds

The next two propositions show that the two bounds—NWJ and InfoNCE—can be applied to the policy-based adaptive BOED setting.

Proposition 2 (NWJ bound for implicit policy-based BOED)**.** *For a design policy π and a critic function $U : \mathcal{H}^T \times \Theta \to \mathbb{R}$, let*

$$\mathcal{L}_T^{NWJ}(\pi, U) \coloneqq \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[U(h_T,\theta)\right] - e^{-1}\mathbb{E}_{p(\theta)p(h_T|\pi)}\left[\exp(U(h_T,\theta))\right], \tag{7}$$

then $\mathcal{I}_T(\pi) \geq \mathcal{L}_T^{NWJ}(\pi, U)$ holds for any U. Further, the inequality is tight for the optimal critic $U^*_{NWJ}(h_T,\theta) = \log p(h_T|\theta,\pi) - \log p(h_T|\pi) + 1$.

Proof. Let $\pi : \mathcal{H}^* \to \Xi$ be any (deterministic) policy taking histories h_t as inputs and returning a design ξ as output, $U : \mathcal{H}^T \times \Theta \to \mathbb{R}$ be any function and define $g(h_T,\theta) \coloneqq \frac{\exp(U(h_T,\theta))}{\mathbb{E}_{p(h_T|\pi)}[\exp(U(h_T,\theta))]}$.

First, we multiply the numerator and denominator of the unified objective (6) by $g(h_T,\theta) > 0$

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)}\right] \tag{34}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)} \log\left[\frac{p(h_T|\theta,\pi)}{p(h_T|\pi)}\frac{g(h_T,\theta)}{g(h_T,\theta)}\right] \tag{35}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log g(h_T,\theta)\right] + \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)g(h_T,\theta)}\right] \tag{36}$$

Next, note that the second term is a KL divergence between two distributions

$$\mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)g(h_T,\theta)}\right] = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(\theta)p(h_T|\theta,\pi)}{p(\theta)p(h_T|\pi)g(h_T,\theta)}\right] \tag{37}$$

$$= KL(p(\theta)p(h_T|\theta,\pi)||\hat{p}(h_T,\theta)) \geq 0 \tag{38}$$

where $\hat{p}(h_T,\theta) = p(\theta)p(h_T|\pi)g(h_T,\theta)$ is a valid distribution since

$$\int p(\theta)p(h_T|\pi)g(h_T,\theta)d\theta dh_T = \mathbb{E}_{p(\theta)p(h_T|\pi)}\frac{\exp(U(h_T,\theta))}{\mathbb{E}_{p(h_T|\pi)}\left[\exp(U(h_T,\theta))\right]} \tag{39}$$

$$= \mathbb{E}_{p(\theta)}1 = 1. \tag{40}$$

Therefore, we have

$$\mathcal{I}_T(\pi) \geq \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log g(h_T,\theta)\right] \tag{41}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}[U(h_T,\theta) - \log \mathbb{E}_{p(h_T|\pi)} \exp(U(h_T,\theta))] \tag{42}$$

$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}[U(h_T,\theta)] - \mathbb{E}_{p(\theta)}\left[\log \mathbb{E}_{p(h_T|\pi)} \exp(U(h_T,\theta))\right] \tag{43}$$

Now using the inequality $\log x \leq e^{-1}x$

$$\geq \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}[U(h_T,\theta)] - e^{-1}\mathbb{E}_{p(\theta)p(h_T|\pi)}\left[\exp(U(h_T,\theta))\right] \tag{44}$$

$$= \mathcal{L}_T^{NWJ}(\pi, U) \tag{45}$$

Finally, substituting $U^*(h_T,\theta) = \log \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)} + 1$ in the bound we get

$$\mathcal{L}_T^{NWJ}(\pi, U^*) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)} + 1\right] - e^{-1}\mathbb{E}_{p(\theta)p(h_T|\pi)}\left[\frac{p(h_T|\theta,\pi)}{p(h_T|\pi)}e^1\right] \tag{46}$$

$$= \mathcal{I}_T(\pi) + 1 - \mathbb{E}_{p(\theta)p(h_T|\pi)}\left[\frac{p(h_T|\theta,\pi)}{p(h_T|\pi)}\right] \tag{47}$$

$$= \mathcal{I}_T(\pi), \tag{48}$$

where we used $\mathbb{E}_{p(\theta)p(h_T|\pi)}\left[\frac{p(h_T|\theta,\pi)}{p(h_T|\pi)}\right] = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[1\right] = 1$, establishing that the bound is tight for the optimal critic.

□

Proposition 3 (InfoNCE bound for implicit policy-based BOED)**.** *Let $\theta_{1:L} \sim p(\theta_{1:L}) = \prod_i p(\theta_i)$ be a set of contrastive samples where $L \geq 1$. For design policy π and critic function $U : \mathcal{H}^T \times \Theta \to \mathbb{R}$, let*

$$\mathcal{L}_T^{NCE}(\pi, U; L) \coloneqq \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)}\mathbb{E}_{p(\theta_{1:L})}\left[\log \frac{\exp(U(h_T,\theta_0))}{\frac{1}{L+1}\sum_{i=0}^{L}\exp(U(h_T,\theta_i))}\right], \tag{8}$$

then $\mathcal{I}_T(\pi) \geq \mathcal{L}_T^{NCE}(\pi, U; L)$ for any U and $L \geq 1$. Further, the optimal critic, $U^*_{NCE}(h_T, \theta) = \log p(h_T|\theta, \pi) + c(h_T)$ where $c(h_T)$ is any arbitrary function depending only on the history, recovers the sPCE bound in (5); the inequality is tight in the limit as $L \to \infty$ for this optimal critic.

Proof. Let $\pi : \mathcal{H}^* \to \Xi$ be any (deterministic) policy taking histories h_t as inputs and returning a design ξ as output. Choose any function (critic) $U : \mathcal{H}^T \times \Theta \to \mathbb{R}$.

We introduce the shorthand

$$g(h_T, \theta_{0:L}) \coloneqq \frac{\exp(U(h_T,\theta_0))}{\frac{1}{L+1}\sum_{i=0}^{L}\exp(U(h_T,\theta_i))} \tag{49}$$

Starting with the definition of the unified objective from Equation (6) we multiply its numerator and denominator by $g(h_T, \theta_{0:L}) > 0$ to get

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)}\right] \tag{50}$$

where $p(\theta_0)p(h_T|\theta_0, \pi) \equiv p(\theta)p(h_T|\theta, \pi)$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)}\right] \tag{51}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi)g(h_T,\theta_{0:L})}{p(h_T|\pi)g(h_T,\theta_{0:L})}\right] \tag{52}$$

We next split the expectation into two terms one of which does not contain the unknown likelihoods and equals $\mathcal{L}^{NCE}$

$$\begin{aligned} &= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)g(h_T,\theta_{0:L})}\right] \\ &\quad + \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log g(h_T,\theta_{0:L})\right] \\ &= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)g(h_T,\theta_{0:L})}\right] + \mathcal{L}^{NCE}(\pi, U; L) \end{aligned} \tag{53}$$

We now show that the first term is a KL divergence and hence non-negative. To see why, first write

$$\mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)g(h_T,\theta_{0:L})}\right] \tag{54}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}{p(\theta_0)p(h_T|\pi)p(\theta_{1:L})g(h_T,\theta_{0:L})}\right] \tag{55}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}{\hat{p}(\theta_{0:L},h_T|\pi)}\right] \tag{56}$$

$$= KL(p(h_T|\theta_0,\pi)p(\theta_{0:L})||\hat{p}(\theta_{0:L},h_T|\pi)). \tag{57}$$

and $\hat{p}(\theta_{0:L}, h_T|\pi)$ is a valid distribution since

$$\int \hat{p}(\theta_{0:L},h_T|\pi)d\theta_{0:L}dh_T = \int p(\theta_0)p(h_T|\pi)p(\theta_{1:L})g(h_T,\theta_{0:L})d\theta_{0:L}dh_T \tag{58}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\frac{\exp(U(h_T,\theta_0))}{\frac{1}{L+1}\sum_{i=0}^{L}\exp(U(h_T,\theta_i))}\right], \tag{59}$$

because of the symmetry $\theta_0 \overset{d}{=} \theta_j \; \forall j = 1, \ldots, L$

$$= \frac{1}{L+1} \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})} \left[\frac{\sum_{j=0}^{L} \exp(U(h_T, \theta_j))}{\frac{1}{L+1}\sum_{i=0}^{L} \exp(U(h_T, \theta_i))} \right] \tag{60}$$

$$= 1. \tag{61}$$

Thus we have established

$$\mathcal{I}_T(\pi) = KL(p(h_T|\theta_0, \pi)p(\theta_{0:L})||\hat{p}(\theta_{0:L}, h_T|\pi)) + \mathcal{L}_T^{NCE}(\pi, U; L) \geq \mathcal{L}_T^{NCE}(\pi, U; L). \tag{62}$$

Next, substituting $U^*(h_T, \theta) = \log p(h_T|\theta, \pi) + c(h_T)$ in the definition of $\mathcal{L}^{NCE}(\pi, U; L)$ we obtain

$$\mathcal{L}_T^{NCE}(\pi, U^*; L) = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})} \left[\frac{p(h_T|\theta_0, \pi)\exp(c(h_T))}{\frac{1}{L+1}\sum_{i=0}^{L} p(h_T|\theta_i, \pi)\exp(c(h_T))} \right] \tag{63}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})} \left[\frac{p(h_T|\theta_0, \pi)}{\frac{1}{L+1}\sum_{i=0}^{L} p(h_T|\theta_i, \pi)} \right], \tag{64}$$

which is exactly the sPCE bound (5), which is monotonically increasing in L and tight in the limit as $L \to \infty$ [see 17, Theorem 2]. □

A.4 A note on optimal critics

An interesting feature of our approach is that, for both the InfoNCE and NWJ bounds, the optimal critics do not depend on the policy. This is because we include the designs as explicit inputs to the critics. Indeed, we have

$$U^*_{\text{NCE}}(h_T, \theta) = \log \left(\prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1}) \right) + c(h_T), \tag{65}$$

$$U^*_{\text{NWJ}}(h_T, \theta) = \log \left(\frac{\prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1})}{\int_\Theta p(\theta) \prod_{t=1}^{T} p(y_t|\theta, \xi_t, h_{t-1})\, d\theta} \right) + 1. \tag{66}$$

In previous work that utilized critics for gradient-based BOED [16, 28], it was typical to not treat the designs $\xi_{1:T}$ as an input to the critic, which renders the optimal critic implicitly dependent on the designs. This makes more sense for static designs, for which the additional design input does not change. Our approach avoids an implicit dependence between policy and optimal critic which may be beneficial for the joint optimization.

B Theoretical Comparison and Additional Bounds

Recently, a number of studies have discussed the challenges of estimating mutual information, an in particular those associated with variational MI estimators [34, 42, 55].

Starting with the InfoNCE bound, it is trivial to show that the bound cannot exceed $\log(L+1)$, where L is the number of contrastive samples used to approximate the marginal in the denominator. Indeed,

$$\mathcal{L}_T^{NCE}(\pi, U; L) = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)} \mathbb{E}_{p(\theta_{1:L})} \left[\log \frac{\exp(U(h_T, \theta_0))}{\frac{1}{L+1}\sum_{i=0}^{L} \exp(U(h_T, \theta_i))} \right] \tag{67}$$

$$\leq \log(L+1) + \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)} \mathbb{E}_{p(\theta_{1:L})} \left[\log \frac{\exp(U(h_T, \theta_0))}{\exp(U(h_T, \theta_0))} \right] \tag{68}$$

$$= \log(L+1) \tag{69}$$

This means that the corresponding Monte Carlo estimator will be highly biases whenever the true mutual information exceeds $\log(L+1)$, regardless of whether we have access to the optimal critic or not. This high bias estimator, however, comes with low variance [see e.g. 42, for discussion]. With

the optimal critic we would require exponential (in the MI) number of samples to accurately estimate the true mutual information.

It might appear at first that the NWJ bound might offer a better trade-off between bias and variance. Recall from the proof of Proposition 2, we have for the *optimal* critic

$$\mathcal{L}_T^{NWJ}(\pi, U^*) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(h_T|\theta,\pi)}{p(h_T|\pi)}\right] + 1 - e^{-1}\mathbb{E}_{p(\theta)p(h_T|\pi)}\left[\frac{p(h_T|\theta,\pi)}{p(h_T|\pi)}e^1\right], \quad (70)$$

of which we form a Monte carlo estimate using N (M) samples for the first (second) term, respectively

$$\approx \frac{1}{N}\sum_{n=1}^{N} \log \frac{p(h_{T,n}|\theta_n,\pi)}{p(h_{T,n}|\pi)} + \left(1 - \frac{1}{M}\sum_{m=1}^{M} \log \frac{p(h_{T,m}|\theta_m,\pi)}{p(h_{T,m}|\pi)}\right), \quad (71)$$

where $\theta_n, h_{T,n} \sim p(\theta)p(h_T|\theta,\pi)$ are samples from the joint distribution and $\theta_m, h_{T,m} \sim p(\theta)p(h_T|\pi)$ are samples from the product of marginals. The first term is a Monte Carlo estimate of the mutual information, while the second has mean zero, meaning that this estimator is unbiased. The second term, however has variance which grows exponentially with the value of the (true) mutual information [see Theorem 2 in 55]. What this means is that even with an optimal critic, we will need an exponential (in the MI) number of samples to control the variance of the NWJ estimator. One might then hope that the variance can be reduced when using a sub-optimal critic at the cost of introducing some (hopefully small) bias. Unfortunately, according to a recent result [see Theorems 3.1 and 4.1 in 34, and the discussion therein], it is not possible to guarantee that a likelihood-free lower bound on the mutual information can exceed $\log(N)$. Indeed, the authors show theoretically and empirically that all high-confidence distribution-free lower bounds on the mutual information require exponential (in the the MI) number of samples.

Constructing a better lower bound on the mutual information—one that does not need exponential number of samples—therefore, requires us to make additional assumptions. Foster et al. [17] propose one such bound, namely the sequential Adaptive Constrative Estimation (sACE). The sACE bound introduces a proposal distribution $q(\theta; h_T)$, which aims to approximate the posterior $p(\theta|h_T)$. Since implicit models were not the focus of the work in [17] the proposed bound, relies on analytically available likelihood. The following proposition shows we can derive a likelihood-free version of the sACE bound.

Proposition 4 (Sequential Likelihood-free ACE)**.** *For a design function π, a critic function U, a number of contrastive samples $L \geq 1$, and a proposal $q(\theta; h_T)$, we have the sequential Likelihood-free Adaptive Contrastive Estimation (sLACE) lower bound*

$$\mathcal{L}_T^{sLACE}(\pi, U, q; L) \coloneqq \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)}\left[\log \frac{U(h_T,\theta_0)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{U(h_T,\theta_\ell)p(\theta_\ell)}{q(\theta_\ell;h_T)}}\right] \leq \mathcal{I}_T(\pi). \quad (72)$$

The bound is tight as $L \to \infty$ for the optimal critic $U^*(h_T,\theta) = \log p(h_T|\theta,\pi) + c(h_T)$, where $c(h_T)$ is arbitrary. In addition, if $q(\theta; h_T) = p(\theta|h_T)$, the bound is tight for the optimal critic $U^*(h_T,\theta)$ with any $L \geq 0$.

Proof. The proof follows similar arguments to the ones in Propositions 2 and 3. First let

$$g(h_T, \theta_{0:L}) \coloneqq \frac{U(h_T,\theta_0)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{U(h_T,\theta_\ell)p(\theta_\ell)}{q(\theta_\ell;h_T)}} \quad (73)$$

Starting with the definition of the EIG:

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)}\right] \quad (74)$$

since $q(\theta_i; h_T)$ is a valid density

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)}\left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)}\right] \quad (75)$$

multiplying its numerator and denominator inside the log by $g(h_T, \theta_{0:L}) > 0$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)} \left[\log \frac{p(h_T|\theta_0,\pi)g(h_T,\theta_{0:L})}{p(h_T|\pi)g(h_T,\theta_{0:L})}\right] \tag{76}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)} \left[\log g(h_T,\theta_{0:L})\right]$$

$$+ \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)} \left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)g(h_T,\theta_{0:L})}\right] \tag{77}$$

The first term is exactly the sLACE bound, $\mathcal{L}_T^{\text{sLACE}}(\pi, U, q; L)$. We now show that the second term is a KL divergence between two distribitions and hence non-negative. To see this

$$\mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)} \left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)g(h_T,\theta_{0:L})}\right] \tag{78}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)} \left[\log \frac{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)}{p(h_T|\pi)g(h_T,\theta_{0:L})p(\theta_0)q(\theta_{1:L};h_T)}\right] \tag{79}$$

$$= KL(p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)||\hat{p}(h_T,\theta_{0:L})), \tag{80}$$

since $\hat{p}(h_T, \theta_{0:L}) \coloneqq p(h_T|\pi)g(h_T,\theta_{0:L})p(\theta_0)q(\theta_{1:L};h_T)$ is a valid density. Indeed:

$$\int \hat{p}(h_T,\theta_{0:L})dh_T d\theta_{0:L} = \mathbb{E}_{q(\theta_{1:L};h_T)p(h_T|\pi)}\left[p(\theta_0)g(h_T,\theta_{0:L})\right] \tag{81}$$

$$= \mathbb{E}_{q(\theta_{1:L};h_T)p(h_T|\pi)} \left[p(\theta_0)\frac{U(h_T,\theta_0)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{U(h_T,\theta_\ell)p(\theta_\ell)}{q(\theta_\ell;h_T)}}\right] \tag{82}$$

$$= \mathbb{E}_{q(\theta_{0:L};h_T)p(h_T|\pi)} \left[\frac{\frac{U(h_T,\theta_0)p(\theta_0)}{q(\theta_0;h_T)}}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{U(h_T,\theta_\ell)p(\theta_\ell)}{q(\theta_\ell;h_T)}}\right] \tag{83}$$

by symmetry

$$= \mathbb{E}_{q(\theta_{0:L};h_T)p(h_T|\pi)} \left[\frac{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{U(h_T,\theta_\ell)p(\theta_\ell)}{q(\theta_\ell;h_T)}}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{U(h_T,\theta_\ell)p(\theta_\ell)}{q(\theta_\ell;h_T)}}\right] \tag{84}$$

$$= 1. \tag{85}$$

With the optimal critic we recover the sACE bound from [17], which under mild conditions converges to the mutual information $\mathcal{I}_T(\pi)$. To see that start by writing

$$\mathcal{L}_T^{\text{sLACE}}(\pi, U^*, q; L) = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)q(\theta_{1:L};h_T)} \left[\log \frac{p\left(h_T|\theta_0,\pi\right)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{q(\theta_\ell;h_T)}}\right]. \tag{86}$$

The denominator is a consistent estimator of the marginal, provided that each term in the sum is bounded, and so by the Strong Law of Large Numbers we have

$$\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p\left(h_T|\theta_\ell,\pi\right)p\left(\theta_\ell\right)}{q\left(\theta_\ell;h_T\right)} \to p(h_T|\pi)\ a.s. \quad \text{as } L\to\infty, \tag{87}$$

which establishes point-wise convergence of the integrand to $p(h_T|\theta_0,\pi)/p(h_T|\pi)$. We can apply Bounded convergence theorem to establish $\mathcal{L}_T^{\text{sACE}}(\pi, U^*, q; L) \to \mathcal{I}_T(\pi)$ as $L \to \infty$.

If in addition $q(\theta; h_T) = p(\theta|h_T)$ we have by Bayes rule:

$$\mathcal{L}_T^{\text{sLACE}}(\pi, U^*, q; L) = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L}|h_T)} \left[\log \frac{p(h_T|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(h_T|\theta_\ell,\pi)p(\theta_\ell)}{p(\theta_\ell|h_T)}}\right] \tag{88}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L}|h_T)} \left[\log \frac{p(h_T|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}p(h_T|\pi)}\right] \tag{89}$$

$$= \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)} \left[\log \frac{p(h_T|\theta_0,\pi)}{p(h_T|\pi)}\right] \tag{90}$$

$$= \mathcal{I}_T(\pi) \quad \forall L \geq 0. \tag{91}$$

□

In practice, we parameterize the policy, the critic and the density of the proposal distribution by neural networks π_ϕ, U_ψ and q_ζ and optimize $\mathcal{L}_T^{\text{sLACE}}$ with respect to the parameters of these networks, ϕ, ψ and ζ with SGA. As before, optimizing with respect to ϕ improves the quality of the designs, proposed by the policy, whilst optimizing with respect to ψ and ζ tightens the bound. If the parametric density q_ζ and the critic U_ψ are expressive enough, so that we can recover the optimal critic and the true posterior, then the bound is tight for any number of contrastive samples L. If, on the other hand, we fix $q_\zeta(\theta; h_T) = p(\theta)$ instead of training it, then we recover the InfoNCE bound. Therefore, as long as q_ζ approximates the posterior better than the prior, then even an imperfect proposal q_ζ can benefit training.

In addition to introducing another set of optimizable parameters, ζ, the sLACE bound assumes that we know the prior $p(\theta)$ and can evaluate its density.

C Neural architecture

C.1 Permutation invariance of the critic for exchangeable experiments

We show that if the BOED problem is exchangeable then the critic function U should be permutation-invariant.

Proposition 5 (Permutation invariance)**.** *Let σ be a permutation acting on a history h_T^1 yielding $h_T^2 = \{(\xi_{\sigma(i)}, y_{\sigma(i)})\}_{i=1}^T$. If the data generating process is conditionally independent of its past given θ, then the optimal critics for both (7) and (8) are invariant under permutations of the history, i.e.*

$$p(\theta)\prod_{t=1}^T p(y_t|\theta, \xi_t(h_{t-1}), h_{t-1}) = p(\theta)\prod_{t=1}^T p(y_t|\theta, \xi_t) \implies U^*(h_T^1, \theta) = U^*(h_T^2, \theta). \tag{92}$$

Proof. This is a direct consequence from the form of the optimal critics. To see this formally, let h_T^1 be a history and h_T^2 be a permutation of it.

