Technical Note—Knowledge Gradient for Selection with Covariates: Consistency and Computation

Liang Ding*   L. Jeff Hong[†]   Haihui Shen[‡⊠]   Xiaowei Zhang[§]

Knowledge gradient is a design principle for developing Bayesian sequential sampling policies to solve optimization problems. In this paper we consider the ranking and selection problem in the presence of covariates, where the best alternative is not universal but depends on the covariates. In this context, we prove that under minimal assumptions, the sampling policy based on knowledge gradient is consistent, in the sense that following the policy the best alternative as a function of the covariates will be identified almost surely as the number of samples grows. We also propose a stochastic gradient ascent algorithm for computing the sampling policy and demonstrate its performance via numerical experiments.

KEYWORDS: selection of the best; covariates; knowledge gradient; consistency

1. Introduction

We consider the ranking and selection (R&S) problem in the presence of covariates. A decision maker is presented with a finite collection of alternatives. The performance of each alternative is unknown and depends on the covariates. Suppose that the decision maker has access to noisy samples of each alternative for any chosen value of the covariates, but the samples are expensive to acquire. Given a finite sampling budget, the goal is to develop an efficient sampling policy indicating locations as to which alternative and what value of the covariates to sample from, so that upon termination of the sampling, the decision maker can identify a decision rule that accurately specifies the best alternative as a function of the covariates.

The problem of R&S with covariates emerges naturally as the popularization of data and decision analytics in recent years. In clinical and medical research, for many diseases the effect of a treatment may be

---
*Department of Industrial Systems Engineering, Texas A&M University, College Station, TX, USA
[†]School of Management and School of Data Science, Fudan University, Shanghai, China
[‡]Sino-US Global Logistics Institute, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
[§]Faculty of Business and Economics, University of Hong Kong, Pok Fu Lam, Hong Kong SAR
[⊠]Corresponding Author, shenhaihui@sjtu.edu.cn

substantially different across patients, depending on their biometric characteristics (i.e., the covariates), including age, weight, lifestyle habits such as smoking and alcohol use, etc. (Kim et al. 2011). A treatment regime that works for a majority of patients might not work for the others. Samples needed for estimating treatment effects may be collected from clinical trials or computer simulation. For example, in Hur et al. (2004) and Choi et al. (2014), a simulation model is developed to simulate the effect of several treatment regimens for Barrett's esophagus, a precursor to esophageal cancer, for patients with different biometric characteristics. Personalized medicine can then be developed to determine the best treatment regime that is customized to the particular characteristics of each individual patient. Similar customized decision-making can be found in online advertising (Arora et al. 2008), where advertisements are displayed depending on consumers' web browsing history or buying behavior to increase the revenue of the advertising platform as well as to improve consumers' shopping experience.

Being a classic problem in the area of stochastic simulation, R&S has a vast literature. We refer to Kim and Nelson (2006) and Chen et al. (2015) for reviews on the subject with emphasis on frequentist and Bayesian approaches, respectively. Most of the prior work, however, does not consider the presence of the covariates, and thus the best alternative to select is universal rather than varies as a function of the covariates. There are several exceptions, including Hu and Ludkovski (2017), Pearce and Branke (2017), and Shen et al. (2021). Among them Shen et al. (2021) take a frequentist approach to solve R&S with covariates, whereas the other two a Bayesian approach. The present paper adopts a Bayesian perspective as well.

This paper considers a sampling policy based on knowledge gradient (KG) for R&S with covariates. KG, introduced in Frazier et al. (2008), is a design principle that has been widely used for developing Bayesian sequential sampling policies to solve a variety of optimization problems, including R&S, in which evaluation of the objective function is noisy and expensive. In its basic form, KG begins with assigning a multivariate normal prior on the unknown constant performance of all alternatives. In each iteration, it chooses the sampling location by maximizing the increment in the expected value of the information that would be gained by taking a sample from the location. Then, the posterior is updated upon observing the noisy sample from the chosen location. The sampling efficiency of KG-type policies is often competitive with or outperforms other sampling policies; see Frazier et al. (2009), Scott et al. (2011), Ryzhov (2016), and Pearce and Branke (2018) among others.