Starting with the InfoNCE bound we have

$$\begin{aligned}
U^*_{\text{NCE}}(h_T^1, \theta) &= \log p(h_T^1|\theta, \pi) + c(h_T^1) && (93)\\
&= \log \prod_{t=1}^T p(y_t|\theta, \xi_t) + c(\{(\xi_t, y_t)\}_{t=1}^T) && (94)
\end{aligned}$$

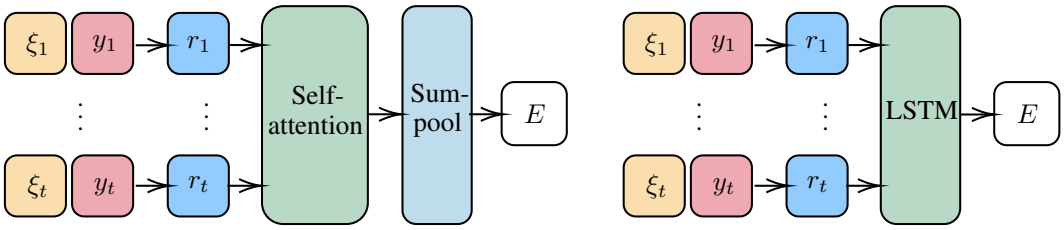
since $c(h_T)$ is arbitrary, we can choose it to be permutation invariant

$$\begin{aligned}
&= \log \prod_{t=1}^T p(y_{\sigma(t)}|\theta, \xi_{\sigma(t)}) + c(\{(\xi_{\sigma(t)}, y_{\sigma(t)})\}_{t=1}^T) && (95)\\
&= \log p(h_T^2|\theta, \pi) + c(h_T^2) && (96)\\
&= U^*_{\text{NCE}}(h_T^2, \theta) && (97)
\end{aligned}$$

Similarly, for the optimal critic of the NWJ bound we have

$$\begin{aligned}
U^*_{\text{NWJ}}(h_T^1, \theta) &= \log \frac{p(h_T^1|\theta, \pi)}{p(h_T^1|\pi)} + 1 && (98)\\
&= \log \frac{\prod_{t=1}^T p(y_t|\theta, \xi_t)}{\mathbb{E}_{p(\theta)}\left[\prod_{s=1}^T p(y_s|\theta, \xi_s)\right]} + 1 && (99)\\
&= \log \frac{\prod_{t=1}^T p(y_{\sigma(t)}|\theta, \xi_{\sigma(t)})}{\mathbb{E}_{p(\theta)}\left[\prod_{s=1}^T p(y_{\sigma(s)}|\theta, \xi_{\sigma(s)})\right]} + 1 && (100)\\
&= \log \frac{p(h_T^2|\theta, \pi)}{p(h_T^2|\pi)} + 1 = U^*_{\text{NWJ}}(h_T^2, \theta). && (101)
\end{aligned}$$

□



(a) Exchangeable models (b) Non-exchangeable models

Figure 6: History encoder architectures for different classes of models. When conditional independence of the experiments holds, we use self-attention, followed by sum-pooling, making the history encoder permutation invariant. When experiments are not conditionally independent we use LSTM and only keep its last hidden state. We train two separate history encoders—one for the design network π_ϕ and one for the critic network U_ψ, although we note that all the weights except those in the head layers can be shared.

To the best of our knowledge, we are the first to propose a critic architecture that is tailored to BOED problems with exchangeable models. Previous work in the static BOED setting, where MI information objective is optimized with variational lower bounds and thus require the training of critics [e.g. 28, 63], did not discuss what an appropriate critic architecture might be. In particular, in all experiments [28, 63] use a generic architecture for both exchangeable and non-exchangeable problems. An expressive enough generic architecture should be able to obtain the optimal critic, and thus achieve a tight bound, however, the optimisation process will be considerably more difficult as the network needs to learn this key invariance structure. We therefore recommend using permutation invariant architectures whenever the model is exchangeable, especially if achieving tight bounds (and therefore learning an optimal critic) is of importance.

C.2 Further details on the history encoder

Figure 6 shows the history encoders we use in the policy network π_ϕ and the critic network U_ψ. First, we encode the individual design-outcome pairs, (ξ_t, y_t), with an MLP, which gives us a vector of representations $r_t \in \mathbb{R}^m$, where m is the encoding dimension we have selected. The representations $\{r_i\}_{i=1}^t$ are row-stacked into a matrix R of dimension $t \times m$, which we then aggregate back to a vector of size m by an appropriate layer(s).

When conditional independence of the experiments holds, we apply 8-head self-attention, based on the Image Transformer [40] and as implemented by [14]. Applying self-attention leaves the dimension of the matrix R unchanged. We then apply sum-pooling across time t, which gives us the final encoding vector $E \in \mathbb{R}^m$.

When experiments are not conditionally independent, we pass the matrix R though an LSTM with two hidden layers and hidden state of size m (see the LSTM module in Pytorch for more details). The LSTM returns hidden state vectors associated with the history h_t for each t; we keep the last hidden state of the last layer, which is our final encoding vector $E \in R^m$.

In both cases the resulting encoding E is a vector of size m. It is passed through final fully connected "head" layers, which output either a design (in the case of the policy) or a vector (in the case of the critic). We train two separate history encoders—one for the design network π_ϕ and one for the critic network U_ψ, although we note that all the weights except those in the head layers can be shared.

D Experiments

D.1 Computational resources

All of the experiments were implemented in Python using open-source software. All estimators and models were implemented in PyTorch [41] (BSD license) and Pyro [6] (Apache License Version 2.0), whilst MlFlow [61] (Apache License Version 2.0) was used for experiment tracking and management. The self-attention architecture from [14] was used to implement the self-attention mechanisms in the

design and critic networks. For full details on package versions, environment set-up and commands for running the code, see instructions in the `README.md` file.

Experiments were ran on internal GPU clusters, consisting of GeForce RTX 3090 (24GB memory), GeForce RTX 2080 Ti (11GB memory) and GeForce GTX 1080 Ti GPUs (11GB memory).

The deployment-time of iDAD (Table 4) was estimated on a lightweight CPU machine with the following specifications

Processor	2.8 GHz Quad-Core Intel Core i7
Memory	16 GB
Operating system	macOS Big Sur v11.2.3

D.2 CO2 Emission Related to Experiments

Experiments were conducted using a private infrastructure, which has an estimated carbon efficiency of 0.432 kgCO_2eq/kWh. A cumulative of 160 hours of computation was performed on hardware of type RTX 2080 Ti (TDP of 250W), or similar. The training time of each experiment (including the baselines that require optimization), took on average between 1-3 GPU hours, depending on the number of experiments T.

Total emissions are estimated to be 17.28 kgCO_2eq of which 0% was directly offset.

Estimations were conducted using the Machine Learning Impact calculator presented in [31].

D.3 Traditional sequential BOED with variational posterior estimator

The variational posterior estimator from [15] is based on the Barbar-Agakov lower bound [4], which takes the form

$$\mathcal{L}^{\text{post}}(\xi, q_\psi) = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[\log \frac{q_\psi(\theta; y, \xi)}{p(\theta)}\right] \leq \mathcal{I}(\xi), \tag{102}$$

where q_ψ is any normalized distribution over the parameters θ. The bound is tight when $q_\psi(\theta; y, \xi) = p(\theta|y, \xi)$, i.e. if we can recover the true posterior. We assume mean-field variational family and optimize the parameters ψ by maximizing the bound (102) using stochastic gradient schemes. Simultaneously we optimize the bound with respect to the design variable ξ to select the optimal design ξ^*. At the inference stage, denoting by y^* the outcome of experiment ξ^*, we obtain an approximate posterior by evaluating $q_\psi(\theta; y^*, \xi^*)$, i.e. we reuse the learnt variational posterior. We repeat this process at each stage of the experiments by substituting the the approximate posterior, $q_\psi(\theta; y^*, \xi^*)$, as the prior in (102).

D.4 Location Finding

In this experiment we have K hidden objects (*sources*) in $\mathbb{R}^2$ and we wish to learn their locations, $\theta = \{\theta_1, \ldots, \theta_K\}$. The number of sources, K, is assumed to be known. Each source emits a signal with intensity obeying the inverse-square law. Put differently, if a source is located at θ_k and we perform a measurement at a point ξ, the signal strength emitted from that source only will be proportional to $\frac{1}{\|\theta_k - \xi\|^2}$. The total intensity at location ξ, emitted from all K sources, is a superposition of the individual ones

$$\mu(\theta, \xi) = b + \sum_{k=1}^{K} \frac{\alpha_k}{m + \|\theta_k - \xi\|^2}, \tag{103}$$

where α_k can be known constants or random variables, $b > 0$ is a constant background signal and m is a constant, controlling the maximum signal.

We place a standard normal prior on each of the location parameters θ_k and we observe the log-total intensity with some Gaussian noise. We therefore have the following prior and likelihood:

$$\theta_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, I_d) \quad \log y \mid \theta, \quad \xi \sim \mathcal{N}(\log \mu(\theta, \xi), \sigma^2) \tag{104}$$

D.4.1 Training details

All our experiments are performed with the following model hyperparameters

Parameter	Value
Number of sources, K	2
α_k	$1\ \forall k$
Max signal, m	10^{-4}
Base signal, b	10^{-1}
Observation noise scale, σ	0.5

The architecture of the design network π_ϕ used in Table 2 and all its hyperparameters are in the following tables. For the encoder of the design-outcome pairs we used the following:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	ξ, y	3	3	-
H1	Fully connected	64	64	ReLU
H2	Fully connected	512	512	ReLU
Output	Fully connected	64	64	-
Attention	8 heads	64	64	-

The output of the encoder, $R(h_t)$, is fed into an emitter network, for which we used the following:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	$R(h_t)$	64	64	-
H1	Fully connected	256	256	ReLU
H2	Fully connected	64	64	ReLU
Output	Fully connected	2	2	-

The architecture of the critic network U_ψ used in Table 2 and all its hyperparameters are in the tables that follow. First, the encoder network of the latent variables is:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	θ	4	4	-
H1	Fully connected	16	16	ReLU
H2	Fully connected	64	64	ReLU
H3	Fully connected	512	512	ReLU
Output	Fully connected	64	64	-

For the design-outcome pairs encoder we use the same architecture as in the design network, namely:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	ξ, y	3	3	-
H1	Fully connected	64	64	ReLU
H2	Fully connected	512	512	ReLU
Output	Fully connected	64	64	-
Attention	8 heads	64	64	-

The output of the encoder, $R(h_t)$, is fed into fully connected head layers:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	$R(h_t)$	64	64	-
H1	Fully connected	1024	1024	ReLU
H2	Fully connected	512	512	ReLU
H3	Fully connected	512	512	ReLU
Output	Fully connected	64	64	-

The optimisation was performed with Adam [26] with `ReduceLROnPlateau` learning rate scheduler, with the following hyperparameters:

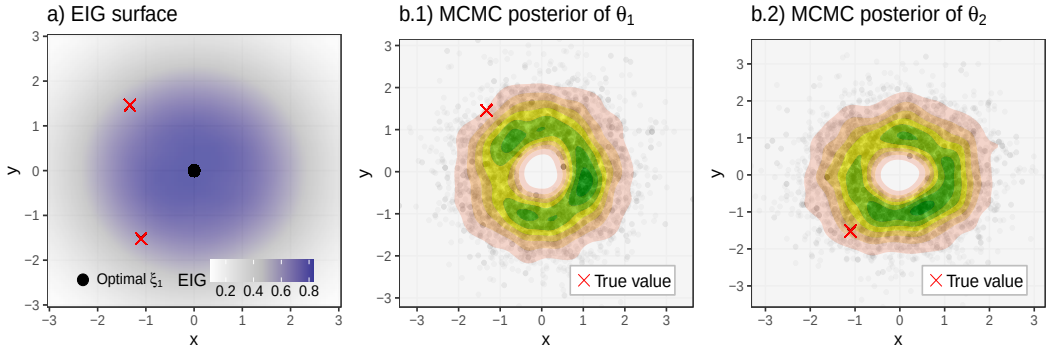


Figure 7: a): EIG surface induced by the prior; b) Samples from $p(\theta|\xi_1, y_1)$—the posterior distribution of the locations, after performing experiment ξ_1 and observing y_1, along with a KDE.

Parameter	iDAD, InfoNCE	iDAD, NWJ
Batch size	2048	2048
Number of contrastive/negative samples	2047	2047
Number of gradient steps	100000	100000
Initial learning rate (LR)	0.0005	0.0005
LR annealing factor	0.8	0.8
LR annealing frequency (if no improvement)	2000	2000

D.4.2 Performance of the variational baseline

As we saw in Table 2, this variational approach to (myopic) adaptive BOED performed very poorly, despite its large computational budget. The likely reason for that is that the mean-field variational approximation cannot adequately capture the complex non-Gaussian posterior of this problem. Figure 7 clearly demonstrates this: before any data is observed it is optimal to sample at the origin (since the prior is centered at it). After observing a low signal (the locations in this example are not close to the origin), we can only conclude that the sources are not within a small radius of the origin, but anywhere outside of it would be a plausible location, as indeed indicated by the fitted posteriors.

D.4.3 Hyperparameter selection

We did not perform extensive hyperparameters search; in particular, the network sizes were guided by two hyperparametes: hidden-dimension ($HD = 512$) and encoding dimension ($ED = 64$). We set-up all the networks to scale up with the number of experiments as follows:

- Design-outcome encoder has three hidden layers of sizes $[64, HD, ED]$.
- Design emitter network has three hidden layers of sizes $[HD/2, ED, 2]$, where 2 is the dimension of the design variable.
- The latent encoder for the critic network has four hidden layers of sizes $[16, 64, HD, ED]$.
- The critic design-outcome encoder's head layer has four hidden layers of sizes $[HD \times \log(T), HD \times \log(T)/2, HD, ED]$.

Since our multi-head attention layer has 8 heads, the encoding dimension we use has to be a multiple of 8. In addition to $ED = 64$ we tried $ED = 32$ which provided marginally worse results. We did not try other values for these hyperparameters.

For the learning rate, we tried 0.001, which was too high, as well as 0.0005 (which we selected) and 0.0001 (which yielded very similar results).

We performed similar level of hyperparameter tuning for all trainable baselines as well (DAD, MINEBED and SG-BOED).

Table 6: Upper bounds on the total information, $\mathcal{I}_{10}(\pi)$, for the location finding experiment in Section 5.1. The bounds were estimated using $L = 5 \times 10^5$ contrastive samples. Errors indicate ± 1 s.e. estimated over 4096 histories (128 for variational). Lower bounds are presented in Table 2.

Method \ θ dim.	4D	6D	10D	20D
Random	4.794 ± 0.041	3.506 ± 0.004	1.895 ± 0.003	0.552 ± 0.001
MINEBED	5.522 ± 0.028	4.229 ± 0.029	2.459 ± 0.029	0.801 ± 0.019
SG-BOED	5.549 ± 0.028	4.220 ± 0.030	2.455 ± 0.029	0.803 ± 0.019
Variational	4.644 ± 0.146	3.626 ± 0.167	2.181 ± 0.152	0.669 ± 0.097
iDAD (NWJ)	7.806 ± 0.050	5.851 ± 0.041	$\mathbf{3.264 \pm 0.039}$	0.877 ± 0.022
iDAD (InfoNCE)	$\mathbf{7.863 \pm 0.043}$	$\mathbf{6.068 \pm 0.039}$	$\mathbf{3.257 \pm 0.040}$	$\mathbf{0.872 \pm 0.020}$
DAD	8.034 ± 0.038	6.310 ± 0.031	3.358 ± 0.040	0.953 ± 0.022

Table 7: Upper and lower bounds on the total information, $\mathcal{I}_{20}(\pi)$, for the location finding experiment in 2D from Section 5.1. The bounds were estimated using $L = 5 \times 10^5$ contrastive samples. Errors indicate ± 1 s.e. estimated over 4096 histories.

Method	Lower bound	Upper bound
Random	7.000 ± 0.034	7.020 ± 0.034
MINEBED	7.672 ± 0.030	7.690 ± 0.031
SG-BOED	7.701 ± 0.030	7.728 ± 0.031
iDAD (NWJ)	$\mathbf{9.961 \pm 0.033}$	$\mathbf{10.372 \pm 0.048}$
iDAD (InfoNCE)	$\mathbf{10.075 \pm 0.032}$	$\mathbf{10.463 \pm 0.043}$
DAD	10.424 ± 0.031	10.996 ± 0.049

D.4.4 Further ablation studies

Scalability with number of experiments. We first demonstrate that iDAD can scale to a larger number of experiments T. We train policy networks to perform $T = 20$ experiments and compare them to baselines in Table 7. We omit the variational baseline as it is too computationally costly to run for a large enough number of histories, and as we saw in the previous subsection, it is not particularly suited to this model.

Training stability. To assess the robustness of the results and the stability of the training process, we trained 5 additional iDAD networks with each of the two bounds, using different seeds but the same hyperparameters (described in Subsection D.4.1) we used to produce the results of the location finding experiment in 2D (Table 2 in the main text). We report upper and lower bounds on the mutual information along with their mean and standard error in the table below.

Estimator	Bound	Run 1	Run 2	Run 3	Run 4	Run 5	**Mean**	**SE**
InfoNCE	Lower	7.826	7.682	7.856	7.713	7.804	**7.776**	**0.034**
InfoNCE	Upper	7.933	7.791	7.856	7.807	7.925	**7.862**	**0.029**
NWJ	Lower	7.820	7.545	7.592	7.555	7.691	**7.641**	**0.052**
NWJ	Upper	7.976	7.640	7.669	7.651	7.800	**7.747**	**0.064**

We can see that the iDAD networks trained with InfoNCE are highly stable, with the 5 additional runs achieving very similar mutual information values to each other and to the iDAD network used the report the results in the main paper. The performance of the iDAD networks trained with the NWJ bound is more variable and empirically achieve slightly lower average value of mutual information. This higher variance is in-line with the discussion in Section B.

We similarly verify the robustness of the static baselines, reporting the results in the table below:

Table 8: Ablation study on the performance of iDAD as a function of training time for the location finding experiment.

Training budget	MI lower bound
0.1%	3.38
1.0%	6.09
2.0%	6.46
4.0%	6.81
8.0%	7.08
16.0%	7.33
32.0%	7.56
64.0%	7.78
100.0%	7.82

Estimator	Bound	Run 1	Run 2	Run 3	Run 4	Run 5	**Mean**	**SE**
SG-BOED	Lower	5.537	5.536	5.473	5.523	5.518	**5.517**	**0.013**
SG-BOED	Upper	5.553	5.548	5.491	5.541	5.531	**5.533**	**0.012**
MINEBED	Lower	5.460	5.506	5.553	5.539	5.565	**5.524**	**0.021**
MINEBED	Upper	5.473	5.526	5.567	5.554	5.574	**5.540**	**0.022**

Performance sensitivity to errors in the policy. Finally, we investigate the effect of slight errors in the design policy network. To this end, we look at the performance achieved by partially trained design networks (there will be some errors or inaccuracies in networks that were not trained until convergence). Table 8 shows the performance of iDAD as a function of training time, demonstrating that small errors in the network only lead to small drops in performance.

In detail, our results show that with just 8% of the total training budget, this slightly inaccurate network still performs relatively well, achieving total mutual information of 7.1, compared to the fully trained network that reached 7.8. We also highlight that iDAD outperforms all baselines with as little as 1% of the total training budget (the best performing baseline achieves mutual information of 5.5, see Table 2).

D.5 PK model

The drug concentration z, measured ξ hours after administering it, and the corresponding noisy observation y are given by

$$z(\xi;\theta) = \frac{D_V}{V}\frac{k_\alpha}{k_\alpha - k_e}[e^{-k_e\xi} - e^{-k_\alpha\xi}], \quad y(\xi;\theta) = z(\xi;\theta)(1+\epsilon) + \eta \tag{105}$$

where $\theta = (k_\alpha, k_e, V)$, $D_V = 400$ is a constant, $\epsilon \sim \mathcal{N}(0, 0.01)$ is multiplicative noise to account for heteroscedasticity and $\eta \sim \mathcal{N}(0, 0.1)$ is an additive observation noise. Since both noise sources are Gaussian, the observation likelihood is also Gaussian i.e.

$$y(\xi;\theta) \sim \mathcal{N}(z(\xi;\theta), 0.01z(\xi;\theta)^2 + 0.1) \tag{106}$$

The prior for the parameters θ that we used

$$\log\theta \sim \mathcal{N}\left(\begin{bmatrix}\log 1\\ \log 0.1\\ \log 20\end{bmatrix}, \begin{bmatrix}0.05 & 0 & 0\\ 0 & 0.05 & 0\\ 0 & 0 & 0.05\end{bmatrix}\right) \tag{107}$$

D.5.1 Training details

The architecture of the design network π_ϕ used for Figure 3 and 4 and all its hyperparameters are in the following tables. For the encoder of the design-outcome pairs we used the following:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	ξ, y	2	2	-
H1	Fully connected	64	64	ReLU
H2	Fully connected	512	512	ReLU
Output	Fully connected	32	32	-
Attention	8 heads	32	32	-

The outputs of the encoder, $\{R(h_t)\}_{t=1}^{T}$, are summed and the resulting vector (of dimension 32) is fed into an emitter network, for which we used the following:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	$R(h_t)$	32	32	-
H1	Fully connected	256	256	ReLU
H2	Fully connected	32	32	ReLU
Output	Fully connected	1	1	Sigmoid

The architecture of the critic network U_ψ used in Figures 3 and 4 and all its hyperparameters are in the following tables. For the encoder of the design-outcome pairs we used the same architecture as for the design network, namely:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	ξ, y	2	2	-
H1	Fully connected	64	64	ReLU
H2	Fully connected	512	512	ReLU
Output	Fully connected	32	32	-
Attention	8 heads	32	32	-

The resulting pooled representation, $R(h_T)$ is fed into fully connected critic head layers with the following architecture:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	$R(h_T)$	32	32	-
H1	Fully connected	512	512	ReLU
H2	Fully connected	256	256	ReLU
H3	Fully connected	512	512	ReLU
Output	Fully connected	32	32	-

Finally, for the latent variable encoder network we used:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	θ	3	3	-
H1	Fully connected	8	8	ReLU
H2	Fully connected	64	64	ReLU
H3	Fully connected	512	512	ReLU
Output	Fully connected	32	32	-

The optimisation was performed with Adam [26] with the following hyperparameters:

Parameter	iDAD, InfoNCE	iDAD, NWJ
Batch size	1024	1024
Number of contrastive/negative samples	1023	1023
Number of gradient steps	100000	100000
Initial learning rate (LR)	0.0001	0.0001
LR annealing factor	0.8	0.5
LR annealing frequency (if no improvement)	2000	2000

Table 9: Upper and lower bounds on the total information, $\mathcal{I}_5(\pi)$, for the pharmacokinetic experiment. Errors indicate ± 1 s.e. estimated over 4096 (126 for variational) histories and $L = 5 \times 10^5$.

Method	Lower bound	Upper bound	Deployment time
Random	2.523 ± 0.033	2.523 ± 0.033	N/A
Equal interval	2.651 ± 0.022	2.651 ± 0.022	N/A
MINEBED	2.955 ± 0.030	2.956 ± 0.030	N/A
SG-BOED	2.985 ± 0.027	2.985 ± 0.027	N/A
Variational	2.683 ± 0.093	2.683 ± 0.093	$505.4 \pm 1\%$
IDAD (NWJ)	$\mathbf{3.163 \pm 0.023}$	$\mathbf{3.163 \pm 0.023}$	$0.007 \pm 7\%$
IDAD (InfoNCE)	$\mathbf{3.200 \pm 0.024}$	$\mathbf{3.200 \pm 0.024}$	$0.007 \pm 8\%$
DAD	3.234 ± 0.023	3.234 ± 0.023	$0.002 \pm 7\%$

Table 10: Upper and lower bounds on the total information, $\mathcal{I}_{10}(\pi)$, for the pharmacokinetic experiment. Errors indicate ± 1 s.e. estimated over 4096 (126 for variational) histories and $L = 5 \times 10^5$.

Method	Lower bound	Upper bound	Deployment time
Random	3.344 ± 0.034	3.345 ± 0.034	N/A
Equal interval	3.422 ± 0.026	3.423 ± 0.026	N/A
MINEBED	3.849 ± 0.034	3.849 ± 0.034	N/A
SG-BOED	3.824 ± 0.034	3.824 ± 0.034	N/A
Variational	3.624 ± 0.099	3.624 ± 0.099	$1055.2 \pm 8\%$
IDAD (NWJ)	$\mathbf{4.034 \pm 0.025}$	$\mathbf{4.034 \pm 0.025}$	$0.007 \pm 6\%$
IDAD (InfoNCE)	$\mathbf{4.045 \pm 0.026}$	$\mathbf{4.045 \pm 0.026}$	$0.007 \pm 5\%$
DAD	4.116 ± 0.024	4.117 ± 0.024	$0.007 \pm 8\%$

D.5.2 Hyperparameter selection

Hyperparameter selection was done in a way similar to the Location Finding experiment (see D.4.3). We tried encodin dimensions $ED = 32, 64$ and selected the smaller size as there were no clear benefits to larger networks (relatively speaking, this is an easier model that the location finding). We used the same hidden dimension, i.e. $HD = 512$. In terms of learning rates, we tried 0.0001, 0.0005 and 0.001; we found 0.0001 to be appropriate, although NWJ bound was exhibiting high variance, so used a smaller learning rate annealing factor for that network (0.5 vs 0.8 for InfoNCE). We performed similar level of hyperparameter tuning for all trainable baselines as well (DAD, MINEBED and SG-BOED).

D.5.3 Further results

Table 9 reports the results shown in Figure 3c), along with the corresponding upper bounds and deployment times, while Table 10 reports the results for $T = 10$.

Training stability. To assess the robustness of the results and the stability of the training process, we trained 5 additional iDAD networks with each of the two bounds, using different seeds but the same hyperparameters we used to produce the results of the pharmacokinetic experiment (Figure 3c) and corresponding Table 9). We report upper and lower bounds on the mutual information along with their mean and standard error in the table below.

Method	Bound	Run 1	Run 2	Run 3	Run 4	Run 5	**Mean**	**SE**
iDAD, InfoNCE	Lower	3.209	3.165	3.198	3.221	3.128	**3.185**	**0.019**
iDAD, InfoNCE	Upper	3.210	3.166	3.201	3.223	3.130	**3.186**	**0.019**
iDAD, NWJ	Lower	3.034	3.049	2.608	3.149	3.082	**3.034**	**0.107**
iDAD, NWJ	Upper	3.034	3.049	2.609	3.150	3.083	**3.034**	**0.107**

We repeat the same procedure for the static baselines. The results reported in the table below demonstrate the training stability of these baselines as well.

Method	Bound	Run 1	Run 2	Run 3	Run 4	Run 5	**Mean**	**SE**
SG-BOED	Lower	2.932	2.452	2.448	2.991	2.962	**2.757**	**0.140**
SG-BOED	Upper	2.932	2.453	2.449	2.992	2.962	**2.757**	**0.140**
MINEBED	Lower	2.912	2.213	3.014	2.092	2.941	**2.634**	**0.221**
MINEBED	Upper	2.914	2.213	3.015	2.092	2.942	**2.635**	**0.222**

D.6 SIR Model

Generally speaking, the SIR model advocates that, within a fixed population of size N, susceptible individuals $S(\tau)$, where τ is time, can become infected and move to an infected state $I(\tau)$. The infected individuals can then recover from the disease and move to the recovered state $R(\tau)$. The dynamics of these events are governed by the infection rate β and recovery rate γ, which define the particular disease in question. In the context of BOED, the aim is generally to estimate these two model parameters by observing state populations at particular measurement times τ, which are the experimental design variables. The SIR model has been studied extensively in the context of BOED, e.g. in [12, 27, 29, 30].

Stochastic versions of the SIR model are usually formulated via continuous-time Markov chains (CTMC), which can be simulated from via the Gillespie algorithm [2], yielding discrete state populations. However, iDAD requires us to differentiate through the sampling path of the state populations to the experimental designs, which is impossible if the simulated data is discrete as gradients are undefined. Thus, we here implement an alternative formulation of the stochastic SIR model that is based on stochastic differential equations (SDEs), as studied in [29], which yields continuous state populations that can be differentiated.

Following [29], let us first define a state population vector $\mathbf{X}(\tau) = (S(\tau), I(\tau))^\top$, where we can safely ignore the population of recovered $R(\tau)$ for modelling purposes because we assume that the total population stays fixed. The system of Itô SDEs that defines the stochastic SIR model is given by

$$\mathrm{d}\mathbf{X}(\tau) = \mathbf{f}(\mathbf{X}(\tau))\mathrm{d}\tau + \mathbf{G}(\mathbf{X}(\tau))\mathrm{d}\mathbf{W}(\tau), \tag{108}$$

where $\mathbf{f}$ is a drift vector, $\mathbf{G}$ is a diffusion matrix and $\mathbf{W}(\tau)$ is a vector of independent Wiener processes. [29] showed that the drift vector and diffusion matrix are given by

$$\mathbf{f}(\mathbf{X}(\tau)) = \begin{pmatrix} -\beta \frac{S(\tau)I(\tau)}{N} \\ \beta \frac{S(\tau)I(\tau)}{N} - \gamma I(\tau) \end{pmatrix} \quad \text{and} \quad \mathbf{G}(\mathbf{X}(t)) = \begin{pmatrix} -\sqrt{\beta \frac{S(\tau)I(\tau)}{N}} & 0 \\ \sqrt{\beta \frac{S(\tau)I(\tau)}{N}} & -\sqrt{\gamma I(\tau)} \end{pmatrix}. \tag{109}$$

Given the system of Itô SDEs in (108), as well as the above drift vector and diffusion matrix, we can then simulate state populations $\mathbf{X}(\tau)$ by solving the SDE using finite-difference methods, such as e.g. the Euler-Maruyama method. See [29] for more information on the SDE-based SIR model, including derivations of the drift vector and diffusion matrix.

Importantly, we note that [29] further used the solutions of (108) as an input to a Poisson observation model, which increases the noise in simulated data. We here opt to simply use the solutions of (108) as data and do not consider an additional Poisson observational model.

D.6.1 Training details

As previously mentioned, the design variable for this model is the measurement time $\tau \in [0, 100]$. When solving the SDE with the Euler-Maruyama method, we discretize the time domain with a resolution of $\Delta\tau = 10^{-2}$. We here only use the number of infected $I(\tau)$ as the observed data, as others might be difficult to measure in reality. The total population is fixed at $N = 500$ and the initial conditions are $\mathbf{X}(\tau = 0) = (0, 2)^\top$. The model parameters β and γ have log-normal priors, i.e. $p(\beta) = \text{Lognorm}(0.50, 0.50^2)$ and $p(\gamma) = \text{Lognorm}(0.10, 0.50^2)$. Importantly, because solving SDEs is expensive, we pre-simulate our data on a time grid, store it in memory and then access the relevant data during training.

We present the network architectures and hyper-parameters corresponding to the $T = 5$ iDAD results shown in Table 5 of the main text. For the encoder of the design-outcome pairs we used:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	ξ, y	2	2	-
H1	Fully connected	8	8	ReLU
H2	Fully connected	64	64	ReLU
H3	Fully connected	512	512	ReLU
Output	Fully connected	32	32	-

The resulting representations, $\{R(h_t)\}_{t=1}^{T-1}$, are stacked into a matrix (as new design–outcome pairs are obtained) and fed into an emitter network, which contains an LSTM cell with two hidden layers. We only keep the last hidden state of the LSTM's output and pass it through a final FC layer:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	$\{R(h_t)\}_{t=1}^{T-1}$	$32 \times$ t	$32 \times$ t	-
H1 & H2	LSTM	32	32	-
H3	Fully connected	16	16	ReLU
Output	Fully connected	1	1	-

The architecture of the critic network U_ψ used in Table 5 and all its hyper-parameters are in the tables that follow. First, the encoder network of the latent variables is:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	θ	2	2	-
H1	Fully connected	8	8	ReLU
H2	Fully connected	64	64	ReLU
H3	Fully connected	512	512	ReLU
Output	Fully connected	32	32	-

For the design-outcome pairs encoder we use the same architecture as in the design network, namely:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	ξ, y	2	2	-
H1	Fully connected	8	8	ReLU
H2	Fully connected	64	64	ReLU
H3	Fully connected	512	512	ReLU
Output	Fully connected	32	32	-

The outputs of the encoder, $\{R(h_t)\}_t$, are stacked and fed into an LSTM cell with two hidden layers. We only keep the last hidden state of the LSTM's output and pass it through a FC layer:

Layer	Description	iDAD, InfoNCE	iDAD, NWJ	Activation
Input	$\{R(h_t)\}_{t=1}^{T-1}$	$32 \times$ t	$32 \times$ t	-
H1 & H2	LSTM	32	32	-
H3	Fully connected	16	16	ReLU
Output	Fully connected	32	32	-

The optimization was performed with Adam [26] with learning rate annealing with the following hyper-parameters:

Parameter	iDAD InfoNCE	iDAD, NWJ
Batch size	512	512
Number of contrastive/negative samples	511	511
Number of gradient steps	100000	100000
Initial learning rate (LR)	0.0005	0.0005
LR annealing factor	0.96	0.96
LR annealing frequency	1000	1000

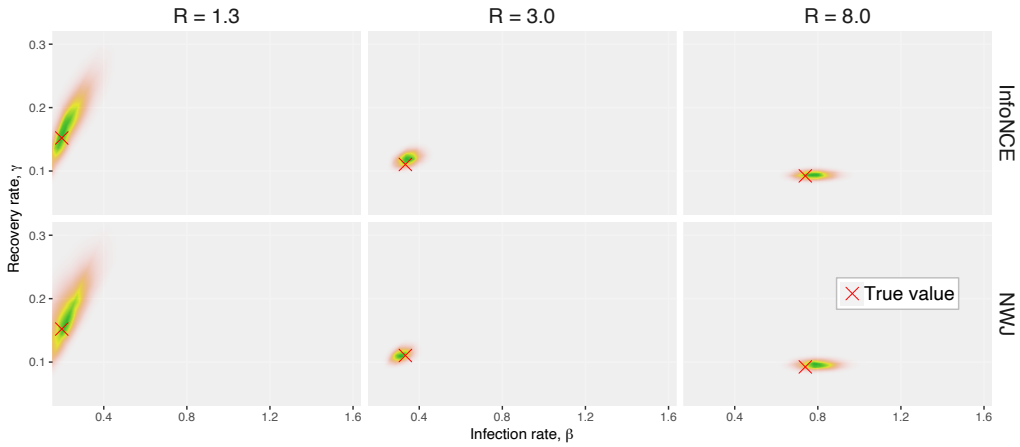


Figure 8: Approximate posteriors for the SIR model.

D.6.2 Further results

Different number of experiments T. In Table 11 we show lower bound estimates when applying iDAD with the InfoNCE lower bound to the SDE-based SIR model for different number of measurements T. The design network and critic architectures are the same as for $T = 5$. Table 11 shows that more measurements yield higher expected information gains, as one might intuitively expect. Furthermore, the increase in expected information gain saturates with increasing T, which is why we presented the results for $T = 5$ in the main text. The biggest increase, however, occurs from $T = 1$ to $T = 2$. This is intuitive, because the SIR model has two model parameters that we wish to estimate but we only gather one data point with one measurement. Hence, in order to accurately estimate both of these parameters, we would need at least 2 measurements, which is reflected in Table 11. We note that all of these numbers, with the exception of $T = 1$, are larger than those found by [29]. This increase in expected information gain may be explained by the fact that [29] use an additional Poisson observation model, which means that the resulting data are inherently noisier and less informative.

Table 11: InfoNCE lower bound estimates ($\pm$ s.e.) when applying iDAD to the SDE-based SIR model for different number of measurements T.

T	iDAD, InfoNCE	iDAD, NWJ
1	1.396 ± 0.018	1.417 ± 0.001
2	2.714 ± 0.019	2.699 ± 0.001
3	3.554 ± 0.021	3.515 ± 0.001
4	3.600 ± 0.018	3.749 ± 0.001
5	3.915 ± 0.020	3.869 ± 0.001
7	4.027 ± 0.019	3.911 ± 0.001
10	4.100 ± 0.020	4.019 ± 0.001

Training stability. To assess the robustness of the results and the stability of the training process, we trained 5 additional iDAD networks with each of the two bounds, using different seeds but the same hyperparameters we used to produce the results of Table 5 in the main text. We report upper and lower bounds on the mutual information along with their mean and standard error in the table below.

Method	Bound	Run 1	Run 2	Run 3	Run 4	Run 5	**Mean**	**SE**
iDAD, InfoNCE	Lower	3.900	3.919	3.919	3.901	3.887	**3.906**	**0.007**
iDAD, NWJ	Lower	3.872	3.838	3.854	3.883	3.848	**3.859**	**0.009**

We repeat the same procedure for the static baselines. The results reported in the table below demonstrate the training stability of these baselines as well.

Method	Bound	Run 1	Run 2	Run 3	Run 4	Run 5	**Mean**	**SE**
SG-BOED	Lower	3.713	3.765	3.767	3.764	3.739	**3.749**	**0.012**
MINEBED	Lower	3.373	3.438	3.376	3.379	3.420	**3.397**	**0.015**

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Implicit Deep Adaptive Design: Policy-Based Experimental Design without Likelihoods
Publication Status	Accepted for Publication
Publication Details	Desi R. Ivanova, Adam Foster Steven Kleinegesse, Michael U. Gutmann, Tom Rainforth (2021). Implicit Deep Adaptive Design: Policy-Based Experimental Design without Likelihoods. 35th Conference on Neural Information Processing Systems (NeurIPS 2021) (to appear).

Student Confirmation

Student Name:	Adam Foster		
Contribution to the Paper	Second author. Contributed to the development of the methods and theory of the paper. Contributed significantly to the writing of the paper.		
Signature		Date	25/10/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Dr Tom Rainforth		
Supervisor comments I verify Adam's above account		
Signature 	Date	26/10/21

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 7

Discussion

This discussion is broken into a number of self-contained essays that delve into specific aspects of the concepts laid out in the earlier chapters. In Section 1, we focus on elucidating the ideas of Chapters 2 and 3 by focusing on the specific example use case of Bayesian *model selection*. We examine how the variational estimators of Chapter 2 look in this example, highlighting connections to other work, we also note how the approach of Chapter 3 translates to this specific case. In Section 2, we focus on comparing our work, particularly the estimators of Chapter 3 to work in the field of Bayesian active learning, drawing out the deep connections between experimental design and active learning. In Section 3, we draw another connection, this time between the sequential experiment methods in Chapters 5 and 6 and the field of Bayesian reinforcement learning. We show that reinforcement learning provides a natural language to express the sequential experimental design problem. In Section 4, we investigate some of the statistical properties of various estimators discussed in this thesis. We focus on the NMC estimator, the PCE estimator introduced in Chapter 3. We connect these more basic estimators with MLMC estimators, creating a stronger connection between earlier chapters and Chapter 4. In Section 5, we present new results on mutual information bounds. This can be seen a generalising theory for some of the bounds derived in Chapters 2, 3, 5 and 6.

1 Bayesian experimental design for model selection: variational and classification approaches

1.1 Introduction

Bayesian experimental design for model selection is a important and well-studied problem (Cavagnaro et al., 2010; Vanlier et al., 2014; Hainy et al., 2018). In this essay, we tackle two questions that are relevant to this problem. First, how do recently proposed variational methods for experimental design (Foster et al., 2019, 2020) translate into the model selection context? Second, how do these methods intersect with recently proposed classification-driven approaches to experimental design for model selection (Hainy et al., 2018)?

We begin by elucidating the key features of the model selection problem—it turns out that we can characterise the set-up as a semi-implicit model with a discrete latent variable of interest. The posterior or Barber–Agakov approach of Foster et al. (2019) involves training an amortised inference network from data simulated from the model. We find that, for model selection, this network is exactly a (neural) classifier that predicts the true model that synthesised an observation from that synthetic experimental observation. The marginal + likelihood method of Foster et al. (2019) also translates into the model selection case. This method involves variational density estimation of experimental outcomes for each possible model. In other words, it involves approximating the model evidence of the data for each possible model. Finally, we examine how the stochastic gradient design approach of Foster et al. (2020) applies here. This approach can build off the back of the Barber–Agakov bound, so it also utilises a classifier. The key difference here is that we differentiate the classifier output with respect to its input to learn the design at the same time as the classifier network parameters. This bears some similarities with adversarial approaches to neural network robustness (Carlini et al., 2019). Finally, we compare and contrast the variational approach with other classification driven approaches in the literature.

1.2 Characterising the problem

We denote experimental designs by ξ and experimental observations as y. Suppose there are K competing models $\{m_1, \dots, m_k\}$ and we have a prior distribution $p(m)$ on which model we think is likely to be correct. Given the choice of model, there are other model parameters $\psi \sim p(\psi|m)$. Conditional on the model, and on its parameters, we have a likelihood for the experiment $p(y|m, \psi, \xi)$ which we assume is known in closed form.

One important feature of the model selection problem is that we do *not* have a likelihood that directly relates the design ξ, observation y and the latent variable of interest m. Instead, we have to account for the auxiliary latent variable ψ. Indeed, we actually have $p(y|m, \xi) = \int_\Psi p(y|m, \psi, \xi)p(\psi|m)d\psi$. This case, where we have a closed form likelihood but for a larger set of variable, is referred to as a *semi-implicit* model.

In this essay, we focus on experimental design with the expected information gain (EIG) criterion, also called mutual information utility, that aims to reduce Shannon entropy in our beliefs about m. The EIG-optimal design is specifically,

$$\xi^* = \arg\max_\xi \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)} \left[\log \frac{p(m|y, \xi)}{p(m)} \right]. \tag{1}$$

Finding ξ^* amounts to estimating the EIG objective function and optimising over the space of possible designs.

If we have already observed some data $\mathcal{D} = \{(\xi_1, y_1), \dots, (\xi_T, y_T)\}$, then we fit model-specific posteriors for the auxiliary variable ψ for each model $p(\psi|m, \mathcal{D})$, and we compute the posterior over models $p(m|\mathcal{D}) \propto p(m)p(\mathcal{D}|m)$. Thus, we update our priors $p(m)$ and $p(\psi|m)$ on the basis of past data.

1.3 The variational approach

1.3.1 Posterior lower bound

Foster et al. (2019) considered variational estimation of the EIG. Their general strategy was to optimise variational upper or lower bounds on the EIG. Their simplest bound was the posterior lower bound (also called the Barber–Agakov bound after Barber and Agakov (2003)). With the variables we have in this model, the bound would be expressed as

$$\mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log \frac{p(m|y,\xi)}{p(m)}\right] \geq \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log \frac{q_\phi(m|y)}{p(m)}\right]. \tag{2}$$

The new term $q_\phi(m|y,\xi)$ was generically referred to as the *amortised approximate posterior* with variational parameters ϕ. It is an approximate posterior distribution on the latent variable m of interest. The amortisation here refers to the fact that we learn a function from y to a distribution over m (for different ξ, we would train separate functions). For the model selection approach, then, q_ϕ is a function from y to a distribution over the discrete model indicator m. First, since m is discrete, the choice of variational family is moot, because every distribution over m can be finitely represented. Second, q_ϕ has a very simple interpretation. It is a classifier that attempts to predict, on the basis of input y, which model of $m_1, \ldots, m_k$ generated that data, specifically trying to estimate the posterior probability $p(m|y,\xi)$ over the k different possibilities for m. Importantly though, rather than just attempting to predict the correct model that was responsible for generating the data y, it is essential that we have a *probabilistic* classifier that assigns probabilities to each possible model. For this probabilistic classifier, the issue of calibration becomes central, as we hope that our classifier probabilities will approach $p(m|y,\xi)$ during training.

We have established that q_ϕ is simply a probabilistic classifier for the model selection case. How should this classifier be trained? In general, Foster et al. (2019) proposed training q_ϕ by stochastic gradient methods (Robbins and Monro, 1951; Kingma and Ba, 2014) to maximise the lower bound with respect to ϕ

$$\phi^* = \arg\max_\phi \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log \frac{q_\phi(m|y)}{p(m)}\right] \tag{3}$$

In model selection, training ϕ simply means training the parameters of the classifier. Maximising the posterior lower bound is equivalent to simply maximising the expected log likelihood under q, i.e.

$$\phi^* = \arg\max_\phi \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log q_\phi(m|y)\right]. \tag{4}$$

This is true because $p(m)$ has no dependence on ϕ. So, we see that training q_ϕ to maximise the variational posterior lower bound amounts to maximum likelihood training of a neural classifier when we are in the setting of model selection. (Care may be needed to ensure the classifier produces good *probabilistic uncertainty*, as well as getting good predictions, as these probabilities are central to our method.)

In fact, we have an enhanced setting in which we can draw an infinite amount of training data by simulating from $p(m)p(\psi|m)p(y|m,\psi,\xi)$. To do this, we sample a random model m from its prior, then a random set of parameters $\psi \sim p(\psi|m)$ for the chosen model, and then simulate an experimental outcome under design ξ. Importantly, we do not need to draw a fixed training or test set, and we never need to show the classifier the same examples twice, we instead draw new batches on the fly. One particularly important consequence of this is that the spectre of *over-fitting* is much reduced in our case, as there is no fixed training set to overfit to.