A KG-based sampling policy for R&S with covariates is also proposed in Pearce and Branke (2017). The main difference here is that our treatment is more general. First, we allow the sampling noise to be heteroscedastic, whereas it is assumed to be constant for different locations of the same alternative in their work. Heteroscedasticity is of particular significance for simulation applications such as queueing systems. Second, we take into account possible variations in sampling cost at different locations, whereas the sampling cost is simply treated as constant everywhere in Pearce and Branke (2017). Hence, our policy, which we refer to as integrated knowledge gradient (IKG), attempts in each iteration to maximize a "cost-adjusted" increment in the expected value of information. These generalizations are straightforward when the variance of the sampling noise and the sampling cost are assumed to be known. We also briefly discuss and show how to deal with the case where they are unknown.

The first main contribution of this paper is to provide a theoretical analysis of the asymptotic behavior

of the IKG policy, whereas Pearce and Branke (2017) conducted only numerical investigation. In particular, we prove that IKG is consistent in the sense that for any value of the covariates, the selected alternative upon termination of the sampling will converge to the true best almost surely as the sampling budget grows to infinity. Moreover, we consider a practical variant—termed quasi-IKG—which does not require the intermediate optimization problem in each iteration of IKG to be solved exactly, and prove its consistency under mild conditions.

Consistency of KG-type policies has been established in various settings, mostly for problems where the number of feasible solutions is finite, including R&S (Frazier et al. 2008, 2009, Frazier and Powell 2011, Mes et al. 2011), and discrete optimization via simulation (Xie et al. 2016). KG is also used for Bayesian optimization of continuous functions in Wu and Frazier (2016), Poloczek et al. (2017), and Wu et al. (2017). However, in these papers the continuous domain is discretized first, which effectively reduces the problem to one with finite feasible solutions, in order to facilitate their asymptotic analysis. The finiteness of the domain is critical in the aforementioned papers, because the asymptotic analysis there boils down to proving that each feasible solution can be sampled infinitely often. This, by the law of large numbers, implies that the variance of the objective value estimate of each solution will converge to zero. Thus, the optimal solution will be identified ultimately since the uncertainty about the performances of the solutions will be removed completely in the end.

By contrast, proving consistency of KG-type policies for continuous solution domains demands a fundamentally different approach, since most solutions in a continuous domain would hardly be sampled even once after all. Among the several related papers, Scott et al. (2011) studies a KG-type policy for Bayesian optimization of continuous functions. Assigning a Gaussian process prior on the objective function, they established the consistency of the KG-type policy basically by leveraging the continuity of the covariance function of the Gaussian process, which intuitively suggests that if the variance at one location is small, then the variance in its neighborhood ought to be small too. Toscano-Palmerin and Frazier (2018) prove the consistency of a KG-type policy on a more general problem that can reduce to the problem in Scott et al. (2011), for both discrete and continuous domains.

We cast R&S with covariates to a problem of ranking a finite number of Gaussian processes, thereby having both discrete and continuous elements structurally. As a result, we establish the consistency of the proposed IKG policy by proving the following two facts – (i) each Gaussian process is sampled infinitely often, and (ii) the infinitely many samples assigned to a given Gaussian process drives its posterior variance at any location to zero, thanks to the assumed continuity of its covariance function. The theoretical analysis in this paper is partly built on the ideas developed for discrete and continuous problems, respectively, in Frazier et al. (2008) and Scott et al. (2011) in a federated manner.

Although our proofs share similar structures to those in Scott et al. (2011), our assumptions are substantially simpler and minimal. By contrast, for the proof in Scott et al. (2011) to be valid, technical conditions are imposed to regulate the asymptotic behavior of the posterior mean function and the posterior covariance function of the underlying Gaussian process. Nevertheless, the two conditions are difficult to verify. We do not impose such conditions. We achieve the substantial simplification of the assumptions by leveraging the reproducing kernel Hilbert space (RKHS) theory. The theory has been used widely in

machine learning (Steinwart and Christmann 2008). But its use in the analysis of KG-type policies is less common. We develop several technical results based on RKHS theory to facilitate analysis of the asymptotic behavior of the posterior covariance function.[1]

The second main contribution of this paper is that we develop an algorithm to solve a stochastic optimization problem that determines the sampling decision of the IKG policy in its each iteration. In Pearce and Branke (2017), this optimization problem is addressed by the sample average approximation method with a derivative-free optimization solver. Instead, we propose a stochastic gradient ascent (SGA) algorithm, taking advantage of the fact that a gradient estimator can be derived analytically for many popular covariance functions. Numerical experiments demonstrate the finite-sample performance of the IKG policy in conjunction with the SGA algorithm.