We now see another important point—the negative log-likelihood loss of the classifier is essentially an estimate of the EIG, up to a constant. Suppose we have completed training and reached parameters $\hat{\phi}$. Then the EIG estimate is

$$\mathrm{EIG}(\xi) \approx \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log \frac{q_{\hat{\phi}}(m|y)}{p(m)}\right] = \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log q_{\hat{\phi}}(m|y)\right] + H[p(m)] \tag{5}$$

and we can estimate the expectation with new, independent batches simulated from the model.

In summary, the posterior lower bound method for model selection amounts to training a classifier on (infinite) simulated data to predict m from y. The optimal design ξ^* will be approximated by the classifier which has the best (lowest) validation loss, which is a good approximation of having the highest EIG.

1.3.2 Marginal + likelihood estimator

The posterior lower bound is not the only way to estimate the EIG proposed by Foster et al. (2019). Both the marginal and the VNMC methods require an explicit likelihood, so they are not suitable for the semi-implicit model selection scenario. The marginal + likelihood estimator is

$$\mathrm{EIG}(\xi) \approx \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log \frac{q_\ell(y|m,\xi)}{q_p(y|\xi)}\right]. \tag{6}$$

This estimator translates, with some simplification, into the model selection setting. The 'approximate likelihood' $q_\ell(y|m,\xi)$ in the model selection setting is an approximation of the model evidence $q_\ell(y|m,\xi) \approx p(y|m,\xi)$. For model selection when m is discrete, we do not need to separately estimate q_p and q_ℓ, we can instead sum over m to obtain

$$q_p(y|\xi) = \sum_m p(m) q_\ell(y|m,\xi). \tag{7}$$

As shown in Appendix A.4 of Foster et al. (2019), the estimator actually becomes a lower bound

$$\mathrm{EIG}(\xi) \geq \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)}\left[\log \frac{q_\ell(y|m,\xi)}{\sum_{m'} p(m') q_\ell(y|m',\xi)}\right] \tag{8}$$

on the EIG in this case, which is not generally the case for the marginal + likelihood method. (In fact, this lower bound is itself a special case of the likelihood-free ACE lower bound introduced in Foster et al. (2020). Indeed, if we take the prior as the variational posterior and let $L \to \infty$ in the LF-ACE bound, we recover this lower bound.)

This lower bound also has a nice interpretation in the model selection scenario. The best design will be the one where the lower bound is largest, which happens, loosely speaking, when $q_\ell(y|m,\xi)$ is much larger than $\sum_m p(m) q_\ell(y|m,\xi)$. That means the approximate model evidence for the observation y under the correct model m is much larger than its evidence under other models. Thus, using the experiment with design ξ and observing y will allow us to easily discriminate between models.

To explicitly use this method, we need to choose trainable density estimators for $q_\ell(y|m,\xi;\phi)$ with parameters ϕ. The simplest method would be to have a distinct set of variational parameters for each value of m and ξ. Whilst it is possible to use a Gaussian density model, we could use more sophisticated methods such as normalising flows (Rezende and Mohamed, 2015). The training approach is similar to that for the posterior method. We use infinite simulated data, and maximise the variational lower bound using stochastic gradient optimisers.

The last two sections highlight a general feature of the variational methods of Foster et al. (2019)—we can either make variational approximations to densities over m or over y. Both lead to valid bounds.

1.4 Stochastic gradient optimisation of the design

So far, we have focused on variational estimation of the EIG. As shown in Foster et al. (2020), it is only a short jump from variational estimation of the EIG to stochastic gradient optimisation of the design using a variational lower bound on EIG. The benefit here, of course, is that we do not have conduct a grid search, co-ordinate exchange or similar algorithm over the design space. What we require instead is a continuous design space and the ability to differentiate observations with respect to designs.

Whilst Foster et al. (2020) focused on explicit likelihood models, both the posterior (Barber–Agakov) lower bound and the LF-ACE bound are applicable to the semi-implicit model selection setting. There is just one thing to check, which is that we can compute a derivative $\partial y/\partial \xi$. In the semi-implicit case, this is often fine.

For example, if $p(y|m, \psi, \xi)$ takes the form $y = g(m, \psi, \xi, \epsilon)$ for a differentiable g and an independent noise random variable ϵ.

Assuming this is the case, we can train ξ by stochastic gradient using either the posterior bound or the simplified LF-ACE bound that was derived in equation (8). We focus on the posterior lower bound for simplicity. Recall that, for the posterior bound, we are training a classifier to predict m from y. We have

$$\text{EIG}(\xi) \geq \mathbb{E}_{p(m)p(\psi|m)p(y|m,\psi,\xi)} \left[\log q_\phi(m|y)\right] + H[p(m)] \tag{9}$$

where q_ϕ is the classifier. One thing that we skimmed over slightly in the previous section was that ϕ implicitly depends on ξ via the training data, and different ξ will have different classifiers with different optimal values of the classifier parameters ϕ.

In Foster et al. (2020), rather than training separate classifiers with different designs ξ, we update ξ and ϕ togther in one stochastic gradient optimisation over the combined set of variables (ξ, ϕ). To explicitly write down the ξ gradient here, let's assume that we do have $y = g(m, \psi, \xi, \epsilon)$, so we can write

$$\mathcal{L}(\xi, \phi) = \mathbb{E}_{p(m)p(\psi|m)p(\epsilon)} \left[\log q_\phi(m|g(m, \psi, \xi, \epsilon))\right] + H[p(m)]. \tag{10}$$

In this form, the ξ gradient can be simply calculated as

$$\frac{\partial \mathcal{L}}{\partial \xi} = \mathbb{E}_{p(m)p(\psi|m)p(\epsilon)} \left[\left.\frac{\partial \log q_\phi}{\partial y}\right|_{m,g(m,\psi,\xi,\epsilon)} \left.\frac{\partial g}{\partial \xi}\right|_{m,\psi,\xi,\epsilon} \right]. \tag{11}$$

The beauty of modern auto-diff frameworks, of course, means that we do not even need to calculate this explicitly ourselves.

For model selection, equation (11) has a natural interpretation. We want to increase the lower bound $\mathcal{L}$ by moving to regions in which the classifier can confidently predict the correct model label m. This corresponds to moving y into regions in which $\log q_\phi(m|y)$ is larger *for the model that actually generated* y. In other words, we want the input to the classifier y to be pushed to regions where the classifier already finds it easy to classify correctly. That is, regions where deciding which model is correct is easier. We then exploit the differentiable relationship between ξ and y, and use this signal to 'improve' the input to the classifier by adjusting the design ξ to that such datasets y are more likely to be synthesised.

At the same time, we are constantly making gradient updates on the classifier parameters ϕ. This means that, as the distribution of (m, y) changes, the classifier can adjust accordingly.

If this sounds dubious, it is worth taking a step back. We are quite simply optimising the lower bound $\mathcal{L}(\xi, \phi)$ jointly with respect to ξ and ϕ, in the hopes that this global maximum may closely correspond to the EIG maximiser ξ^*. We actually have a guarantee that the value of $\mathcal{L}$ at our final trained variables $\hat{\xi}, \hat{\phi}$ is a lower bound on $\text{EIG}(\hat{\xi})$, i.e. the true value of $\hat{\xi}$ cannot be worse than the value we estimate for it.

Whilst the method is approximate, because we cannot quantify the discrepancy between $\mathcal{L}$ and the true EIG, it is highly scalable to very large design spaces. Other bounds presented in Foster et al. (2020) have the added benefit that they become equal to the EIG in a limit, providing some assurances that the global maximum of $\mathcal{L}$ is a good design. Foster et al. (2020) also introduced the evaluation method of establishing *lower and upper* bounds on chosen designs. This numerically bounds the discrepancy between the training objective $\mathcal{L}$ and the true EIG objective. Sadly, the upper bounds are only valid for explicit likelihood models; they don't work in the semi-implicit model selection case.

Finally, all of the above discussion carries over if we were to use the lower bound of equation (8) instead of the posterior bound.

1.5 Comparing with other classification approaches

We have established that the variational posterior approach of Foster et al. (2019) instructs us to learn a classifier to predict m from y and use the log probabilities $q_\phi(m|y)$ to estimate EIG. Other authors have considered supervised classification as a means to perform Bayesian experimental design for model selection.

Here, we focus on Hainy et al. (2018), which is "the first approach using supervised learning methods for optimal Bayesian design." This method trains a classifier that predicts m using y, with separate classifiers for different ξ. They focus on training decision trees and random forest classifiers (Breiman, 2001). Since random forests are not generally trained by stochastic gradient methods, this means that they fall back on simulating fixed training and test datasets of samples $(m_j, y_j)_{j=1}^J$ from $p(m)p(\psi|m)p(y|m, \psi, \xi)$. The training dataset is used to train the classifier model, whilst the test dataset gives unbiased estimates of the posterior loss. There is a danger that the classifier may overfit to the training set in this case. Compare this with the training of stochastic gradient classifiers in our previous sections—here we can draw fresh training batches on the fly, and avoid overfitting to a training set.

Decision trees and random forests do provide estimates of the class probabilities $q(m|y)$, but they are relatively noisy. For this reason, Hainy et al. (2018) focus on the 0–1 loss to evaluate designs. In the language of classification, therefore, they choose the design which gives the best *test accuracy*. Again, this is different to the variational approach which fits a neural classifier that automatically provides smooth probability estimates $q_\phi(m|y)$. The latter case was applied to estimate the information gain, which we showed is equivalent to choosing the design which gives the best *test loss*, assuming a negative log-likelihood loss function.

The trade-offs between these methods are clear when we consider optimising over a large design space. For the variational method, we have to train a number of neural networks to convergence. For the classification approach of Hainy et al. (2018), we train a number of random forest classifiers—this may be significantly more computationally efficient. Hainy et al. (2018) propose embedding their 0–1 loss estimation within a co-ordinate exchange algorithm (Meyer and Nachtsheim, 1995) to optimise over designs. The variational method, on the other hand, can naturally be embedded in a unified stochastic gradient optimisation to find the optimal design through stochastic gradient optimisation. The former may be more effective when the design space is not continuous, the latter can work well in a high-dimensional design space that is difficult to search using discrete methods.

2 Bayesian active learning by disagreement and Bayesian experimental design

2.1 Introduction

The purpose of this essay is to highlight the connection between the Bayesian Active Learning by Disagreement (BALD) score as estimated by Gal et al. (2017) and the Prior Contrastive Estimation (PCE) bound of Foster et al. (2020). There is a deep connection between Bayesian experimental design and Bayesian active learning. A significant touchpoint is the use of the mutual information score (Lindley, 1956)

$$I(\xi) = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}\left[H[p(\theta)] - H[p(\theta|y,\xi)]\right]. \tag{12}$$

to acquire new information in a Bayesian model with parameters θ where, y is the as yet unobserved outcome, and ξ is the design to be chosen.

2.2 Bayesian Active Learning by Disagreement

One of the computational challenges inherent in estimating equation (12) directly is that it involves repeated estimation of posterior distributions $p(\theta|y,\xi)$ for different simulated observations y. To remove this particular bottleneck, Houlsby et al. (2011) introduced a rewriting of the mutual information score using Bayes rule

$$I(\xi) = H[p(y|\xi)] - \mathbb{E}_{p(\theta)}\left[H[p(y|\theta,\xi)]\right]. \tag{13}$$

Whilst this is exactly equal to the original mutual information score, the new way of expressing I removes the requirement to estimate posterior distributions over θ. They termed equation (13) the Bayesian Active Learning by Disagreement (BALD) score.

Unfortunately, the story does not end with the BALD score because it still typically involves some intractable computations that must be estimated. For example, Houlsby et al. (2011) focused on approximations for Gaussian Process models (Williams and Rasmussen, 2006).

The more recent work by Gal et al. (2017) estimated the BALD score in the context of Bayesian deep learning classifiers. In such a model, θ represents the parameters of a classification model, and $p(y|\theta,\xi)$ is a probability distribution over classes $y \in \{c_1, \ldots, c_k\}$. Computing $p(y|\theta,\xi)$ involves a forward pass through the classifier with input ξ and parameters θ, the network generally ends in a softmax activation to produce a normalised distribution. To sample different values of θ, Gal et al. (2017) employed Monte Carlo Dropout (Gal and Ghahramani, 2016). Given independent samples $\theta_1, \ldots, \theta_M$ from $p(\theta)$, they proposed the following Deep BALD (DBALD) estimator of $I(\xi)$

$$I(\xi) \approx \hat{I}_{\text{DBALD}}(\xi) = H\left[\frac{1}{M}\sum_{i=1}^{M} p(y|\theta_i,\xi)\right] - \frac{1}{M}\sum_{i=1}^{M} H[p(y|\theta_i,\xi)] \tag{14}$$

where $H[P(y)] = -\sum_c P(y=c)\log P(y=c)$.

Notation For comparison with the original paper, we used θ in place of $\boldsymbol{\omega}$, ξ in place of $\mathbf{x}$, M in place of T and $p(\theta)$ is used in place of $q^*_\theta(\boldsymbol{\omega})$.

2.3 Prior Contrastive Estimation

In the context of stochastic gradient optimisation of Bayesian experimental designs, Foster et al. (2020) also considered the mutual information score $I(\xi)$ and the rearrangement equation (13). They proved the following Prior Contrastive Estimation (PCE) lower bound on $I(\xi)$

$$I(\xi) \geq \mathbb{E}_{p(\theta_0)p(y|\theta_0,\xi)p(\theta_1)\ldots p(\theta_L)}\left[\log \frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(y|\theta_\ell,\xi)}\right] \tag{15}$$

and used this bound to optimise ξ by stochastic gradient. One approach to estimate this bound using finite samples is the estimator

$$\hat{I}_{\text{PCE-naive}}(\xi) = \frac{1}{M}\sum_{m=1}^{M} \log \frac{p(y_m|\theta_{m0},\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(y_m|\theta_{m\ell},\xi)}. \tag{16}$$

where $y_m, \theta_{m0} \sim p(y,\theta|\xi)$ and $\theta_{m\ell} \sim p(\theta)$ for $\ell \geq 1$. However, we can also re-use samples more efficiently to give the estimator

$$\hat{I}_{\text{PCE}}(\xi) = \frac{1}{M}\sum_{m=1}^{M} \log \frac{p(y_m|\theta_m,\xi)}{\frac{1}{M}\sum_{\ell=1}^{M} p(y_m|\theta_\ell,\xi)}. \tag{17}$$

where $y_m, \theta_m \sim p(y,\theta|\xi)$. (To check the expectation of this version matches the PCE bound with $L = M-1$, we simply move the $\mathbb{E}$ sign inside of the summation.) Finally, Foster et al. (2020) discussed a speed-up that is possible when y is a discrete random variable taking values in $\{c_1, \ldots, c_k\}$. In this case, we can integrate out y by summing over it, rather than by drawing random samples of y. This method, called Rao-Blackwellisation, results in the estimator

$$\hat{I}_{\text{PCE-RB}}(\xi) = \frac{1}{M}\sum_{m=1}^{M}\sum_{c} p(y=c|\theta_m,\xi) \log \frac{p(y=c|\theta_m,\xi)}{\frac{1}{M}\sum_{\ell=1}^{M} p(y=c|\theta_\ell,\xi)}. \tag{18}$$

2.4 PCE and DBALD equivalence

We have looked at two parallel ways of approximating $I(\xi)$. The interesting result is that *the Rao-Blackwellised PCE estimator and the DBALD estimator are the same.* We can see this by direct calculation

$$\begin{aligned}
\hat{I}_{\text{PCE-RB}}(\xi) &= \frac{1}{M}\sum_{m=1}^{M}\sum_{c} p(y=c|\theta_m,\xi) \log \frac{p(y=c|\theta_m,\xi)}{\frac{1}{M}\sum_{\ell=1}^{M} p(y=c|\theta_\ell,\xi)} && (19)\\
&= \frac{1}{M}\sum_{m=1}^{M}\sum_{c} p(y=c|\theta_m,\xi) \log p(y=c|\theta_m,\xi) \\
&\quad - \frac{1}{M}\sum_{m=1}^{M}\sum_{c} p(y=c|\theta_m,\xi) \log\left(\frac{1}{M}\sum_{\ell=1}^{M} p(y=c|\theta_\ell,\xi)\right) && (20)\\
&= -\frac{1}{M}\sum_{m=1}^{M} H[p(y|\theta_m,\xi)] - \frac{1}{M}\sum_{m=1}^{M}\sum_{c} p(y=c|\theta_m,\xi) \log\left(\frac{1}{M}\sum_{\ell=1}^{M} p(y=c|\theta_\ell,\xi)\right) && (21)\\
&= -\frac{1}{M}\sum_{m=1}^{M} H[p(y|\theta_m,\xi)] - \sum_{c}\left(\frac{1}{M}\sum_{m=1}^{M} p(y=c|\theta_m,\xi)\right) \log\left(\frac{1}{M}\sum_{\ell=1}^{M} p(y=c|\theta_\ell,\xi)\right) && (22)\\
&= -\frac{1}{M}\sum_{m=1}^{M} H[p(y|\theta_m,\xi)] + H\left[\frac{1}{M}\sum_{m=1}^{M} p(y|\theta_m,\xi)\right] && (23)\\
&= \hat{I}_{\text{DBALD}}(\xi). && (24)
\end{aligned}$$

A major consequence of this result is that *the expectation of the DBALD score is a lower bound on the true mutual information score.* We also note that this estimator has been used by Vincent and Rainforth (2017) in the context of Bayesian experimental design, although they did not show that it was a stochastic lower bound.

2.5 New diagnostic for the DBALD score

One advantage of making this connection is that we can bring certain diagnostics that were applied by Foster et al. (2020) over to the active learning setting. In particular, Foster et al. (2020) paired their PCE lower

bound with a complementary *upper bound* on $I(\xi)$. This provides a very useful diagnostic tool to tune the number of samples M used to compute the DBALD score. If the lower bound and upper bound are very close, we know that the difference between the DBALD score and the true mutual information must also be small. On the other hand, if the upper and lower bounds are far apart, then the DBALD score might not yet be close to the true mutual information.

One upper bound upper by Foster et al. (2020) was the Nested Monte Carlo (NMC) (Vincent and Rainforth, 2017) estimator. For the discrete y case with Rao-Blackwellisation, the estimator is

$$\hat{I}_{\text{NMC-RB}}(\xi) = -\frac{1}{M}\sum_{m=1}^{M}\sum_{c} p(y=c|\theta_m,\xi)\log\left(\frac{1}{M-1}\sum_{\ell\neq m} p(y=c|\theta_\ell,\xi)\right) - \frac{1}{M}\sum_{m=1}^{M} H[p(y|\theta_m,\xi)] \tag{25}$$

$$= \frac{1}{M}\sum_{m=1}^{M} H\left[p(y|\theta_m,\xi), \frac{1}{M-1}\sum_{\ell\neq m} p(y|\theta_\ell,\xi)\right] - \frac{1}{M}\sum_{m=1}^{M} H[p(y|\theta_m,\xi)] \tag{26}$$

where $H[p,q]$ is the cross-entropy. The expectation of this mutual information estimator is always an upper bound on $I(\xi)$. So both the DBALD score and the NMC-RB estimator converge to $I(\xi)$ as $M \to \infty$, but from opposite directions. We suggest NMC-RB as a diagnostic for the parameter M.

2.6 BALD estimators for regression

The connection to PCE may also be helpful when considering regression models. The standard parametrisation of a Bayesian neural network for regression is for the output of the network with parameters θ and input ξ to be the predictive mean μ and standard deviations σ of a Gaussian $y|\theta,\xi \sim N(\mu(\theta,\xi),\sigma(\theta,\xi)^2)$. (It is normal for y, μ and σ to be vector-valued and for the Gaussian to have a diagonal covariance matrix.)

For the DBALD estimator for a regression model, the entropy of a Gaussian is known in closed form, so $H[p(y|\theta_i,\xi)] = \frac{1}{2}\log\left(2\pi e\sigma(\theta_i,\xi)^2\right)$. However, the entropy of a mixture of Gaussians $H\left[\frac{1}{M}\sum_{i=1}^{M} p(y|\theta_i,\xi)\right]$ cannot be computed analytically. Instead, we could estimate this mixture of Gaussians entropy using Monte Carlo by sampling $i \in \{1,\ldots,M\}$ uniformly, sampling y from $p(y|\theta_i,\xi)$ and calculating the log-density at y.

Despite the fact that we are using an analytic entropy for one term, and a Monte Carlo estimate for the other, it's easy to see that this new estimator is a *partially Rao-Blackwellised* PCE estimator. (This can be proved starting from equation (17).) That means all the existing facts, such as the estimator being a stochastic lower bound on $I(\xi)$, carry over naturally to the regression case.

3 Deep Adaptive Design and Bayesian reinforcement learning

3.1 Introduction

The purpose of this essay is to discuss the connections between the recently proposed Deep Adaptive Design (DAD) (Foster et al., 2021) method and the field of Bayesian reinforcement learning (Ghavamzadeh et al., 2016). That such a connection exists is hinted at by a high-level appraisal of the DAD method—it solves a sequential decision making problem to optimise a certain objective function, decision optimality is dependent on a *state* which is the experimental data already gathered, and the automated decision maker is a design *policy* network. We begin by showing how the sequential Bayesian experimental design problem solved by DAD can be viewed as a Bayes Adaptive Markov Decision Process (BAMDP) (Ross et al., 2007; Guez et al., 2012), making this connection formally precise. We also isolate some of the key differences between the problem DAD is solving and a conventional Bayesian RL problem, noting that the reward in DAD is intractable. Much of the effort of DAD is in establishing a differentiable surrogate for the true objective. The differentiability of the surrogate reward is also a key feature of the DAD problem, which facilitates the direct policy optimisation approach taken to train the policy that is rarely applicable in standard RL problems. We also highlight other features of the DAD method, such as its avoidance of explicitly estimating any posterior distributions, i.e. the avoidance of explicit belief state estimation.

Having studied DAD in some detail, we consider possible extensions of the method that make use of the RL connection. First, there are rather natural extensions of DAD to more general objective functions that incorporate design costs, terminal decisions and other functionals of the posterior distribution. Second, more standard approaches to (Bayesian) RL, such as Q-learning (Watkins and Dayan, 1992; Dearden et al., 1998) can be applicable to the sequential Bayesian experimental design problem. They may be particularly useful for long- or infinite-horizon problems.

3.2 Background on Bayesian Reinforcement Learning

3.2.1 Markov Decision Processes

The Markov Decision Process (MDP) (Bellman, 1957; Duff, 2002) is a highly successful mathematical framework for sequential decision problems in a known environment. Formally, a MDP consists of a state space S, an action space A, a transition model $\mathcal{P}$, a reward distribution R, a discount factor $0 \leq \gamma \leq 1$ and a time horizon T which may be infinite. An agent operates in the MDP by moving between different states in discrete time. For example, if the agent is in state s_t at time t and chooses to play action a_t, then the next state s_{t+1} will be sampled randomly according to the transition model $s_{t+1} \sim \mathcal{P}(s|s_t, a_t)$. Since the distribution over the next state depends only on s_t and a_t, the transitions are Markovian. Finally, by making the transition $s_t \xrightarrow{a_t} s_{t+1}$, the agent receives a random reward $r_t \sim R(r|s_t, a_t, s_{t+1}) \in \mathbb{R}$. The agent's objective is to maximise the discounted sum of rewards $\sum_{t=0}^{T} \gamma^t r_t$. Given the Markovian nature of the problem, it is sufficient to choose actions according to some *policy* π, where $a_t = \pi(s_t)$. The optimality condition for a policy is

$$\pi^* = \arg\max_{\pi} \mathcal{J}(\pi), \tag{27}$$

where

$$\mathcal{J}(\pi) = \mathbb{E}_{s_0 \sim p(s_0) \prod_{t=0}^{T} a_t = \pi(s_t), s_{t+1} \sim \mathcal{P}(s|s_t, a_t), r_t \sim R(r|s_t, a_t, s_{t+1})} \left[\sum_{t=0}^{T} \gamma^t r_t \right]. \tag{28}$$

In a classical MDP, we assume that $\mathcal{P}$ and R are known during the planning phase, when the agent devises their policy π. Of particular utility in planning a policy is the value function, defined as

$$V^\pi(s) = \mathbb{E}_{s' \sim \mathcal{P}(\cdot|s,\pi(s)), r \sim R(r|s,\pi(s),s')} \left[r + \gamma V^\pi(s') \right] \tag{29}$$

and the Q-function

$$Q^\pi(s, a) = \mathbb{E}_{s' \sim \mathcal{P}(\cdot|s,a), r \sim R(r|s,a,s')} \left[r + \gamma V^\pi(s') \right]. \tag{30}$$

These equations are valid when $T = \infty$, for finite time horizon we also have to take account of time t in state evaluations.

3.2.2 Bayes Adaptive Markov Decision Processes

The BAMDP (Duff, 2002; Ross et al., 2007; Guez et al., 2012; Ghavamzadeh et al., 2016) is one approach to generalising the MDP to deal with unknown transition models. In the BAMDP, the agent retains an explicit posterior distribution over the transition model called a belief state. This allows a formally elegant approach to behaviour under uncertainty which can trade off exploration (learning the transition model) and exploitation (executing actions that receive a high reward).

To set this up formally using the notation of Guez et al. (2012), we begin by considering an outer probabilistic model over the transition probabilities with prior $P(\mathcal{P})$. Given a history of states, actions and rewards $h_t = s_0 a_0 \dots r_{t-1} a_{t-1} s_t$, we can compute a posterior distribution on $\mathcal{P}$ by

$$P(\mathcal{P}|h_t) \propto P(\mathcal{P})P(h_t|\mathcal{P}) = P(\mathcal{P}) \prod_{\tau=0}^{t} \mathcal{P}(s_{\tau+1}|s_\tau, a_\tau). \tag{31}$$

To bring this back into the MDP formulation, we consider an augmented state space S^+ which consists of entire histories, and which encapsulates both the current state and our beliefs about the transition model. Transitions in the augmented state space S^+ are given by integrating over the current beliefs on $\mathcal{P}$

$$\mathcal{P}^+(h_{t+1}|h_t, a_t) = \int P(\mathcal{P}|h_t)\mathcal{P}(s_{t+1}|s_t, a_t)\, d\mathcal{P}. \tag{32}$$

It is also possible for BAMDPs to incorporate unknown reward distributions (see e.g. Zintgraf et al. (2019)), where an outer model over reward distributions is updated on the basis of h_t in the same manner as for the transition probabilities. Specifically, if we have a prior $P(R)$ over reward distributions, then the reward function for playing action a_t in augmented state h_t is

$$R^+(r|h_t, a_t, h_{t+1}) = \int P(R|h_{t+1})R(r|s_t, a_t, s_{t+1})\, dR. \tag{33}$$

Combining these gives a new MDP with state space S^+ of histories, unchanged action space A, augmented transition model $\mathcal{P}^+$, augmented reward distribution R^+, discount factor γ and time horizon T. Optimal action in this new MDP gives the optimal trade-off between exploration and exploitation.

3.3 The Bayesian RL formulation of DAD

In DAD (Foster et al., 2021), we choose a sequence of designs $\xi_1, \dots, \xi_T$ with a view to maximising the expected information gained about a latent parameter of interest θ. To place DAD in a Bayesian RL setting, we begin by associating the design ξ_t chosen before observing an outcome with the action a_{t-1}. The difference in time labels is necessary because ξ_t is chosen before y_t is observed. Since the observation distribution $p(y|\xi, \theta)$ depends on the unknown θ, we are not in a MDP, but rather a BAMDP. As in the previous section, it seems sensible to consider the state space for DAD as the space of histories $h_t = \xi_1 y_1 \dots \xi_t y_t$. Uncertainty over the transition model in DAD is captured by uncertainty in θ. Specifically, we have the following transition distribution for history states

$$p(h_{t+1}|h_t, \xi_{t+1}) = \int p(\theta|h_t)p(y_{t+1}|\xi_{t+1}, \theta)\, d\theta \tag{34}$$

which is the analogue of equation (32), but now expressed in the notation of experimental design. Unlike the standard reinforcement learning setting, there are no external rewards in DAD. Instead, rewards are defined in terms of information gathered about θ. Specifically, we can take the reward distribution on augmented

states $R^+(r|h_t, a_t, h_{t+1})$ to be a deterministic function of h_{t+1} that represents the information gained about θ by moving from h_t to h_{t+1}. This is given by the reduction in entropy

$$R^+(h_t, a_t, h_{t+1}) = H[p(\theta|h_t)] - H[p(\theta|h_{t+1})]. \tag{35}$$

To complete the BAMDP specification, we take $\gamma = 1$ and we use a time horizon of T. This gives the objective function for policies

$$\mathcal{J}(\pi) = \mathbb{E}\left[\sum_{t=1}^{T} r_t\right] = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\sum_{t=1}^{T} H[p(\theta|h_{t-1})] - H[p(\theta|h_t)]\right]. \tag{36}$$

To connect this with the objective that is used in DAD, we apply Theorem 1 of Foster et al. (2021), which tells us that

$$\mathcal{J}(\pi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\sum_{t=1}^{T} H[p(\theta|h_{t-1})] - H[p(\theta|h_t)]\right] \overset{\text{Theorem 1}}{=} \mathcal{I}_T(\pi) \tag{37}$$

where

$$\mathcal{I}_T(\pi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\log \frac{p(h_T|\theta,\pi)}{\mathbb{E}_{p(\theta')}[p(h_T|\theta',\pi)]}\right]. \tag{38}$$

In summary, we can cast the problem that DAD solves as a BAMDP. We identify designs with actions, experimental histories with augmented states, we use the probabilistic model to give a natural transition distribution on these states, we introduce non-random rewards that are one-step information gains, we set $\gamma = 1$ and generally assume a finite number of experiment iterations T.

3.4 What makes the experimental design problem distinctive?

Having established a theoretical connection between sequential Bayesian experimental design and Bayesian RL, one might naturally ask whether there is any reason to develop specialist algorithms for experimental design when general purpose Bayesian RL algorithms are applicable. First, we focus on the reward structure of the Bayesian experimental design problem. The rewards $r_t = H[p(\theta|h_{t-1})] - H[p(\theta|h_t)]$ are generally intractable, requiring Bayesian inference on θ. Rather than attempting to estimate this reward, DAD proposes the sPCE lower bound on the total expected informationn gain under policy π, namely

$$\mathcal{I}_T(\pi) \geq \mathcal{L}_T(\pi, L) = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi)}\right]. \tag{39}$$

Interestingly, there is a way to interpret the sPCE objective within the RL framework. First, we use *root sampling* to sample θ_0 and h_T together. We also fix the contrasts $\theta_{1:L}$. Finally, we use the surrogate rewards

$$\tilde{r}_t = \log \frac{p(h_t|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_t|\theta_\ell,\pi)} - \log \frac{p(h_{t-1}|\theta_0,\pi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_{t-1}|\theta_\ell,\pi)}. \tag{40}$$

Since these rewards depend on θ_0, we can treat them as randomised rewards if we are only conditioning on h_t.

One important feature of these rewards is that, whilst intractable, the surrogate $\mathcal{L}_T(\pi, L)$ is differentiable with respect to the designs $(\xi_t)_{t=1}^T$ and observations $(y_t)_{t=1}^T$. In the simplest form of DAD, we further assume a differentiable relationship between y_t and ξ_t that is encapsulated by a reparametrisable way to sample $p(y|\theta,\xi)$. Concretetly, for example, we might have $y|\theta,\xi = \mu(\theta,\xi) + \sigma(\theta,\xi)\varepsilon$ where $\varepsilon \sim N(0,1)$ and μ and σ are differentiable functions. The result of these assumptions is that we can directly differentiate the surrogate objective $\mathcal{L}_T(\pi, L)$ with respect to the parameters ϕ of the policy network π_ϕ that generates the designs $(\xi_t)_{t=1}^T$ according to the formula $\xi_t = \pi_\phi(h_{t-1})$. DAD optimises the policy π_ϕ directly by gradient descent on $\mathcal{L}_T(\pi, L)$.

Thus, DAD can be characterised in RL language as a direct policy optimisation method. Whilst direct policy optimisation methods (Lorberbom et al., 2019; Howell et al., 2021) are used in RL, they are far

from the norm, with methodologies such as Q-learning (Watkins and Dayan, 1992) and actor–critic (Konda and Tsitsiklis, 2000) being more dominant. This may be because RL does not typically assume that the reward function is differentiable—for example, rewards from a real environment rarely come with gradient information. It may also be because discrete action problems are more the focus.

DAD also contrasts with many approaches to *Bayesian* RL in that it avoids the estimation of the posteriors $p(\theta|h_t, \pi)$. In Bayesian RL, these posterior distributions are referred to as *belief states*. Many methods for tackling Bayesian RL problems utilise the estimation of belief states (Ghavamzadeh et al., 2016; Igl et al., 2018; Zintgraf et al., 2019). DAD instead relies on an approach that is closer to the method of root sampling (Guez et al., 2012). This is also one difference between DAD and the previous approach to non-greedy sequential Bayesian experimental design of Huan and Marzouk (2016).

3.5 New objective functions for DAD

Seeing DAD in the framework of Bayesian RL naturally invites the question of whether the general DAD methodology can be applied to objective functions (rewards) that are not information gains. The preceding discussion suggests that, using root sampling so a dependence on θ is possible, we could consider rewards of the form

$$r_t^{\text{general}} = R(\theta, h_t, \epsilon_t) \tag{41}$$

where R is a known differentiable function and ϵ_t is an independent noise random variable. Clearly, the information gain reward r_t fits this pattern, being a function of h_t only. Combining the differentiable reward function with the reparametrisation assumption would mean that the general reward

$$\mathcal{J}^{\text{general}}(\pi) = \mathbb{E}_{p(\theta)p(h_T)p(\epsilon_{1:T})}\left[\sum_{t=1}^{T} r_t^{\text{general}}\right] \tag{42}$$

can be optimised with respect to π by direct policy gradients. In the experimental design context, this opens the door to two relatively simple extensions of DAD. First, we can assign a (differentiable) cost to each design. Suppose we augment the original expected information gain objective with the negative sum of the costs of the designs. Using λ to trade off cost and information, we arrive at

$$\mathcal{J}^{\text{costed}}(\pi) = \mathcal{I}_T(\pi) - \lambda\mathbb{E}\left[\sum_{t=1}^{T} C(\xi_t)\right] \tag{43}$$

which we can tackle using an approach that is essentially the same as DAD. Second, we can consider different measures of the quality of the final posterior distribution. For instance, with a one-dimensional θ, we might be more interested in reducing posterior *variance* than posterior entropy. We could take the reward function

$$r_t^{\text{variance}} = \text{Var}_{p(\theta|h_{t-1})}[\theta] - \text{Var}_{p(\theta|h_t)}[\theta]. \tag{44}$$

Whilst there are certain reasons why the entropy approach is considered more theoretically well-justified (Lindley, 1956), using a different functional of the posterior distribution as a reward signal does fit relatively naturally into the DAD framework. The remaining piece of the puzzle would be whether that functional could be estimated efficiently as DAD estimates the information gain using sPCE. For the variance, we have

$$\mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[\sum_{t=1}^{T} r_t^{\text{variance}}\right] \geq \text{Var}_{p(\theta)}[\theta] - \mathbb{E}_{p(\theta)p(h_T|\theta,\pi)}\left[(\theta - f_{\phi'}(h_T))^2\right] \tag{45}$$

where $f_{\phi'}$ is a learnable function. Note the similarity with the Barber–Agakov bound (Barber and Agakov, 2003; Foster et al., 2019, 2020).

3.6 RL algorithms for Bayesian experimental design

To conclude, making the formal connection between sequential Bayesian experimental design opens up the possibility of using the vast literature on Bayesian RL and control theory to improve our ability to plan

sequential experiments. Whilst the direct policy optimisation approach of DAD works remarkably well, understanding the connection to RL should aid us when this training method begins to break down. The application of existing Bayesian RL algorithms to experimental design is an exciting area for new research that is well within reach.

A case of potential difficulty for DAD, where such insights may be useful, is in long-horizon experiments. In order to plan effectively for long experiments, DAD simulates thousands of possible experimental trajectories. However, the efficiency of this simulation is likely to drop as T increases. DAD is extremely data hungry—it resimulates completely new trajectories at each gradient step. This avoids any problems of the training data becoming out-of-date, but it increases the training cost.

It is also conceivable that, in some settings, it is impossible to plan for all future eventualities. The RL analogy would be a strongly stochastic environment in which a game is selected at random from a long list at the start of play. The agent, therefore, has to first discover which game it is playing, and then to play it successfully. If all planning is conducted up-front, then the RL agent has to learn how to play every single game well before starting on the real environment. The alternative is to introduce some real data and retrain the policy as we go. In the RL setting, that would mean discovering which game is being played before knowing how to play the games, which could be achieved with a much simpler policy. Once this discovery is made with good confidence, we can retrain to learn to play that specific game. In the experimental design setting, we are often in the 'unknown game' setting. This is because, until we have observed some data, it is almost impossible to know which later experiments will be optimal to run. The DAD approach is to simulate different possibilities and learn to 'play' well across the board. The retraining alternative would be a hybrid approach between the standard greedy method and DAD in which some real data is used to retrain the policy as we progress.

4 Statistical estimation of mutual information

4.1 Introduction

Mutual information is a central statistical quantity that measures the relationship between two random variables. In machine learning, it has found use in blind source separation (Hyvärinen, 1999), representation learning (van den Oord et al., 2018), the information bottleneck (Tishby et al., 2000) and feature selection (Kwak and Choi, 2002). It is also a key quantity in Bayesian experimental design (Lindley, 1956). The mutual information between jointly distributed random variables $\mathbf{x}, \mathbf{y} \sim p(\mathbf{x}, \mathbf{y})$ is defined as

$$I(\mathbf{x}, \mathbf{y}) = \mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})}\right]. \tag{46}$$

In this document, we focus on the estimation of mutual information in the *explicit likelihood* setting in which one of the conditional densities, say $p(\mathbf{y}|\mathbf{x})$ is known in closed form. In this case, asymptotically consistent estimators exist for the mutual information, and we are concerned in studying their convergence rates. In the *implicit likelihood* setting, the standard approach is to introduce a positive, unnormalised function $\kappa(\mathbf{x}, \mathbf{y})$ that is an estimate of the joint $p(\mathbf{x}, \mathbf{y})$. However, estimators that use κ as a surrogate for the true unknown density can only be guaranteed to produce lower bounds on the mutual information in the limit of infinite samples of $\mathbf{x}, \mathbf{y}$. The convergence rates, though, behave similarly.

4.2 Nested Monte Carlo and leave-one-out estimators

The Nested Monte Carlo (NMC) estimator (Ryan, 2003), also called the double loop estimator, for mutual information estimation with an explicit likelihood is defined as

$$A_{n,m} = \frac{1}{n}\sum_{i=1}^{n} \log \frac{p(\mathbf{y}_i|\mathbf{x}_i)}{\frac{1}{m}\sum_{j=1}^{m} p(\mathbf{y}_i|\mathbf{x}_{ij})} \tag{47}$$

where $\mathbf{x}_i, \mathbf{y}_i \overset{\text{i.i.d.}}{\sim} p(\mathbf{x}, \mathbf{y})$ and $\mathbf{x}_{ij} \overset{\text{i.i.d.}}{\sim} p(\mathbf{x})$ are independent. It is also possible to include some correlation in the $\mathbf{x}$ samples, for example we can repeatedly use $(\mathbf{x}_{1j})_{j=1}^{m}$

$$A'_{n,m} = \frac{1}{n}\sum_{i=1}^{n} \log \frac{p(\mathbf{y}_i|\mathbf{x}_i)}{\frac{1}{m}\sum_{j=1}^{m} p(\mathbf{y}_i|\mathbf{x}_{1j})}, \tag{48}$$

and we can use the original n samples, giving the leave-one-out (LOO) estimator (Poole et al., 2019)

$$\tilde{A}_n = \frac{1}{n}\sum_{i=1}^{n} \log \frac{p(\mathbf{y}_i|\mathbf{x}_i)}{\frac{1}{n-1}\sum_{j\neq i} p(\mathbf{y}_i|\mathbf{x}_j)}. \tag{49}$$

Note that $\mathbb{E}[A_{n,m}] = \mathbb{E}[A'_{n,m}]$ and $\mathbb{E}[\tilde{A}_n] = \mathbb{E}[A_{n,n-1}]$, so the correlations only change the variance. Furthermore, estimators $A_{n,m}$ and $A'_{n,m}$ both cost $\mathcal{O}(mn)$ evaluations of the likelihood and $\tilde{A}_n$ costs $\mathcal{O}(n^2)$ evaluations of the likelihood. So, whilst $A'_{n,m}$ and $\tilde{A}_n$ appear more efficient in their use of samples, their theoretical computational complexity is not different to $A_{n,m}$.

Here, we focus on analysing the estimator $A_{n,m}$. Our results reaffirm previous analysis by Rainforth et al. (2018); Zheng et al. (2018); Beck et al. (2018). We focus on a rigorous approach to using Taylor's Theorem for the logarithm. Our techniques can then be used to analyse other estimators.

Theorem 1 (Expectation of $A_{n,m}$)**.** *Suppose there exist Hölder conjugate indices $p, q > 0$ with $1/p + 1/q = 1$ such that*

$$\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right)^{3p}\right] < \infty \text{ and } \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{q}\right] < \infty. \tag{50}$$

Then we have

$$\mathbb{E}[A_{n,m}] = I(\mathbf{x}, \mathbf{y}) + \frac{1}{m}\mathbb{E}_{p(\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x})}[p(\mathbf{y}|\mathbf{x})]}{2p(\mathbf{y})^2}\right] + \mathcal{O}\left(m^{-3/2}\right). \tag{51}$$

Proof. By linearity, $\mathbb{E}[A_{n,m}] = \mathbb{E}[A_{1,m}]$. To compute this expectation, we define

$$U_j = \frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)}. \tag{52}$$

with $E[U_j] = \mathbb{E}[\mathbb{E}[U_j|\mathbf{y}_1]] = 1$. Then,

$$A_{1,m} = \log \frac{p(\mathbf{y}_1|\mathbf{x}_1)}{p(\mathbf{y}_1)} - \log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right), \tag{53}$$

giving

$$\mathbb{E}[A_{n,m}] = I(\mathbf{x},\mathbf{y}) - \mathbb{E}\left[\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)\right]. \tag{54}$$

The standard approach to analysing the second term is to apply Taylor's Theorem to the logarithm function. However, a naive application does not work for several reasons: a) the Taylor series for the logarithm about 1 is convergent only on $(0, 2)$ rather than $(0, \infty)$, b) the derivatives of the logarithm are not bounded at 0, so the classical Delta Method (Lemma 9) does not apply. To get around these problems, we define the partial Taylor series

$$L_k(x) = \sum_{j=1}^{k} \frac{(-1)^{j+1}}{j}(x-1)^j, \tag{55}$$

in Lemma 10, we prove that $|\log x - L_k(x)| \leq |x-1|^{k+1}\max(1, -\log x)$ on $(0, \infty)$. Taking $k = 2$, we have

$$\mathbb{E}\left[\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)\right] = -\frac{1}{2}\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right)^2\right] + \mathbb{E}[\varepsilon] \tag{56}$$

and

$$|\mathbb{E}[\varepsilon]| \leq \mathbb{E}[|\varepsilon|] \leq \mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right|^3 \max\left(1, -\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)\right)\right] \tag{57}$$

applying Hölder's Inequality

$$\leq \mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right|^{3p}\right]^{1/p} \mathbb{E}\left[\max\left(1, -\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)\right)^q\right]^{1/q}. \tag{58}$$

We tackle each term separately. Since the U_j are i.i.d conditional on $\mathbf{y}_1$, we can apply Corollary 8 that uses the Marcinkiewicz–Zygmund Inequality, and the Tower Law to conclude that there is a finite constant D_{3p} such that

$$\mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right|^{3p}\right]^{1/p} \leq D_{3p}^{1/p} m^{-3/2}\mathbb{E}\left[|U_1 - 1|^{3p}\right]^{1/p} \tag{59}$$

and

$$\mathbb{E}\left[|U_1 - 1|^{3p}\right] \leq 1 + \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right)^{3p}\right] < \infty \text{ by assumption.} \tag{60}$$

So this term is $\mathcal{O}(m^{-3/2})$. For the latter term, we use the fact that $x \mapsto \max(1, -\log x)$ is a convex function. Thus

$$\begin{aligned}
\mathbb{E}\left[\max\left(1, -\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)\right)^q\right]^{1/q} &\leq \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\max\left(1, -\log\left(U_j\right)\right)^q\right]^{1/q} && (61)\\
&= \mathbb{E}\left[\max\left(1, -\log\left(U_1\right)\right)^q\right]^{1/q} && (62)\\
&\leq \left(1 + \mathbb{E}[|\log U_1|^q]\right)^{1/q} && (63)\\
&= \left(1 + \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^q\right]\right)^{1/q} && (64)\\
&< \infty \text{ by assumption.} && (65)
\end{aligned}$$

Overall, we have $\mathbb{E}[\varepsilon] = \mathcal{O}(m^{-3/2})$. Finally,

$$\frac{1}{2}\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right)^2\right] = \frac{1}{2m}\mathbb{E}_{p(\mathbf{y})}\left[\mathbb{E}_{p(\mathbf{x})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} - 1\right)^2\right]\right] = \frac{1}{m}\mathbb{E}_{p(\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x})}[p(\mathbf{y}|\mathbf{x})]}{2p(\mathbf{y})^2}\right]. \tag{66}$$

This completes the proof. □

A simple application of Jensen's Inequality further shows that $\mathbb{E}[A_{n,m}] \geq I(\mathbf{x}, \mathbf{y})$ for every value of n and m. Put another way, the NMC estimator is always a stochastic upper bound on the mutual information with bias of order $1/m$. Zheng et al. (2018) showed that the coefficient of the $1/m$ term is

$$\mathbb{E}_{p(\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x})}[p(\mathbf{y}|\mathbf{x})]}{2p(\mathbf{y})^2}\right] = \frac{1}{2}\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} - 1\right)^2\right] \tag{67}$$

which is the χ^2-divergence from $p(\mathbf{x}, \mathbf{y})$ to $p(\mathbf{x})p(\mathbf{y})$.