We conclude the introduction by reviewing briefly the most pertinent literature. A closely related problem is multi-armed bandit (MAB); see Bubeck and Cesa-Bianchi (2012) for a comprehensive review on the subject. The significance of covariates, thereby contextual MAB (or MAB with covariates), has also drawn substantial attention in recent years; see Rusmevichientong and Tsitsiklis (2010), Yang and Zhu (2002), Krause and Ong (2011), and Perchet and Rigollet (2013) among others. There are two critical differences between contextual MAB and R&S with covariates. First, the former generally assumes that the covariates arrive exogenously in a sequential manner, and the decision-maker can choose at which arm (or alternative) to sample but not the value of covariates. By contrast, the latter assumes that the decision-maker is capable of choosing both the alternative and the covariates when specifying sampling locations. A second difference is MAB focuses on minimizing the regret which is caused by choosing inferior alternatives and accumulated during the sampling process, whereas R&S focuses on identifying the best alternative eventually and the regret is not the primary concern.

The rest of the paper is organized as follows. In Section 2 we follow a nonparametric Bayesian approach to formulate the problem of R&S with covariates, introduce the IKG policy, and present the main result. In Section 3 we prove the consistency of our sampling policy in the sense that the estimated best alternative as a function of the covariates converges to the truth with probability one as the number of samples grows to infinity. We then propose to use SGA for computing our sampling policy in Section 4, and demonstrate its performance via numerical experiments in Section 5. We conclude in Section 6 and collect detailed proof and additional technical results and numerical experiments in the Appendix.

2. Problem Formulation

Suppose that a decision maker is presented with M competing alternatives. For each $i = 1, \ldots, M$, the performance of alternative i depends on a vector of *covariates* $\boldsymbol{x} = (x_1, \ldots, x_d)^\intercal$ and is denoted by $\theta_i = \theta_i(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$. The performances are unknown and can only be learned via sampling. In particular, for any i and $\boldsymbol{x}$, one can acquire possibly multiple noisy samples of $\theta_i(\boldsymbol{x})$. The decision maker aims to select the "best" alternative for a given value of $\boldsymbol{x}$, i.e., identify $\mathrm{argmax}_i \, \theta_i(\boldsymbol{x})$. However,

---

[1] Bect et al. (2019) adopt a supermartingale approach to study the asymptotic behavior of a general class of sequential sampling algorithms. Their analysis has a broader scope of applicability but it is technically more involved.

since the sampling is usually expensive in time and/or money, instead of estimating the performances $\{\theta_i(\boldsymbol{x}) : i = 1, \ldots, M\}$ every time a new value of $\boldsymbol{x}$ is observed and then ranking them, it is preferable to learn *offline* the decision rule

$$i^*(\boldsymbol{x}) \in \underset{1 \leq i \leq M}{\operatorname{argmax}} \theta_i(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{X}, \tag{1}$$

as a function of $\boldsymbol{x}$, through a carefully designed sampling process. Equipped with such a decision rule, the decision maker can select the best alternative upon observing the covariates in a timely fashion. In addition, the decision maker may have some knowledge with regard to the covariates. For example, certain values of the covariates may be more important or appear more frequently than others. Suppose that this kind of knowledge is expressed by a probability density function $\gamma(\boldsymbol{x})$ on $\mathcal{X}$.