Theorem 2 (Variance of $A_{n,m}$)**.** *Assume that there exist Hölder conjugate indices $p, q > 0$ such that*

$$\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right)^{3p}\right] < \infty \text{ and } \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{2q}\right] < \infty. \tag{68}$$

Then,

$$\begin{aligned}
\mathrm{Var}[A_{n,m}] = &\frac{1}{n}\mathrm{Var}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right]\\
&+ \frac{1}{nm}\left(\mathbb{E}_{p(\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x})}[p(\mathbf{y}|\mathbf{x})]}{p(\mathbf{y})^2}\right] + \mathrm{Cov}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}, \frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right]\right)\\
&+ \mathcal{O}\left(n^{-1}m^{-3/2}\right).
\end{aligned} \tag{69}$$

Proof. We have

$$\mathrm{Var}[A_{n,m}] = \frac{1}{n}\mathrm{Var}[A_{1,m}]. \tag{70}$$

For the variance of $A_{1,m}$, we use the Tower Law for the Variance

$$\mathrm{Var}[A_{1,m}] = \mathbb{E}[\mathrm{Var}[A_{1,m}|\mathbf{x}_1, \mathbf{y}_1]] + \mathrm{Var}[\mathbb{E}[A_{1,m}|\mathbf{x}_1, \mathbf{y}_1]]. \tag{71}$$

For the conditional variance, we follow the proof of Theorem 1 to see that

$$\mathbb{E}\left[\mathrm{Var}[A_{1,m}|\mathbf{x}_1, \mathbf{y}_1]\right] = \mathbb{E}\left[\mathrm{Var}\left[\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)\middle|\, \mathbf{y}_1\right]\right] \text{ where } U_j = \frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)} \tag{72}$$

We will the form of the variance $\mathrm{Var}[A] = \mathbb{E}[A^2] - \mathbb{E}[A]^2$. We now study the function $x \mapsto \log(x)^2$. Taylor's Theorem suggests that $\log x = (x-1)^2 + ...$, but as before, we aim for a more rigorous approach. We have

$$|\log(x)^2 - (x-1)^2| = |(\log x - x + 1)(\log x + x - 1)| \leq |\log x - x + 1||\log x + x - 1|. \tag{73}$$

Using Lemma 10, we can show $|\log x - x + 1| \leq |x-1|^2 \max(1, -\log x)$. It is also elementary to check that $|\log x + x - 1| \leq 3|x-1| \max(1, -\log x)$. Hence

$$|\log(x)^2 - (x-1)^2| \leq 3|x-1|^3 \max(1, -\log x)^2. \tag{74}$$

We can now return to computing the conditional expectation of equation (72). We have

$$\mathbb{E}\left[\mathbb{E}\left[\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)^2 \middle| \mathbf{y}_1\right]\right] = \mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right)^2\right] + \mathbb{E}[\eta] \tag{75}$$

where our recent result guarantees that

$$|\mathbb{E}[\eta]| \leq \mathbb{E}[|\eta|] \leq 3\mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right|^3 \max\left(1, -\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right)\right)^2\right]. \tag{76}$$

Without reproducing all the details, the approach of Theorem 1 shows us that this error term is $\mathcal{O}(m^{-3/2})$ provided that

$$\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right)^{3p}\right] < \infty \text{ and } \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{2q}\right] < \infty \tag{77}$$

where p, q are Hölder conjugate indices. Theorem 1 also shows that

$$\mathbb{E}\left[\mathbb{E}\left[\log\left(\frac{1}{m}\sum_{j=1}^{m} U_j\right) \middle| \mathbf{y}_1\right]^2\right] = \mathcal{O}(m^{-2}). \tag{78}$$

Putting these pieces together, we have

$$\mathbb{E}\left[\mathrm{Var}[A_{1,m}|\mathbf{x}_1, \mathbf{y}_1]\right] = \frac{1}{m}\mathbb{E}_{p(\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x})}[p(\mathbf{y}|\mathbf{x})]}{p(\mathbf{y})^2}\right] + \mathcal{O}(m^{-3/2}). \tag{79}$$

Turning to the variance of the conditional expectation, recall from Theorem 1 that

$$\mathbb{E}\left[A_{1,m}|\mathbf{x}_1, \mathbf{y}_1\right] = \log\frac{p(\mathbf{y}_1|\mathbf{x}_1)}{p(\mathbf{y}_1)} + \frac{1}{m}\frac{\mathrm{Var}_{p(\mathbf{x})}[p(\mathbf{y}_1|\mathbf{x})]}{2p(\mathbf{y}_1)^2} + \mathcal{O}\left(m^{-3/2}\right). \tag{80}$$

Taking the variance gives

$$\mathrm{Var}\left[\mathbb{E}\left[A_{1,m}|\mathbf{x}_1, \mathbf{y}_1\right]\right] = \mathrm{Var}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right] + \frac{1}{m}\mathrm{Cov}\left(\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}, \frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right) + \mathcal{O}(m^{-3/2}). \tag{81}$$

Thus,

$$\begin{aligned}\mathrm{Var}[A_{1,m}] = &\mathrm{Var}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right] \\ &+ \frac{1}{m}\left(\mathbb{E}_{p(\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x})}[p(\mathbf{y}|\mathbf{x})]}{p(\mathbf{y})^2}\right] + \mathrm{Cov}_{p(\mathbf{x},\mathbf{y})}\left(\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}, \frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right)\right) + \mathcal{O}(m^{-3/2})\end{aligned} \tag{82}$$

and the full result follows. □

Combining the last two theorems establishes that

$$\mathbb{E}\left[|A_{n,m} - I(\mathbf{x},\mathbf{y})|^2\right] = \mathcal{O}\left(\frac{1}{n} + \frac{1}{m^2}\right). \tag{83}$$

The computational cost of $A_{n,m}$ is $\mathcal{O}(mn)$. Thus it is optimal to set $m \propto \sqrt{n}$. Then the estimator converges to $I(\mathbf{x},\mathbf{y})$ at a rate $T^{-1/3}$ in root mean square, where T is the total computational budget.

Finally, in the case that $p(\mathbf{y}|\mathbf{x})$ is not known, we can repeat this analysis using a positive function $\kappa(\mathbf{x},\mathbf{y})$ in its place. In this case,

$$A^{(\kappa)}_{n,m} \to \mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\log \frac{\kappa(\mathbf{x},\mathbf{y})}{\kappa(\mathbf{y})}\right] \text{ as } m,n \to \infty \tag{84}$$

where $\kappa(\mathbf{y}) = \mathbb{E}_{p(\mathbf{x})}[\kappa(\mathbf{x},\mathbf{y})]$. The same convergence rates apply.

4.3 Prior Contrastive Estimation and InfoNCE

We now consider the Prior Contrastive Estimation (PCE) estimator (Foster et al., 2020)

$$B_{n,m} = \frac{1}{n}\sum_{i=1}^{n} \log \frac{p(\mathbf{y}_i|\mathbf{x}_i)}{\frac{1}{m+1}\left(p(\mathbf{y}_i|\mathbf{x}_i) + \sum_{j=1}^{m} p(\mathbf{y}_i|\mathbf{x}_{ij})\right)}. \tag{85}$$

where $\mathbf{x}_i, \mathbf{y}_i \overset{\text{i.i.d.}}{\sim} p(\mathbf{x},\mathbf{y})$ and $\mathbf{x}_{ij} \overset{\text{i.i.d.}}{\sim} p(\mathbf{x})$ are independent. We can also re-use samples to make the variant

$$\tilde{B}_n = \frac{1}{n}\sum_{i=1}^{n} \log \frac{p(\mathbf{y}_i|\mathbf{x}_i)}{\frac{1}{n}\sum_{j=1}^{n} p(\mathbf{y}_i|\mathbf{x}_j)}. \tag{86}$$

It is more common to utilise this estimator in the case that $p(\mathbf{y}|\mathbf{x})$ is not known, leading to the InfoNCE estimator (van den Oord et al., 2018)

$$\tilde{B}^{(\kappa)}_n = \frac{1}{n}\sum_{i=1}^{n} \log \frac{\kappa(\mathbf{x}_i,\mathbf{y}_i)}{\frac{1}{n}\sum_{j=1}^{n} \kappa(\mathbf{x}_j,\mathbf{y}_i)} \tag{87}$$

for some positive function κ. Here, we focus on analysing the estimator $B_{n,m}$.

Before computing the asymptotic expansion of $B_{n,m}$, we present a basic result on its expectation.

Proposition 3 (Bounding the expectation of $B_{n,m}$)**.** *Assume*

$$\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right] < \infty. \tag{88}$$

Then,

$$0 \le I(\mathbf{x},\mathbf{y}) - \mathbb{E}[B_{n,m}] \le \frac{1}{m+1}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} - 1\right]. \tag{89}$$

This shows $B_{m,n}$ is negatively biased with bias of order $1/m$.

Proof. See Theorems 1 and 3 of Foster et al. (2020). □

Theorem 4 (Expectation of $B_{n,m}$)**.** *Suppose there exist Hölder conjugate indices $p, q > 0$ with $1/p + 1/q = 1$ such that*

$$\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right)^{3p}\right] < \infty \text{ and } \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{q}\right] < \infty. \tag{90}$$

Then we have

$$\begin{aligned}\mathbb{E}[B_{n,m}] =& I(\mathbf{x},\mathbf{y}) \\ &- \frac{1}{m}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} - 1\right] + \frac{1}{m}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{2p(\mathbf{y})^2}\right] \\ &+ \mathcal{O}\left(m^{-3/2}\right).\end{aligned} \tag{91}$$

Proof. By linearity, $\mathbb{E}[B_{n,m}] = \mathbb{E}[B_{1,m}]$. To compute this we define U_j as in Theorem 1, and we define $U_0 = p(\mathbf{x}_1|\mathbf{y}_1)/p(\mathbf{y}_1)$. We have

$$\mathbb{E}[B_{1,m}] = I(\mathbf{x}, \mathbf{y}) - \mathbb{E}\left[\log\left(\frac{1}{m+1}\sum_{j=0}^{m} U_j\right)\right]. \tag{92}$$

To reduce this to a more manageable form, we have

$$\mathbb{E}\left[\log\left(\frac{1}{m+1}\sum_{j=0}^{m} U_j\right)\right] = \log\left(\frac{m}{m+1}\right) + \mathbb{E}\left[\log\left(1 + \frac{U_0}{m} + \frac{1}{m}\sum_{j=1}^{m}(U_j - 1)\right)\right] \tag{93}$$

$$= \log\left(\frac{m}{m+1}\right) + \mathbb{E}\left[\log\left(1 + \frac{U_0}{m}\right)\right] + \mathbb{E}\left[\log\left(1 + \frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right)\right]. \tag{94}$$

Here, the third term involves a sum of conditionally i.i.d. random variables with mean zero. We now expand this third term with Taylor's Theorem

$$\mathbb{E}\left[\log\left(1 + \frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right)\right] = -\frac{1}{2}\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right)^2\right] + \mathbb{E}[\zeta] \tag{95}$$

We focus on controlling the ζ term. By Lemma 10 with $k = 2$ we have

$$|\zeta| \leq \left|\frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right|^3 \max\left(1, -\log\left(1 + \frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right)\right). \tag{96}$$

Since $U_0 > 0$, we must have

$$\left|\frac{U_j - 1}{1 + U_0/m}\right| \leq |U_j - 1|, \tag{97}$$

thus we can bound $\mathbb{E}[|\zeta|]$ by the exact error term that was considered in Theorem 1. This shows that $\mathbb{E}[|\zeta|] = \mathcal{O}(m^{-3/2})$. To calculate the expectation, we have

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right)^2\right] = \mathbb{E}\left[\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right)^2 \middle| \mathbf{x}_1, \mathbf{y}_1\right]\right] \tag{98}$$

$$= \frac{1}{m}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{1}{1 + U_0/m}\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right] \tag{99}$$

$$= \frac{1}{m}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{1}{1 + \frac{p(\mathbf{y}|\mathbf{x})}{mp(\mathbf{y})}}\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right], \tag{100}$$

this form offers easy comparison with Theorem 1. However, we have

$$\frac{1}{1 + \frac{p(\mathbf{y}|\mathbf{x})}{mp(\mathbf{y})}} = 1 + \mathcal{O}(m^{-1}) \tag{101}$$

and so we can drop the extract factor, giving

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}\frac{U_j - 1}{1 + U_0/m}\right)^2\right] = \frac{1}{m}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right] + \mathcal{O}(m^{-2}) \tag{102}$$

We also need to expand

$$\log\left(\frac{m}{m+1}\right)+\mathbb{E}\left[\log\left(1+\frac{U_0}{m}\right)\right]=\mathbb{E}\left[\log\left(1+\frac{U_0-1}{m+1}\right)\right] \tag{103}$$

$$=\mathbb{E}\left[\frac{U_0-1}{m+1}\right]+\mathcal{O}(m^{-3/2}) \tag{104}$$

$$=\frac{1}{m+1}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}-1\right]+\mathcal{O}(m^{-3/2}) \tag{105}$$

$$=\frac{1}{m}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}-1\right]+\mathcal{O}(m^{-3/2}) \text{ as the difference is order } m^{-2}. \tag{106}$$

Combining these gives the result. □

Theorem 5 (Variance of $B_{m,n}$)**.** *Assume that there exist Hölder conjugate indices $p,q>0$ such that*

$$\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right)^{3p}\right]<\infty \text{ and } \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{2q}\right]<\infty. \tag{107}$$

Then,

$$\begin{aligned}\mathrm{Var}[B_{n,m}]=&\frac{1}{n}\mathrm{Var}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right]\\&+\frac{1}{nm}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{2p(\mathbf{y})^2}\right]\\&+\frac{1}{nm}\mathrm{Cov}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})},-\frac{2p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}+\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right]\\&+\mathcal{O}\left(n^{-1}m^{-3/2}\right).\end{aligned} \tag{108}$$

Proof. We proceed using the same general strategy as Theorem 2. We have

$$\mathrm{Var}[B_{n,m}]=\frac{1}{n}\mathrm{Var}[B_{1,m}]. \tag{109}$$

By Tower Law,

$$\mathrm{Var}[B_{1,m}]=\mathbb{E}[\mathrm{Var}[B_{1,m}|\mathbf{x}_1,\mathbf{y}_1]]+\mathrm{Var}[\mathbb{E}[B_{1,m}|\mathbf{x}_1,\mathbf{y}_1]]. \tag{110}$$

For the conditional variance, using the notation of Theorem 4 we have

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[B_{1,m}|\mathbf{x}_1,\mathbf{y}_1]]=&\mathbb{E}\left[\mathbb{E}\left[\log\left(1+\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)^2\middle|\mathbf{x}_1,\mathbf{y}_1\right]\right]\\&-\mathbb{E}\left[\mathbb{E}\left[\log\left(1+\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)\middle|\mathbf{x}_1,\mathbf{y}_1\right]^2\right].\end{aligned} \tag{111}$$

For the first term of this variance, we use the analysis of $x\mapsto\log(x)^2$ that was done in Theorem 2 showing

$$|\log(x)^2-(x-1)^2|\leq|x-1|^3\max(1,-\log x)^2. \tag{112}$$

Thus,

$$\mathbb{E}\left[\mathbb{E}\left[\log\left(1+\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)^2\middle|\mathbf{x}_1,\mathbf{y}_1\right]\right]=\mathbb{E}\left[\left(1+\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)^2\right]+\mathbb{E}[\nu] \tag{113}$$

where

$$|\mathbb{E}[\nu]| \leq \mathbb{E}[|\nu|] \leq \mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right|^3 \max\left(1,-\log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)\right)^2\right] \tag{114}$$

$$\overset{\text{Hölder}}{\leq} \mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right|^{3p}\right]^{1/p} \mathbb{E}\left[\max\left(1,-\log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)\right)^{2q}\right]^{1/q} \tag{115}$$

$$\leq \mathbb{E}\left[\left|\frac{1}{m}\sum_{j=1}^{m}U_j-1\right|^{3p}\right]^{1/p} \mathbb{E}\left[\max\left(1,-\log\left(\frac{1}{m}\sum_{j=1}^{m}U_j-1\right)\right)^{2q}\right]^{1/q} \tag{116}$$

$$\overset{\text{Corollary 8}}{\leq} D_{3p}^{1/p}m^{-3/2}\mathbb{E}\left[|U_1-1|^{3p}\right]^{1/p} \mathbb{E}\left[\max\left(1,-\log\left(\frac{1}{m}\sum_{j=1}^{m}U_j-1\right)\right)^{2q}\right]^{1/q} \tag{117}$$

$$\overset{\text{convexity}}{\leq} D_{3p}^{1/p}m^{-3/2}\mathbb{E}\left[|U_1-1|^{3p}\right]^{1/p}\left(1+\mathbb{E}\left[|\log U_1|^{2q}\right]\right)^{1/q} \tag{118}$$

$$\leq D_{3p}^{1/p}m^{-3/2}\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left(\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right)^{3p}\right]^{1/p}\left(1+\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{2q}\right]\right)^{1/q}. \tag{119}$$

We also have, as previously

$$\mathbb{E}\left[\left(1+\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)^2\right] = \frac{1}{m}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{2p(\mathbf{y})^2}\right] + \mathcal{O}(m^{-2}). \tag{120}$$

On the other hand, Theorem 4 shows that

$$\mathbb{E}\left[\mathbb{E}\left[\log\left(1+\frac{1}{m}\sum_{j=1}^{m}\frac{U_j-1}{1+U_0/m}\right)\middle|\mathbf{x}_1,\mathbf{y}_1\right]^2\right] = \mathcal{O}(m^{-2}). \tag{121}$$

We can now turn to the variance of the conditional expectation. From Theorem 4, we know

$$\mathbb{E}[B_{1,m}|\mathbf{x}_1,\mathbf{y}_1] = \log\frac{p(\mathbf{y}_1|\mathbf{x}_1)}{p(\mathbf{y}_1)} + \frac{1}{m}\left(1-\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}+\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{2p(\mathbf{y})^2}\right)+\mathcal{O}\left(m^{-3/2}\right). \tag{122}$$

Thus,

$$\begin{aligned}\mathrm{Var}[\mathbb{E}[B_{1,m}|\mathbf{x}_1,\mathbf{y}_1]] = {}& \mathrm{Var}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right] \\ &+ \frac{1}{m}\mathrm{Cov}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})},-\frac{2p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}+\frac{\mathrm{Var}_{p(\mathbf{x}')}[p(\mathbf{y}|\mathbf{x}')]}{p(\mathbf{y})^2}\right] \\ &+ \mathcal{O}\left(m^{-3/2}\right).\end{aligned} \tag{123}$$

Putting the pieces together gives the final result. □

Finally, we note the key difference between the variance of the NMC and PCE estimators is the term

$$-\frac{1}{nm}\mathrm{Cov}_{p(\mathbf{x},\mathbf{y})}\left[\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})},\frac{2p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right]. \tag{124}$$

We would expect the covariance between a random variable and its logarithm to be positive, indicating that this term as a whole is negative. This, in turn, suggests that the PCE estimator has a lower variance than its NMC counterpart. However, focusing on the dominant terms, we still have the same overall NMC convergence rate of $T^{-1/3}$ in the total computational budget T.

4.4 Multi-level Monte Carlo

The following section covers material in Goda et al. (2020a), with Goda et al. (2020b) covering the extension to gradient estimators.

To begin, we define the random variables using the NMC estimator $A_{n,m}$ as our base

$$P_\ell = A_{1,M_\ell} \tag{125}$$

where M_ℓ is an increasing sequence of positive integers. From previous remarks, we know that $\mathbb{E}[P_\ell] \to I(\mathbf{x}, \mathbf{y})$ as $\ell \to \infty$. We now take $M_\ell = M_0 2^\ell$. We define the random variables Z_ℓ as follows

$$\begin{aligned} Z_\ell = &- \log\left(\frac{1}{M_\ell}\sum_{j=1}^{M_\ell} p(\mathbf{y}_1|\mathbf{x}_{1j})\right) \\ &+ \frac{1}{2}\left[\log\left(\frac{1}{M_{\ell-1}}\sum_{j=1}^{M_{\ell-1}} p(\mathbf{y}_1|\mathbf{x}_{1j})\right) + \log\left(\frac{1}{M_{\ell-1}}\sum_{j=1+M_{\ell-1}}^{M_\ell} p(\mathbf{y}_1|\mathbf{x}_{1j})\right)\right]. \end{aligned} \tag{126}$$

The key property of Z_ℓ is

$$\mathbb{E}[Z_\ell] = \mathbb{E}[P_\ell - P_{\ell-1}] \tag{127}$$

and the cost of computing Z_ℓ is bounded by $c2^\ell$. The main technical challenge is to bound the expectation and variance of Z_ℓ. We have the following theorem.

Theorem 6 (Goda et al. (2020a))**.** *Suppose there exist constants $p, q > 2$ such that $(p-2)(q-2) \geq 4$ such that*

$$\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^p\right] < \infty \quad \textit{and} \quad \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^q\right] < \infty. \tag{128}$$

Then,

$$\mathbb{E}\left[|Z_\ell|\right] = O(2^{-a\ell}), \qquad \mathrm{Var}(Z_\ell) = O(2^{-r\ell}) \tag{129}$$

where $a = \min\left(\frac{p(q-1)}{2q}, 1\right), r = \min\left(\frac{p(q-2)}{2q}, 2\right)$.

Proof. First, define

$$\beta_\ell^{(a)} = \frac{1}{M_\ell}\sum_{j=1}^{M_\ell}\frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)} \tag{130}$$

$$\beta_\ell^{(b)} = \frac{1}{M_\ell}\sum_{j=1+M_\ell}^{M_{\ell+1}}\frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)} \tag{131}$$

$$\text{so} \quad Z_\ell = -\log\beta_\ell^{(a)} + \frac{1}{2}\left(\log\beta_{\ell-1}^{(a)} + \log\beta_{\ell-1}^{(b)}\right) \tag{132}$$

$$\text{and} \quad \beta_\ell^{(a)} = \frac{1}{2}\left(\beta_{\ell-1}^{(a)} + \beta_{\ell-1}^{(b)}\right). \tag{133}$$

We then have

$$Z_\ell = -\log\beta_\ell^{(a)} + \frac{1}{2}\left(\log\beta_{\ell-1}^{(a)} + \log\beta_{\ell-1}^{(b)}\right) \tag{134}$$

$$= -\left(\log \beta_\ell^{(a)} - \beta_\ell^{(a)} + 1\right) + \frac{1}{2}\left(\log \beta_{\ell-1}^{(a)} - \beta_{\ell-1}^{(a)} + 1\right) + \frac{1}{2}\left(\log \beta_{\ell-1}^{(b)} - \beta_{\ell-1}^{(b)} + 1\right) \tag{135}$$

$$= 2\left[-\frac{1}{2}\left(\log \beta_\ell^{(a)} - \beta_\ell^{(a)} + 1\right) + \frac{1}{4}\left(\log \beta_{\ell-1}^{(a)} - \beta_{\ell-1}^{(a)} + 1\right) + \frac{1}{4}\left(\log \beta_{\ell-1}^{(b)} - \beta_{\ell-1}^{(b)} + 1\right)\right] \tag{136}$$

By convexity of $x \mapsto |x|^2$, we have

$$|Z_\ell|^2 \leq 2\left|\log \beta_\ell^{(a)} - \beta_\ell^{(a)} + 1\right|^2 + \left|\log \beta_{\ell-1}^{(a)} - \beta_{\ell-1}^{(a)} + 1\right|^2 + \left|\log \beta_{\ell-1}^{(b)} - \beta_{\ell-1}^{(b)} + 1\right|^2. \tag{137}$$

We use the following elementary inequality that holds for $1 \leq r \leq 2$

$$|\log x - x + 1| \leq |x-1|^r \max(-\log x, 1) \tag{138}$$

which gives

$$\left|\log \beta_\ell^{(a)} - \beta_\ell^{(a)} + 1\right|^2 \leq \left|\beta_\ell^{(a)} - 1\right|^{2r} \left(\max\left(-\log \beta_\ell^{(a)}, 1\right)\right)^2. \tag{139}$$

We now take the expectation and apply Hölder's Inequality with $1/s + 1/t = 1$, giving

$$\mathbb{E}\left[\left|\log \beta_\ell^{(a)} - \beta_\ell^{(a)} + 1\right|^2\right] \leq \left\|\left|\beta_\ell^{(a)} - 1\right|^{2r}\right\|_{L^s} \left\|\left(\max\left(-\log \beta_\ell^{(a)}, 1\right)\right)^2\right\|_{L^t}. \tag{140}$$

For the first term, we apply Corollary 8 to conclude that

$$\left\|\left|\beta_\ell^{(a)} - 1\right|^{2r}\right\|_{L^s} \leq D_{2rs}^{1/s} \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{2sr}\right]^{1/s} (M_0 2^\ell)^{-r}, \tag{141}$$

for the second term, we use the fact that the functions $x \mapsto \max(-\log x, 1)$ and $x \mapsto x^{2t}$ are convex to give

$$\left\|\left(\max\left(-\log \beta_\ell^{(a)}, 1\right)\right)^2\right\|_{L^t} \leq \left\|\left(\frac{1}{M_\ell}\sum_{j=1}^{M_\ell} \max\left(-\log \frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)}, 1\right)\right)^2\right\|_{L^t} \tag{142}$$

$$= \left(\mathbb{E}\left[\left(\frac{1}{M_\ell}\sum_{j=1}^{M_\ell} \max\left(-\log \frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)}, 1\right)\right)^{2t}\right]\right)^{1/t} \tag{143}$$

$$\leq \left(\frac{1}{M_\ell}\mathbb{E}\left[\sum_{j=1}^{M_\ell} \max\left(-\log \frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)}, 1\right)^{2t}\right]\right)^{1/t} \tag{144}$$

$$\leq \left(\frac{1}{M_\ell}\mathbb{E}\left[\sum_{j=1}^{M_\ell} \left|\log \frac{p(\mathbf{y}_1|\mathbf{x}_{1j})}{p(\mathbf{y}_1)}\right|^{2t} + 1\right]\right)^{1/t} \tag{145}$$

$$= \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^{2t} + 1\right]^{1/t}. \tag{146}$$

We now choose $s = q/(q-2), t = q/2$ and $r = \min(p(q-2)/2q, 2)$. This gives

$$\mathbb{E}\left[\left|\log \beta_\ell^{(a)} - \beta_\ell^{(a)} + 1\right|^2\right] \leq A_0 2^{-r\ell} \tag{147}$$

where

$$A_0 = D_{2rs}^{1/s} \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^p\right]^{1/s} \mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}\left[\left|\log \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right|^q + 1\right]^{1/t} M_0^{-r}. \tag{148}$$

Since we can bound the other two terms of (137) in a similar way, we obtain a bound on $\mathrm{Var}(Z_\ell)$ that is of order $2^{-r\ell}$. A similar proof gives the bound for $\mathbb{E}[|Z_\ell|]$. □

An important result of this theorem is that we can obtain a MLMC estimator of $I(\mathbf{x}, \mathbf{y})$ with total cost T that converges at a rate $\mathcal{O}(T^{-1/2})$ in root mean square. This is achieved using standard MLMC technology (Giles, 2008). We define, analogously to the NMC case

$$\begin{aligned} Z_{n,\ell} =& \frac{1}{n}\sum_{i=1}^{n}\left[-\log\left(\frac{1}{M_\ell}\sum_{j=1}^{M_\ell} p(\mathbf{y}_i|\mathbf{x}_{ij})\right)\right. \\ &\left. + \frac{1}{2}\left[\log\left(\frac{1}{M_{\ell-1}}\sum_{j=1}^{M_{\ell-1}} p(\mathbf{y}_i|\mathbf{x}_{ij})\right) + \log\left(\frac{1}{M_{\ell-1}}\sum_{j=1+M_{\ell-1}}^{M_\ell} p(\mathbf{y}_i|\mathbf{x}_{ij})\right)\right]\right]. \end{aligned} \tag{149}$$

Then

$$Z_L^{\mathrm{MLMC}} = \sum_{\ell=0}^{L} Z_{N_\ell,\ell} \tag{150}$$

with

$$\mathbb{E}\left[\left|Z_L^{\mathrm{MLMC}} - I(\mathbf{x},\mathbf{y})\right|^2\right] = \sum_{\ell=0}^{L}\frac{\mathrm{Var}[Z_\ell]}{N_\ell} + \left[\mathbb{E}[P_L] - I(\mathbf{x},\mathbf{y})\right]^2. \tag{151}$$

The cost of the estimator Z_L^{MLMC} is $\mathcal{O}(N_L M_L)$. What Theorem 6 shows is that the bias and variance of the Z_ℓ decay fast enough to offset the growth in cost. For full details, see Goda et al. (2020a).

4.5 A rigorous delta method for the natural logarithm

This self-contained section includes some of the mathematical machinery that is relied upon by the rest of this work. As previously mentioned, most analyses of mutual information estimators (Zheng et al., 2018; Beck et al., 2018; Rainforth et al., 2018) utilise the delta method for moments. Unfortunately, the standard delta method that we derive here in Lemma 9 is not valid for the natural logarithm function, because none of its derivatives are bounded on $(0, \infty)$. In this section, we derive a rigorous delta method for the logarithm. Whilst this is *not* sufficient for all the Theorems in the preceding sections, it highlights and essentialises the key technical pieces required.

We begin with the Marcinkiewicz–Zygmund Inequality, which is used to derive the standard delta method.

Lemma 7 (Marcinkiewicz and Zygmund (1937))**.** *Let $X_1, \ldots, X_m$ be independent random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{E}\left[|X_i|^p\right] < \infty$. Then there exists a constant D_p such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^{m}(X_i - \mu)\right|^p\right) \leq D_p \mathbb{E}\left(\left(\sum_{i=1}^{m}|X_i|^2\right)^{p/2}\right) \tag{152}$$

Corollary 8. *Let $X_1, \ldots, X_m$ be i.i.d. random variables with $\mathbb{E}[X_1] = \mu$ and $\mathbb{E}\left[|X_1|^p\right] < \infty$. Then there exists a constant D_p such that*

$$\mathbb{E}\left(\left|\frac{1}{m}\sum_{i=1}^{m}(X_i - \mu)\right|^p\right) \leq D_p m^{-p/2}\mathbb{E}\left[|X_1|^p\right] \tag{153}$$

Proof. Applying the Marcinkiewicz–Zygmund Inequality, we have

$$\mathbb{E}\left(\left|\frac{1}{m}\sum_{i=1}^{m}(X_i - \mu)\right|^p\right) \leq D_p m^{-p/2}\mathbb{E}\left(\left(\frac{1}{m}\sum_{i=1}^{m}|X_i|^2\right)^{p/2}\right), \tag{154}$$

by the convexity of $x \mapsto x^{p/2}$ on $(0, \infty)$, we have

$$\leq D_p m^{-p/2}\mathbb{E}\left(\frac{1}{m}\sum_{i=1}^{m}|X_i|^p\right) \tag{155}$$

$$= D_p m^{-p/2} \mathbb{E}\left[|X_1|^p\right]. \tag{156}$$

□

Notice that Corollary 8 essentially gives the asymptotic moments that would be expected from the Central Limit Theorem, although they cannot be derived from the standard Central Limit Theorem which gives convergence *in distribution* only.

Lemma 9 (Delta method of order k)**.** *Let X_i be a sequence of i.i.d. random variables with mean μ and $\mathbb{E}\left[|X_1|^{k+1}\right] < \infty$, and let f be a smooth function with $\|f^{(k+1)}\|_\infty = M < \infty$. Then*

$$\mathbb{E}\left[f\left(\frac{1}{m}\sum_{i=1}^m X_i\right)\right] = \sum_{j=0}^k \frac{f^{(j)}(\mu)}{j!}\mathbb{E}\left[\left(\frac{1}{m}\sum_{i=1}^m (X_i-\mu)\right)^j\right] + \mathcal{O}\left(m^{-(k+1)/2}\right). \tag{157}$$

Proof. By Taylor's Theorem with Lagrange's form of the remainder, we have for any x and for some ξ between x and μ

$$f(x) = \sum_{j=0}^k \frac{f^{(j)}(\mu)}{j!}(x-\mu)^k + \frac{f^{(k+1)}(\xi)}{(k+1)!}(x-\mu)^{k+1}. \tag{158}$$

Applying this to $\frac{1}{m}\sum_{i=1}^m X_i$ and taking the expectation gives

$$\mathbb{E}\left[f\left(\frac{1}{m}\sum_{i=1}^m X_i\right)\right] = \sum_{j=0}^k \frac{f^{(j)}(\mu)}{j!}\mathbb{E}\left[\left(\frac{1}{m}\sum_{i=1}^m (X_i-\mu)\right)^j\right] + \mathbb{E}\left[\frac{f^{(k+1)}(\Xi)}{(k+1)!}\left(\frac{1}{m}\sum_{i=1}^m (X_i-\mu)\right)^{k+1}\right] \tag{159}$$

where Ξ is a random variable between μ and $\frac{1}{m}\sum_i X_i$. By assumption, we have $f^{(k+1)}(\Xi) \le M$. By Corollary 8, we have

$$\mathbb{E}\left(\left|\sum_{i=1} X_i - \mu\right|^{k+1}\right) \le D_{k+1} m^{(k+1)/2} \mathbb{E}\left[|X_1|^{k+1}\right]. \tag{160}$$

Hence we conclude that

$$\left|\mathbb{E}\left[\frac{f^{(k+1)}(\Xi)}{(k+1)!}\left(\frac{1}{m}\sum_{i=1}^m (X_i-\mu)\right)^{k+1}\right]\right| \le \frac{M D_{k+1}\mathbb{E}\left[|X_1|^{k+1}\right] m^{-(k+1)/2}}{(k+1)!} = \mathcal{O}(m^{-(k+1)/2}). \tag{161}$$

□

We now turn to the logarithm function in particular, bounding the difference between the function and its series approximation.

Lemma 10. *Define*

$$L_k(x) = \sum_{j=1}^k \frac{(-1)^{j+1}}{j}(x-1)^j. \tag{162}$$

Then $|\log x - L_k(x)| \le |x-1|^{k+1}\max(1, -\log x)$ for $0 < x < \infty$.

Proof. By Taylor's Theorem with Cauchy's form of the remainder, for any $0 < x < \infty$ there exists ξ that is between 1 and x such that

$$\log x = L_k(x) + \frac{(-1)^{k+2}}{\xi^{k+1}}(x-\xi)^k(x-1) \tag{163}$$

For $x > 1$, we must have $\xi^{k+1} > 1$, so $|\log x - L_k(x)| < |x-\xi|^k|x-1| < |x-1|^{k+1}$.

For $x \leq 1$, we have

$$\frac{\xi - x}{\xi} = 1 - x/\xi \text{ and } 0 \leq 1 - x/\xi \leq 1 - x \text{ since } x \leq \xi \leq 1. \tag{164}$$

Thus, the magnitude of the remainder term becomes

$$\left|\frac{(-1)^{k+2}}{\xi^{k+1}}(x-\xi)^k(x-1)\right| = \left|\left(\frac{\xi - x}{\xi}\right)^k \frac{x-1}{\xi}\right| \leq (1-x)^k \left|\frac{x-1}{\xi}\right| \leq \frac{(1-x)^{k+1}}{x} \tag{165}$$

which shows that the Taylor series for the logarithm is convergent on $(0, 1]$. Therefore, we have

$$\log x - L_k(x) = \sum_{j=k+1}^{\infty} \frac{(-1)^{j+1}}{j}(x-1)^j \tag{166}$$

$$= (x-1)^{k+1}(-1)^k \left(\frac{1}{k+1} - \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{k+1+j}(x-1)^j\right) \tag{167}$$

noting that $x - 1 \leq 0$ we see that each term of the sum has the same sign, giving

$$= -|x-1|^{k+1} \left(\frac{1}{k+1} + \sum_{j=1}^{\infty} \frac{1}{k+1+j}|x-1|^j\right). \tag{168}$$

If $x \geq e^{-1}$, we have

$$\frac{1}{k+1} + \sum_{j=1}^{\infty} \frac{1}{k+1+j}|x-1|^j \leq \frac{1}{k+1} + \sum_{j=1}^{\infty} \frac{1}{k+1+j}|e-1|^j \tag{169}$$

by monotonicity. If $x \leq e^{-1}$, we have

$$\frac{1}{k+1} + \sum_{j=1}^{\infty} \frac{1}{k+1+j}|x-1|^j \leq \frac{1}{k+1} + \frac{|x-1|}{k+2} + \sum_{j=2}^{\infty} \frac{|x-1|^j}{k+1+j} \tag{170}$$

$$\leq |x-1| + \sum_{j=2}^{\infty} \frac{|x-1|^j}{j} = -\log x, \tag{171}$$

for any $k \geq 1$. Combining these, we have

$$\frac{1}{k+1} + \sum_{j=1}^{\infty} \frac{1}{k+1+j}|x-1|^j \leq \max(-\log x, \log e) = \max(-\log x, 1). \tag{172}$$

□

For the following Proposition, the logic is inspired by Goda et al. (2020a).

Proposition 11 (Rigorous delta method for the logarithm)**.** *Let $U_1, \ldots, U_m$ be a sequence of i.i.d. positive random variables with $\mathbb{E}[U_1] = 1$. Fix a natural number $k \geq 1$. Suppose that for Hölder conjugate indices $p, q > 0$ with $1/p + 1/q = 1$, we have $\mathbb{E}\left[U_1^{(k+1)p}\right] < \infty$ and $\mathbb{E}[|\log U_1|^q] < \infty$. Then,*

$$\mathbb{E}\left[\log\left(\frac{1}{m}\sum_{i=1}^{m} U_i\right)\right] = \sum_{j=2}^{k} \frac{(-1)^{j+1}}{j}\mathbb{E}\left[\left(\frac{1}{m}\sum_{i=1}^{m}(U_i - 1)\right)^j\right] + E_k \tag{173}$$

where $E_k = \mathcal{O}\left(m^{-(k+1)/2}\right)$.

Proof. Define L_k as in Lemma 10. By that Lemma, we have

$$\left|\log\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right) - L_k\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right)\right| \leq \left|\frac{1}{m}\sum_{i=1}^{m}U_i\right|^{k+1} \max\left(-\log\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right), 1\right). \tag{174}$$

We see that

$$\mathbb{E}\left[L_k\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right)\right] = \sum_{j=1}^{k}\frac{(-1)^{j+1}}{j}\mathbb{E}\left[\left(\frac{1}{m}\sum_{i=1}^{m}(U_i-1)\right)^j\right] \tag{175}$$

and $\mathbb{E}[U_i - 1] = 0$.

The error term E_k is bounded in L_1 by

$$\mathbb{E}[|E_k|] \leq \mathbb{E}\left[\left|\frac{1}{m}\sum_{i=1}^{m}U_i\right|^{k+1} \max\left(-\log\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right), 1\right)\right] \tag{176}$$

apply Hölder's Inequality to give

$$\leq \mathbb{E}\left[\left|\frac{1}{m}\sum_{i=1}^{m}U_i\right|^{p(k+1)}\right]^{1/p} \mathbb{E}\left[\max\left(-\log\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right), 1\right)^q\right]^{1/q}. \tag{177}$$

For the first term, Corollary 8 shows that

$$\mathbb{E}\left[\left|\frac{1}{m}\sum_{i=1}^{m}U_i\right|^{p(k+1)}\right]^{1/p} \leq D_{(k+1)p}^{1/p} m^{-(k+1)/2}\mathbb{E}\left[U_1^{(k+1)p}\right]^{1/p} \tag{178}$$

for the second term we use the fact that $x \mapsto \max(-\log x, 1)$ is a convex function, so

$$\max\left(-\log\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right), 1\right)^q \leq \frac{1}{m}\sum_{i=1}^{m}\max\left(-\log\left(U_i\right), 1\right)^q \tag{179}$$

$$\leq \frac{1}{m}\sum_{i=1}^{m}\left(|\log U_i| + 1\right)^q, \tag{180}$$

hence

$$\mathbb{E}\left[\max\left(-\log\left(\frac{1}{m}\sum_{i=1}^{m}U_i\right), 1\right)^q\right]^{1/q} \leq \left(\mathbb{E}\left[|\log U_1|^q\right] + 1\right)^{1/q}. \tag{181}$$

By assumption, we have $\mathbb{E}\left[U_1^{(k+1)p}\right] < \infty$ and $\mathbb{E}\left[|\log U_1|^q\right] < \infty$. Putting the pieces together, we have

$$\mathbb{E}[|E_k|] \leq m^{-(k+1)/2} D_{(k+1)p}^{1/p}\mathbb{E}\left[U_1^{(k+1)p}\right]^{1/p}\left(\mathbb{E}\left[|\log U_1|^q\right] + 1\right)^{1/q}, \tag{182}$$

so E_k is $\mathcal{O}\left(m^{-(k+1)/2}\right)$ as required. □

Notice that we recover the regular delta method with $\log U_i$ bounded if $p = 1, q = \infty$.

5 The generalized Donsker-Varadhan representation

5.1 Introduction

In this essay, we present a generalization of the classical Donsker-Varadhan representation of the KL-divergence (Donsker and Varadhan, 1975). Our purpose is twofold. Firstly, the new representation of the KL-divergence sheds further light on mutual information and may motivate the development of new statistical estimators of information. Secondly, our new representation is a powerful tool that connects a number of *existing* mutual information estimators under one umbrella. An important feature of the generalized Donsker-Varadhan representation is that it includes self-normalized bounds such as InfoNCE (van den Oord et al., 2018) as a special case, something which is not true of the classical Donsker-Varadhan representation.

5.2 Information-theoretic quantities

Throughout machine learning, we have cause to consider the entropy of probability measure p

$$\mathrm{H}\left(p\right) = \mathbb{E}_{p(\mathbf{x})}[-\log p(\mathbf{x})], \tag{183}$$

the KL divergence between two probability measures $p \ll q$

$$\mathrm{KL}\left(p \,\|\, q\right) = \mathbb{E}_{p(\mathbf{x})}\left[\log \frac{p(\mathbf{x})}{q(\mathbf{x})}\right] \tag{184}$$

and the mutual information between jointly distributed random variables $\mathbf{x}, \mathbf{y} \sim p(\mathbf{x}, \mathbf{y})$

$$I(\mathbf{x}, \mathbf{y}) = \mathrm{KL}\left(p(\mathbf{x}, \mathbf{y}) \,\|\, p(\mathbf{x})p(\mathbf{y})\right). \tag{185}$$

These are foundational quantities in information theory (Shannon, 1948), Bayesian experimental design (Lindley, 1956) and deep learning (Linsker, 1988). A key result in information theory is the following.

Theorem 12 (Gibbs' Inequality)**.** *For any probability measures $p \ll q$, $\mathrm{KL}\left(p \,\|\, q\right) \geq 0$.*

5.3 The Donsker-Varadhan representation

An important lower bound on the KL divergence is the Donsker-Varadhan (DV) representation.

Theorem 13 (Donsker and Varadhan (1975))**.** *Let $p \ll q$ be probability measures on $\mathcal{X}$, then*

$$\mathrm{KL}\left(p \,\|\, q\right) = \sup_{T:\mathcal{X}\to\mathbb{R} \text{ measurable}} \mathbb{E}_{p(\mathbf{x})}[T(\mathbf{x})] - \log\left(\mathbb{E}_{q(\mathbf{x})}\left[\exp(T(\mathbf{x}))\right]\right) \tag{186}$$

One important bound that can be obtained as a consequence of the Donsker-Varadhan representation is the following.