During the offline learning period, we need to make a sequence of sampling decisions $\{(a^n, \boldsymbol{v}^n) : n = 0, 1, \ldots\}$, where $(a^n, \boldsymbol{v}^n)$ means that the $(n+1)$-th sample, denoted by y^{n+1}, is taken from alternative a^n with covariates value $\boldsymbol{v}^n$ (refer it as location $\boldsymbol{v}^n$ for simplicity). We assume that given $\theta_{a^n}(\boldsymbol{v}^n)$, y^{n+1} is an unbiased sample having a normal distribution, i.e.,

$$y^{n+1} \mid \theta_{a^n}(\boldsymbol{v}^n) \sim \mathcal{N}(\theta_{a^n}(\boldsymbol{v}^n), \lambda_{a^n}(\boldsymbol{v}^n)),$$

where $y^n \mid \theta_i(\boldsymbol{x})$ is independent of $y^{n'} \mid \theta_{i'}(\boldsymbol{x}')$ for $(i, \boldsymbol{x}, n) \neq (i', \boldsymbol{x}', n')$. Here, $\lambda_i(\boldsymbol{x})$ is the variance of a sample of $\theta_i(\boldsymbol{x})$ given $\theta_i(\boldsymbol{x})$ and is assumed to be *known*. Moreover, suppose that the cost of taking a sample from alternative i at location $\boldsymbol{x}$ is $c_i(\boldsymbol{x}) > 0$, which is also assumed to be *known*. Suppose that the total sampling budget for offline learning is $B > 0$, and the sampling process is terminated when the budget is exhausted. Mathematically, we will stop with the $N(B)$-th sample, where

$$N(B) := \min \left\{ N : \sum_{n=0}^{N} c_{a^n}(\boldsymbol{v}^n) > B \right\}. \tag{2}$$

Consequently, the sampling decisions are $\{(a^n, \boldsymbol{v}^n) : n = 0, \ldots, N(B) - 1\}$ and the samples taken during the process are $\{y^{n+1} : n = 0, \ldots, N(B) - 1\}$. Notice that $N(B) = B$ if $c_i(\boldsymbol{x}) \equiv 1$ for $i = 1, \ldots, M$, in which case the sampling budget is reduced to the number of samples.

Remark 1. The assumption of known $\lambda_i(\boldsymbol{x})$ is critical to the theoretical analysis in this paper. As we will see shortly, with known $\lambda_i(\boldsymbol{x})$, if we impose a Gaussian process as prior for θ_i, then its posterior will still be a Gaussian process, which makes the asymptotic analysis tractable. It would not be the case if $\lambda_i(\boldsymbol{x})$ also needs to be estimated. In practice, $\lambda_i(\boldsymbol{x})$ is usually unknown and it is a common issue in the experiment design. We suggest to follow the approach in Ankenman et al. (2010), which fits the surfaces of $\lambda_i(\boldsymbol{x})$ by running multiple simulations at certain design points and using the sample variances. See more details in the numerical experiments in the Appendix. The unknown sampling cost $c_i(\boldsymbol{x})$ in practice can be handled similarly.

We follow a nonparametric Bayesian approach to model the unknown functions $\{\theta_1, \ldots, \theta_M\}$ as well as to design the sampling policy. We treat θ_i's as random functions and impose a prior on them under which they are mutually independent, although this assumption may be relaxed. Suppose that $\boldsymbol{x}$ takes

continuous values and that under the prior, θ_i is a Gaussian process with mean function $\mu_i^0(\boldsymbol{x}) := \mathbb{E}[\theta_i(\boldsymbol{x})]$ and covariance function $k_i^0(\boldsymbol{x}, \boldsymbol{x}') := \mathrm{Cov}[\theta_i(\boldsymbol{x}), \theta_i(\boldsymbol{x}')]$ that satisfies the following assumption.

Assumption 1. For each $i = 1, \ldots, M$, there exists a constant $\tau_i > 0$ and a positive continuous function $\rho_i : \mathbb{R}^d \mapsto \mathbb{R}_+$ such that $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = \tau_i^2 \rho_i(\boldsymbol{x} - \boldsymbol{x}')$. Moreover,

(i) $\rho_i(|\boldsymbol{\delta}|) = \rho_i(\boldsymbol{\delta})$, where $|\cdot|$ means taking the absolute value component-wise;

(ii) $\rho_i(\boldsymbol{\delta})$ is decreasing in $\boldsymbol{\delta}$ component-wise for $\boldsymbol{\delta} \geq \boldsymbol{0}$;

(iii) $\rho_i(\boldsymbol{0}) = 1$, $\rho_i(\boldsymbol{\delta}) \to 0$ as $\|\boldsymbol{\delta}\| \to \infty$, where $\|\cdot\|$ denotes the Euclidean norm;

(iv) there exist some $0 < C_i < \infty$ and $\varepsilon_i, u_i > 0$ such that

$$1 - \rho_i(\boldsymbol{\delta}) \leq \frac{C_i}{|\log(\|\boldsymbol{\delta}\|)|^{1+\varepsilon_i}},$$

for all $\boldsymbol{\delta}$ such that $\|\boldsymbol{\delta}\| < u_i$.

Remark 2. Assumption 1 stipulates that k_i^0 is *stationary*, i.e., it depends on $\boldsymbol{x}$ and $\boldsymbol{x}'$ only through the difference $\boldsymbol{x} - \boldsymbol{x}'$. In addition, τ_i^2 can be interpreted as the prior variance of $\theta_i(\boldsymbol{x})$ for all $\boldsymbol{x}$, and $\rho_i(\boldsymbol{x} - \boldsymbol{x}')$ as the prior correlation between $\theta_i(\boldsymbol{x})$ and $\theta_i(\boldsymbol{x}')$ which increases to 1 as $\|\boldsymbol{x} - \boldsymbol{x}'\|$ decreases to 0. The condition in part (iv) of Assumption 1 is weak. In conjunction with the continuity assumption of ρ_i, it implies that the sample paths of the Gaussian process θ_i are continuous almost surely if the mean function $\mu_i^0(\boldsymbol{x})$ is continuous; see, e.g., Adler and Taylor (2007, Theorem 1.4.1). The sample path continuity will be used to establish the uniform convergence of the posterior mean functions.

A variety of covariance functions satisfy Assumption 1. Notable examples include the squared exponential (SE) covariance function

$$k_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \tau^2 \exp\left(-r^2(\boldsymbol{x} - \boldsymbol{x}')\right),$$

where $r(\boldsymbol{\delta}) = \sqrt{\sum_{j=1}^d \alpha_j \delta_j^2}$ and α_j's are positive parameters, and the Matérn covariance function

$$k_{\mathrm{Matérn}}(\boldsymbol{x}, \boldsymbol{x}') = \tau^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} r(\boldsymbol{x} - \boldsymbol{x}')\right)^\nu K_\nu\left(\sqrt{2\nu} r(\boldsymbol{x} - \boldsymbol{x}')\right),$$

where ν is a positive parameter that is typically taken as half-integer (i.e., $\nu = p + 1/2$ for some nonnegative integer p), Γ is the gamma function, and K_ν is the modified Bessel function of the second kind. The covariance function reflects one's prior belief about the unknown functions. We refer to Rasmussen and Williams (2006, Chapter 4) for more types of covariance functions.

2.1. Bayesian Updating Equations

For each $n = 1, 2, \ldots$, let $\mathcal{F}^n$ denote the σ-algebra generated by $(a^0, \boldsymbol{v}^0), y^1, \ldots, (a^{n-1}, \boldsymbol{v}^{n-1}), y^n$, the sampling decisions and the samples collected up to time n. Suppose that $(a^n, \boldsymbol{v}^n) \in \mathcal{F}^n$, that is, $(a^n, \boldsymbol{v}^n)$

depends only on the information available at time n. In addition, we use the notation $\mathbb{E}^n[\cdot] := \mathbb{E}[\cdot|\mathcal{F}^n]$, and define $\mathrm{Var}^n[\cdot]$ and $\mathrm{Cov}^n[\cdot]$ likewise.