Corollary 14 (Barber and Agakov (2003))**.** *Let $q(\mathbf{y}|\mathbf{x})$ be a conditional distribution. Then*

$$I(\mathbf{x}, \mathbf{y}) \geq \mathbb{E}_{p(\mathbf{x},\mathbf{y})}\left[\log \frac{q(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right] \tag{187}$$

Proof. Since mutual information is defined as a KL divergence, the DV representation is applicable. Let $T(\mathbf{x}, \mathbf{y}) = \log q(\mathbf{x}, \mathbf{y})/p(\mathbf{y})$ in Theorem 13. We have

$$\mathbb{E}_{p(\mathbf{x})p(\mathbf{y})}[q(\mathbf{y}|\mathbf{x})/p(\mathbf{y})] = 1 \tag{188}$$

so the bound is *self-normalized*. The result follows. □

The Barber-Agakov bound can be written as

$$I(\mathbf{x}, \mathbf{y}) \geq \mathbb{E}_{p(\mathbf{x},\mathbf{y})} \left[\log q(\mathbf{y}|\mathbf{x})\right] + \mathrm{H}\left(p(\mathbf{y})\right) \tag{189}$$

which can be helpful in cases in which the $\mathrm{H}\left(p(\mathbf{y})\right)$ term is unknown but also unneeded for e.g. gradient estimation. Another bound, that appears in Nguyen et al. (2010); Nowozin et al. (2016); Belghazi et al. (2018) has a connection to the theory of f-divergences. Applying the inequality $\log x \leq e^{-1}x$ to Theorem 13 gives the **NWJ bound**

$$I(\mathbf{x}, \mathbf{y}) \geq \mathbb{E}_{p(\mathbf{x})}[T(\mathbf{x})] - e^{-1}\mathbb{E}_{q(\mathbf{x})}\left[\exp(T(\mathbf{x}))\right]. \tag{190}$$

An advantage of this looser bound is that it can be directly estimated by samples.

5.4 A generalization of the Donsker-Varadhan representation

To generalize Theorem 13, suppose we extend the sample space to $\mathcal{X} \times \mathcal{S}$, where $\mathcal{S}$ represents 'side-information'. Suppose we have a conditional distribution $p(\mathbf{s}|\mathbf{x})$. Then we can extend the Donsker-Varadhan representation as follows.

Theorem 15 (Generalized Donsker-Varadhan representation)**.** *Under the assumptions of Theorem 13, let $p(\mathbf{s}|\mathbf{x})$ be a valid conditional distribution for each $\mathbf{x} \in \mathcal{X}$. Then,*

$$\mathrm{KL}\left(p \,\|\, q\right) = \sup_{U:\mathcal{X}\times\mathcal{S}\to\mathbb{R} \text{ measurable}} \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[U(\mathbf{x}, \mathbf{s})] - \log\left(\mathbb{E}_{q(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\left[\exp(U(\mathbf{x}, \mathbf{s}))\right]\right) \tag{191}$$

Proof. Since any function $T : \mathcal{X} \to \mathbb{R}$ can be extended to a new function on $\mathcal{X} \times \mathcal{S}$ by ignoring the side information, Theorem 13 immediately tells us that

$$\mathrm{KL}\left(p \,\|\, q\right) \leq \sup_{U:\mathcal{X}\times\mathcal{S}\to\mathbb{R} \text{ measurable}} \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[U(\mathbf{x}, \mathbf{s})] - \log\left(\mathbb{E}_{q(\mathbf{x})p(\mathbf{s},\mathbf{x})}\left[\exp(U(\mathbf{x}, \mathbf{s}))\right]\right). \tag{192}$$

To prove the $\geq$ inequality, we consider some measurable $U : \mathcal{X} \times \mathcal{S} \to \mathbb{R}$. We have

$$\mathrm{KL}\left(p \,\|\, q\right) = \mathbb{E}_{p(\mathbf{x})}\left[\log \frac{p(\mathbf{x})}{q(\mathbf{x})}\right] \tag{193}$$

$$= \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\left[\log \frac{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}{q(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\right] \tag{194}$$

define $V(\mathbf{x}, \mathbf{s}) = \exp(U(\mathbf{x}, \mathbf{s}))/\mathbb{E}_{q(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[\exp(U(\mathbf{x}, \mathbf{s}))]$

$$= \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\left[\log \frac{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}{q(\mathbf{x})p(\mathbf{s}|\mathbf{x})V(\mathbf{x}, \mathbf{s})}\right] + \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[\log V(\mathbf{x}, \mathbf{s})] \tag{195}$$

now note that by definition of V, $\int_{\mathcal{X}\times\mathcal{S}} q(\mathbf{x})p(\mathbf{x}|\mathbf{s})V(\mathbf{x}, \mathbf{s}) = 1$, so $q(\mathbf{x})p(\mathbf{x}|\mathbf{s})V(\mathbf{x}, \mathbf{s})$ is a probability measure

$$= \mathrm{KL}\left(p(\mathbf{x})p(\mathbf{s}|\mathbf{x}) \,\|\, q(\mathbf{x})p(\mathbf{s}|\mathbf{x})V(\mathbf{x}|\mathbf{s})\right) + \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[\log V(\mathbf{x}, \mathbf{s})] \tag{196}$$

now by Gibbs' Inequality

$$\geq \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[\log V(\mathbf{x}, \mathbf{s})] \tag{197}$$

$$= \mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[U(\mathbf{x}, \mathbf{s})] - \log\left(\mathbb{E}_{q(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\left[\exp(U(\mathbf{x}, \mathbf{s}))\right]\right). \tag{198}$$

This completes the proof. □

5.5 Self-normalized bounds

One particular use of Theorem 15 is for cases in which $\mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\left[\exp(U(\mathbf{x},\mathbf{s}))\right] = 1$. For such a self-normalized bound, the task of estimating the potentially high-dimensional term $\mathbb{E}_{q(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\left[\exp(U(\mathbf{x},\mathbf{s}))\right]$ is removed, and the bound reduces to $\mathbb{E}_{p(\mathbf{x})p(\mathbf{s}|\mathbf{x})}[U(\mathbf{x},\mathbf{s})]$ for which unbiased estimators can be constructed directly from samples.

Theorem 16 (Self-normalized KL bound)**.** *Let $k : \mathcal{X} \to \mathbb{R}$ be any measurable function. Then we have the following bound on the KL divergence*

$$\mathrm{KL}\left(p \,\|\, q\right) \leq \mathbb{E}_{p(\mathbf{x}_1)q(\mathbf{x}_2)\ldots q(\mathbf{x}_m)}\left[\log \frac{\exp(k(\mathbf{x}_1))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i))}\right]. \tag{199}$$

Proof. We apply Theorem 15 with $\mathbf{x} = \mathbf{x}_1$, $\mathcal{S} = \mathcal{X}^{m-1}$, $\mathbf{s} = (\mathbf{x}_2, ..., \mathbf{x}_m)$ and $p(\mathbf{s}|\mathbf{x}) = q(\mathbf{x}_2) \cdot \ldots \cdot q(\mathbf{x}_m)$ is independent of $\mathbf{x}_1$. We have

$$U(\mathbf{x},\mathbf{s}) = \log \frac{\exp(k(\mathbf{x}))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i))} \tag{200}$$

To apply the theorem, we consider

$$\mathbb{E}_{q(\mathbf{x})p(\mathbf{s}|\mathbf{x})}\left[\exp(U(\mathbf{x},\mathbf{s}))\right] = \mathbb{E}_{q(\mathbf{x}_1)\ldots q(\mathbf{x}_m)}\left[\frac{\exp(k(\mathbf{x}))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i))}\right]. \tag{201}$$

Since the $\mathbf{x}_1, ..., \mathbf{x}_m$ are all equal in distribution, we can replace the index of the sample used in the numerator by any $j \in \{1, ..., m\}$

$$= \mathbb{E}_{q(\mathbf{x}_1)\ldots q(\mathbf{x}_m)}\left[\frac{\exp(k(\mathbf{x}_j))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i))}\right] \tag{202}$$

we can take the mean over all possible values of j

$$= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{q(\mathbf{x}_1)\ldots q(\mathbf{x}_m)}\left[\frac{\exp(k(\mathbf{x}_j))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i))}\right] \tag{203}$$

now by linearity of the expectation we have

$$= \mathbb{E}_{q(\mathbf{x}_1)\ldots q(\mathbf{x}_m)}\left[\frac{\frac{1}{m}\sum_{j=1}^{m}\exp(k(\mathbf{x}_j))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i))}\right] \tag{204}$$

$$= 1. \tag{205}$$

Thus the bound is self-normalized and the result follows. □

We note that this bound cannot typically recover the KL divergence, because

$$\log \frac{\exp(k(\mathbf{x}_1))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i))} \leq \log \frac{\exp(k(\mathbf{x}))}{\frac{1}{m}\exp(k(\mathbf{x}))} = \log m. \tag{206}$$

We can apply a related idea to mutual information. The following theorem provides a self-normalized bound on $I(\mathbf{x},\mathbf{y})$ that is closely related to the popular InfoNCE (van den Oord et al., 2018) bound.

Theorem 17 (Self-normalized information bound)**.** *Let $k : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be any measurable function. Then we have the following bound on the mutual information*

$$I(\mathbf{x},\mathbf{y}) \leq \mathbb{E}_{p(\mathbf{x}_1,\mathbf{y}_1)p(\mathbf{x}_2)\ldots p(\mathbf{x}_m)}\left[\log \frac{\exp(k(\mathbf{x}_1,\mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^{m}\exp(k(\mathbf{x}_i,\mathbf{y}_1))}\right]. \tag{207}$$

Proof. Since $I(\mathbf{x}, \mathbf{y}) = \mathrm{KL}\left(p(\mathbf{x}, \mathbf{y}) \,\|\, p(\mathbf{x})p(\mathbf{y})\right)$, we can apply Theorem 15. We set $\mathcal{S} = \mathcal{X}^{m-1}$ and $\mathbf{s} = (\mathbf{x}_2, ..., \mathbf{x}_m)$. We have

$$U((\mathbf{x}_1, \mathbf{y}_1), \mathbf{s}) = \log \frac{\exp(k(\mathbf{x}_1, \mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^m \exp(k(\mathbf{x}_i, \mathbf{y}_1))}. \tag{208}$$

To show that this bound is self-normalized, we consider

$$\mathbb{E}_{p(\mathbf{x}_1)p(\mathbf{y}_1)p(\mathbf{s})}[\exp(U((\mathbf{x}_1, \mathbf{y}_1), \mathbf{s}))] = \mathbb{E}_{p(\mathbf{x}_1)...p(\mathbf{x}_m)p(\mathbf{y}_1)}\left[\frac{\exp(k(\mathbf{x}_1, \mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^m \exp(k(\mathbf{x}_i, \mathbf{y}_1))}\right], \tag{209}$$

for any $\ell \in \{1, ..., m\}$, we have

$$= \mathbb{E}_{p(\mathbf{x}_1)...p(\mathbf{x}_m)p(\mathbf{y}_1)}\left[\frac{\exp(k(\mathbf{x}_\ell, \mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^m \exp(k(\mathbf{x}_i, \mathbf{y}_1))}\right] \tag{210}$$

since the $\mathbf{x}_i$ are all equal in distribution. Then,

$$= \frac{1}{m}\sum_{\ell=1}^m \mathbb{E}_{p(\mathbf{x}_1)...p(\mathbf{x}_m)p(\mathbf{y}_1)}\left[\frac{\exp(k(\mathbf{x}_\ell, \mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^m \exp(k(\mathbf{x}_i, \mathbf{y}_1))}\right] \tag{211}$$

$$= \mathbb{E}_{p(\mathbf{x}_1)...p(\mathbf{x}_m)p(\mathbf{y}_1)}\left[\frac{\frac{1}{m}\sum_{\ell=1}^m \exp(k(\mathbf{x}_\ell, \mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^m \exp(k(\mathbf{x}_i, \mathbf{y}_1))}\right] \tag{212}$$

$$= 1. \tag{213}$$

This completes the proof. □

Finally, it is possible to change the distribution that is used to generate $\mathbf{s}$ as long as we compensate with importance weighting. The following theorem gives a bound that is closely connected to the likelihood-free Adaptive Contrastive Estimation bound of Foster et al. (2020) eq. (14).

Theorem 18 (Importance weighted self-normalized information bound)**.** *Let $k : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be any measurable function. Consider a conditional distribution $q(\mathbf{x}'|\mathbf{y})$ on $\mathcal{X}$. Then we have the following bound on the mutual information*

$$I(\mathbf{x}, \mathbf{y}) \leq \mathbb{E}_{p(\mathbf{x}_1, \mathbf{y}_1)q(\mathbf{x}_2|\mathbf{y}_1)...q(\mathbf{x}_m|\mathbf{y}_1)}\left[\log \frac{\exp(k(\mathbf{x}_1, \mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^m \frac{\exp(k(\mathbf{x}_i, \mathbf{y}_1))p(\mathbf{x}_i)}{q(\mathbf{x}_i|\mathbf{y}_1)}}\right]. \tag{214}$$

Proof. Following the same strategy as the previous two proofs, we consider

$$\mathbb{E}_{p(\mathbf{x}_1)p(\mathbf{y}_1)p(\mathbf{s})}[\exp(U((\mathbf{x}_1, \mathbf{y}_1), \mathbf{s}))] = \mathbb{E}_{p(\mathbf{x}_1)p(\mathbf{y}_1)q(\mathbf{x}_{2:m}|\mathbf{y}_1)}\left[\frac{\exp(k(\mathbf{x}_1, \mathbf{y}_1))}{\frac{1}{m}\sum_{i=1}^m \frac{\exp(k(\mathbf{x}_i, \mathbf{y}_1))p(\mathbf{x}_i)}{q(\mathbf{x}_i|\mathbf{y}_1)}}\right] \tag{215}$$

$$= \mathbb{E}_{p(\mathbf{y}_1)q(\mathbf{x}_{1:m}|\mathbf{y}_1)}\left[\frac{\frac{\exp(k(\mathbf{x}_1, \mathbf{y}_1))p(\mathbf{x}_1)}{q(\mathbf{x}_1|\mathbf{y}_1)}}{\frac{1}{m}\sum_{i=1}^m \frac{\exp(k(\mathbf{x}_i, \mathbf{y}_1))p(\mathbf{x}_i)}{q(\mathbf{x}_i|\mathbf{y}_1)}}\right] \tag{216}$$

for any $\ell \in \{1, ..., m\}$, we have

$$= \mathbb{E}_{p(\mathbf{y}_1)q(\mathbf{x}_{1:m}|\mathbf{y}_1)}\left[\frac{\frac{\exp(k(\mathbf{x}_\ell, \mathbf{y}_1))p(\mathbf{x}_\ell)}{q(\mathbf{x}_\ell|\mathbf{y}_1)}}{\frac{1}{m}\sum_{i=1}^m \frac{\exp(k(\mathbf{x}_i, \mathbf{y}_1))p(\mathbf{x}_i)}{q(\mathbf{x}_i|\mathbf{y}_1)}}\right] \tag{217}$$

since the $\mathbf{x}_i$ are all now equal in distribution. Then,

$$= \frac{1}{m}\sum_{\ell=1}^m \mathbb{E}_{p(\mathbf{y}_1)q(\mathbf{x}_{1:m}|\mathbf{y}_1)}\left[\frac{\frac{\exp(k(\mathbf{x}_\ell, \mathbf{y}_1))p(\mathbf{x}_\ell)}{q(\mathbf{x}_\ell|\mathbf{y}_1)}}{\frac{1}{m}\sum_{i=1}^m \frac{\exp(k(\mathbf{x}_i, \mathbf{y}_1))p(\mathbf{x}_i)}{q(\mathbf{x}_i|\mathbf{y}_1)}}\right] \tag{218}$$

$$= \mathbb{E}_{p(\mathbf{y}_1)q(\mathbf{x}_{1:m}|\mathbf{y}_1)} \left[\frac{\frac{1}{m}\sum_{\ell=1}^{m} \frac{\exp(k(\mathbf{x}_\ell,\mathbf{y}_1))p(\mathbf{x}_\ell)}{q(\mathbf{x}_\ell|\mathbf{y}_1)}}{\frac{1}{m}\sum_{i=1} \frac{\exp(k(\mathbf{x}_i,\mathbf{y}_1))p(\mathbf{x}_i)}{q(\mathbf{x}_i|\mathbf{y}_1)}} \right] \tag{219}$$

$$= 1. \tag{220}$$

This completes the proof. □

A limitation of this bound is that we need to know the density $p(\mathbf{x})$.

References

David Barber and Felix Agakov. The IM algorithm: a variational approach to information maximization. *Advances in Neural Information Processing Systems*, 16:201–208, 2003.

Joakim Beck, Ben Mansour Dia, Luis FR Espath, Quan Long, and Raul Tempone. Fast bayesian experimental design: Laplace-based importance sampling for the expected information gain. *Computer Methods in Applied Mechanics and Engineering*, 334:523–553, 2018.

Ishmael Belghazi, Sai Rajeswar, Aristide Baratin, R Devon Hjelm, and Aaron Courville. MINE: mutual information neural estimation. *arXiv preprint arXiv:1801.04062*, 2018.

Richard Bellman. A markovian decision process. *Journal of mathematics and mechanics*, 6(5):679–684, 1957.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Nicholas Carlini, Anish Athalye, Nicolas Papernot, Wieland Brendel, Jonas Rauber, Dimitris Tsipras, Ian Goodfellow, Aleksander Madry, and Alexey Kurakin. On evaluating adversarial robustness. *arXiv preprint arXiv:1902.06705*, 2019.

Daniel R Cavagnaro, Jay I Myung, Mark A Pitt, and Janne V Kujala. Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural computation*, 22 (4):887–905, 2010.

Richard Dearden, Nir Friedman, and Stuart Russell. Bayesian q-learning. In *Aaai/iaai*, pages 761–768, 1998.

Monroe D Donsker and SR Srinivasa Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.

Michael O'Gordon Duff. *Optimal Learning: Computational procedures for Bayes-adaptive Markov decision processes*. University of Massachusetts Amherst, 2002.

Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational Bayesian Optimal Experimental Design. In *Advances in Neural Information Processing Systems 32*, pages 14036–14047. Curran Associates, Inc., 2019.

Adam Foster, Martin Jankowiak, Matthew O'Meara, Yee Whye Teh, and Tom Rainforth. A unified stochastic gradient approach to designing bayesian-optimal experiments. In *International Conference on Artificial Intelligence and Statistics*, pages 2959–2969. PMLR, 2020.

Adam Foster, Desi R Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. *arXiv preprint arXiv:2103.02438*, 2021.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pages 1183–1192. PMLR, 2017.

Mohammad Ghavamzadeh, Shie Mannor, Joelle Pineau, and Aviv Tamar. Bayesian reinforcement learning: A survey. *arXiv preprint arXiv:1609.04436*, 2016.

Michael B Giles. Multilevel monte carlo path simulation. *Operations research*, 56(3):607–617, 2008.

Takashi Goda, Tomohiko Hironaka, and Takeru Iwamoto. Multilevel Monte Carlo estimation of expected information gains. *Stochastic Analysis and Applications*, 38(4):581–600, 2020a.

Takashi Goda, Tomohiko Hironaka, and Wataru Kitade. Unbiased mlmc stochastic gradient-based optimization of bayesian experimental designs. *arXiv preprint arXiv:2005.08414*, 2020b.

Arthur Guez, David Silver, and Peter Dayan. Efficient bayes-adaptive reinforcement learning using sample-based search. In *Advances in neural information processing systems*, pages 1025–1033, 2012.

Markus Hainy, David J Price, Olivier Restif, and Christopher Drovandi. Optimal bayesian design for model discrimination via classification. *arXiv preprint arXiv:1809.05301*, 2018.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Taylor A Howell, Chunjiang Fu, and Zachary Manchester. Direct policy optimization using deterministic sampling and collocation. *IEEE Robotics and Automation Letters*, 6(3):5324–5331, 2021.

Xun Huan and Youssef M Marzouk. Sequential bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*, 2016.

Aapo Hyvärinen. Survey on independent component analysis. 1999.

Maximilian Igl, Luisa Zintgraf, Tuan Anh Le, Frank Wood, and Shimon Whiteson. Deep variational reinforcement learning for pomdps. In *International Conference on Machine Learning*, pages 2117–2126. PMLR, 2018.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Vijay R Konda and John N Tsitsiklis. Actor-critic algorithms. In *Advances in neural information processing systems*, pages 1008–1014, 2000.

Nojun Kwak and Chong-Ho Choi. Input feature selection by mutual information based on parzen window. *IEEE transactions on pattern analysis and machine intelligence*, 24(12):1667–1671, 2002.

Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005, 1956.

Ralph Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, 1988.

Guy Lorberbom, Chris J Maddison, Nicolas Heess, Tamir Hazan, and Daniel Tarlow. Direct policy gradients: Direct optimization of policies in discrete action spaces. *arXiv preprint arXiv:1906.06062*, 2019.

Józef Marcinkiewicz and Antoni Zygmund. Quelques théoremes sur les fonctions indépendantes. *Fund. Math*, 29:60–90, 1937.

Ruth K Meyer and Christopher J Nachtsheim. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69, 1995.

XuanLong Nguyen, Martin J Wainwright, and Michael I Jordan. Estimating divergence functionals and the likelihood ratio by convex risk minimization. *IEEE Transactions on Information Theory*, 56(11): 5847–5861, 2010.

Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. f-gan: Training generative neural samplers using variational divergence minimization. *arXiv preprint arXiv:1606.00709*, 2016.

Ben Poole, Sherjil Ozair, Aäron van den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International Conference on Machine Learning*, pages 5171–5180, 2019.

Tom Rainforth, Rob Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting monte carlo estimators. In *International Conference on Machine Learning*, pages 4267–4276. PMLR, 2018.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

Stephane Ross, Brahim Chaib-draa, and Joelle Pineau. Bayes-Adaptive POMDPs. In *NIPS*, pages 1225–1232, 2007.

Kenneth J Ryan. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12(3):585–603, 2003.

Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27 (3):379–423, 1948.

Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Joep Vanlier, Christian A Tiemann, Peter AJ Hilbers, and Natal AW van Riel. Optimal experiment design for model selection in biochemical networks. *BMC systems biology*, 8(1):1–16, 2014.

Benjamin T Vincent and Tom Rainforth. The DARC toolbox: automated, flexible, and efficient delayed and risky choice experiments using bayesian adaptive design. *Retrieved from psyarxiv.com/yehjb*, 2017.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

Christopher K Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

Sue Zheng, Jason Pacheco, and John Fisher. A robust approach to sequential information theoretic planning. In *International Conference on Machine Learning*, pages 5941–5949, 2018.

Luisa Zintgraf, Kyriacos Shiarlis, Maximilian Igl, Sebastian Schulze, Yarin Gal, Katja Hofmann, and Shimon Whiteson. Varibad: A very good method for bayes-adaptive deep rl via meta-learning. *arXiv preprint arXiv:1910.08348*, 2019.