Given the setup of our model, it is easy to derive that $\{\theta_1, \ldots, \theta_M\}$ are independent Gaussian processes under the posterior distribution conditioned on $\mathcal{F}^n$, $n = 1, \ldots, N(B)$. In particular, under the prior mutual independence, taking samples from one unknown function does not provide information on another. Let $\boldsymbol{V}_i^n := \{\boldsymbol{v}^\ell : a^\ell = i, \ell = 0, \ldots, n-1\}$ denote the set of the locations of the samples taken from θ_i up to time n and define $\boldsymbol{y}_i^n := \{y^{\ell+1} : a^\ell = i, \ell = 0, \ldots, n-1\}$ likewise. With slight abuse of notation, when necessary, we will also treat $\boldsymbol{V}_i^n$ as a matrix wherein the columns are corresponding to the points in the set and arranged in the order of appearance, and $\boldsymbol{y}_i^n$ as a column vector with elements also arranged in the order of appearance. Then, the posterior mean and covariance functions of θ_i are given by

$$\mu_i^n(\boldsymbol{x}) := \mathbb{E}^n[\theta_i(\boldsymbol{x})] = \mu_i^0(\boldsymbol{x}) + k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i(\boldsymbol{V}_i^n)]^{-1}[\boldsymbol{y}_i^n - \mu_i^0(\boldsymbol{V}_i^n)], \tag{3}$$

$$k_i^n(\boldsymbol{x}, \boldsymbol{x}') := \mathrm{Cov}^n[\theta_i(\boldsymbol{x}), \theta_i(\boldsymbol{x}')] = k_i^0(\boldsymbol{x}, \boldsymbol{x}') - k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i(\boldsymbol{V}_i^n)]^{-1}k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}'), \tag{4}$$

where for two sets $\boldsymbol{V}$ and $\boldsymbol{V}'$, $k_i^0(\boldsymbol{V}, \boldsymbol{V}') = [k_i^0(\boldsymbol{x}, \boldsymbol{x}')]_{\boldsymbol{x} \in \boldsymbol{V}, \boldsymbol{x}' \in \boldsymbol{V}'}$ is a matrix of size $|\boldsymbol{V}| \times |\boldsymbol{V}'|$, $\lambda_i(\boldsymbol{V}) = \mathrm{diag}\{\lambda_i(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{V}\}$ is a diagonal matrix of size $|\boldsymbol{V}| \times |\boldsymbol{V}|$, and $\mu_i^0(\boldsymbol{V}) = (\mu_i^0(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{V})$ is a column vector of size $|\boldsymbol{V}| \times 1$. Here $|\cdot|$ denotes the cardinality of a set. We refer to, for example, Scott et al. (2011, Section 3.2) for details. Further, the following updating equation can be derived

$$\mu_i^{n+1}(\boldsymbol{x}) = \mu_i^n(\boldsymbol{x}) + \sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n)Z^{n+1}, \tag{5}$$

$$k_i^{n+1}(\boldsymbol{x}, \boldsymbol{x}') = k_i^n(\boldsymbol{x}, \boldsymbol{x}') - \sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n)\sigma_i^n(\boldsymbol{x}', \boldsymbol{v}^n), \tag{6}$$

where Z^{n+1} is a standard normal random variable independent to everything else, and

$$\sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n) := \begin{cases} \tilde{\sigma}_i^n(\boldsymbol{x}, \boldsymbol{v}^n), & \text{if } i = a^n, \\ 0, & \text{if } i \neq a^n, \end{cases} \quad \text{and} \quad \tilde{\sigma}_i^n(\boldsymbol{x}, \boldsymbol{v}) := \frac{k_i^n(\boldsymbol{x}, \boldsymbol{v})}{\sqrt{k_i^n(\boldsymbol{v}, \boldsymbol{v}) + \lambda_i(\boldsymbol{v})}}. \tag{7}$$

In particular, conditioned on $\mathcal{F}^n$ and prior to taking a sample at $(a^n, \boldsymbol{v}^n)$, the predictive distribution of $\mu_i^{n+1}(\boldsymbol{x})$ is normal with mean $\mu_i^n(\boldsymbol{x})$ and standard deviation $\sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n)$. Moreover, notice that

$$\mathrm{Var}^{n+1}[\theta_i(\boldsymbol{x})] = k_i^{n+1}(\boldsymbol{x}, \boldsymbol{x}) = k_i^n(\boldsymbol{x}, \boldsymbol{x}) - [\sigma_i^n(\boldsymbol{x}, \boldsymbol{v}^n)]^2 \leq \mathrm{Var}^n[\theta_i(\boldsymbol{x})]. \tag{8}$$

(Note that eqs. (5)–(8) are still valid even if $k_i^n(\boldsymbol{v}^n, \boldsymbol{v}^n) = 0$, and/or $\lambda_i(\boldsymbol{v}^n) = 0$.) Hence, $\mathrm{Var}^n[\theta_i(\boldsymbol{x})]$ is non-increasing in n. This basically suggests that the uncertainty about each unknown function under the posterior decreases as more samples from it are collected. It is thus both desirable and practically meaningful that such uncertainty would be completely eliminated if the sampling budget is unlimited, in which case one would be able to identify the decision rule eq. (1) perfectly. In particular, we define consistency of a sampling policy as follows.

Definition 1. A sampling policy is said to be *consistent* if it ensures that

$$\lim_{B \to \infty} \operatorname*{argmax}_{1 \le i \le M} \mu_i^{N(B)}(\boldsymbol{x}) = \operatorname*{argmax}_{1 \le i \le M} \theta_i(\boldsymbol{x}), \tag{9}$$

almost surely (a.s.) for all $\boldsymbol{x} \in \mathcal{X}$.

Remark 3. Under the assumption that $\{\theta_1, \ldots, \theta_M\}$ are independent under the prior, collecting samples from θ_i does not provide information about θ_j if $i \ne j$. Therefore, a consistent policy under the independence assumption ought to ensure that the number of samples taken from *each* θ_i grows without bounds.

2.2. Knowledge Gradient Policy

We first assume temporarily that $\boldsymbol{x}$ is given and fixed, and that $c_i(\boldsymbol{x}) = 1$ for $i = 1, \ldots, M$. Then, solving $\max_i \theta_i(\boldsymbol{x})$ is a selection of the best problem having finite alternatives, and each sampling decision is reduced to choosing an alternative i to take a sample of $\theta_i(\boldsymbol{x})$. The knowledge gradient (KG) policy introduced in Frazier et al. (2008) is designed exactly to solve such a problem assuming an independent normal prior. Specifically, the knowledge gradient at i is defined there as the increment in the expected value of the information about the maximum at $\boldsymbol{x}$ gained by taking a sample at i, that is,

$$\mathrm{KG}^n(i; \boldsymbol{x}) := \mathbb{E}\left[\max_{1 \le a \le M} \mu_a^{n+1}(\boldsymbol{x}) \,\Big|\, \mathcal{F}^n, a^n = i\right] - \max_{1 \le a \le M} \mu_a^n(\boldsymbol{x}). \tag{10}$$

Then, each time the alternative i that has the largest value of $\mathrm{KG}(i; \boldsymbol{x})$ is selected to generate a sample of $\theta_i(\boldsymbol{x})$.

Let us now return to our context where (1) the covariates are present, (2) each sampling decision consists of both i and $\boldsymbol{x}$, and (3) each sampling decision may induce a different sampling cost. Since a sample of $\theta_i(\boldsymbol{x})$ would alter the posterior belief about $\theta_i(\boldsymbol{x}')$, we generalize eq. (10) and define

$$\mathrm{KG}^n(i, \boldsymbol{x}; \boldsymbol{v}) := \frac{1}{c_i(\boldsymbol{x})}\left\{\mathbb{E}\left[\max_{1 \le a \le M} \mu_a^{n+1}(\boldsymbol{v}) \,\Big|\, \mathcal{F}^n, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] - \max_{1 \le a \le M} \mu_a^n(\boldsymbol{v})\right\}, \tag{11}$$

which can be interpreted as the increment in the expected value of the information about the maximum at $\boldsymbol{v}$ gained per unit of sampling cost by taking a sample at $(i, \boldsymbol{x})$. Then, we consider the following *integrated KG* (IKG)

$$\mathrm{IKG}^n(i, \boldsymbol{x}) := \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \left\{\mathbb{E}\left[\max_{1 \le a \le M} \mu_a^{n+1}(\boldsymbol{v}) \,\Big|\, \mathcal{F}^n, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] - \max_{1 \le a \le M} \mu_a^n(\boldsymbol{v})\right\} \gamma(\boldsymbol{v}) \,\mathrm{d}\boldsymbol{v}, \tag{12}$$

and define the IKG sampling policy as

$$(a^n, \boldsymbol{v}^n) \in \operatorname*{argmax}_{1 \le i \le M, \boldsymbol{x} \in \mathcal{X}} \mathrm{IKG}^n(i, \boldsymbol{x}). \tag{13}$$

The integrand of eq. (12) can be calculated analytically, as shown in Lemma 1, whose proof is deferred to the Appendix.

Lemma 1. *For all $i = 1, \ldots, M$ and $\boldsymbol{x} \in \mathcal{X}$,*

$$\mathrm{IKG}^n(i, \boldsymbol{x}) = \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \left[|\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| \phi\left(\left| \frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})} \right| \right) - |\Delta_i^n(\boldsymbol{v})| \Phi\left(-\left| \frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})} \right| \right) \right] \gamma(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}, \qquad (14)$$

where $\Delta_i^n(\boldsymbol{v}) := \mu_i^n(\boldsymbol{v}) - \max_{a \neq i} \mu_a^n(\boldsymbol{v})$, Φ is the standard normal distribution function, and ϕ is its density function.

We solve eq. (13) by first solving $\max_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$ for all i and then enumerating the results. The computational challenge in the former lies in the numerical integration in eq. (14). Notice that $\max_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$ is in fact a stochastic optimization problem if we view the integration in eq. (14) as an expectation with respect to the probability density $\gamma(\boldsymbol{x})$ on $\mathcal{X}$. One might apply the sample average approximation method to solve $\max_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$, but it would be computationally prohibitive if $\mathcal{X}$ is high-dimensional. Instead, we show in Section 4 that the gradient of the integrand in eq. (14) with respect to $\boldsymbol{x}$ can be calculated explicitly, which is an unbiased estimator of $\nabla_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$ under regularity conditions, thereby leading to a stochastic gradient ascent method (Kushner and Yin 2003).

We now present our main theoretical result — the IKG policy is consistent under simple assumptions. The proof will be sketched in Section 3 and all details are collected in the Appendix.

Assumption 2. The design space $\mathcal{X}$ is a compact set in $\mathbb{R}^d$ with nonempty interior.

Assumption 3. For each $i = 1, \ldots, M$, $\mu_i^0(\cdot)$, $\lambda_i(\cdot) > 0$ and $c_i(\cdot) > 0$ are all continuous on $\mathcal{X}$, and $\gamma(\cdot) > 0$ on $\mathcal{X}$.

Under Assumptions 1–3, the IKG policy (13) is well defined. This can be seen by noting that the maximum of $\mathrm{IKG}^n(i, \boldsymbol{x})$ over $\boldsymbol{x} \in \mathcal{X}$ is attainable since $\mathrm{IKG}^n(i, \boldsymbol{x})$ is continuous in $\boldsymbol{x}$ by Assumptions 1 and 3 together with Lemma 1, and $\mathcal{X}$ is compact by Assumption 2. Moreover, the IKG policy (13) is consistent as formally stated in the following Theorem 1.

Theorem 1. *If Assumptions 1-3 hold, then the IKG policy (13) is consistent, that is, under the IKG policy,*

(i) $k_i^{N(B)}(\boldsymbol{x}, \boldsymbol{x}) \to 0$ *a.s. as $B \to \infty$ for all $\boldsymbol{x} \in \mathcal{X}$ and $i = 1, \ldots, M$;*

(ii) $\mu_i^{N(B)}(\boldsymbol{x}) \to \theta_i(\boldsymbol{x})$ *a.s. as $B \to \infty$ for all $\boldsymbol{x} \in \mathcal{X}$ and $i = 1, \ldots, M$;*

(iii) $\mathrm{argmax}_{1 \leq i \leq M} \mu_i^{N(B)}(\boldsymbol{x}) \to \mathrm{argmax}_{1 \leq i \leq M} \theta_i(\boldsymbol{x})$ *a.s. as $B \to \infty$ for all $\boldsymbol{x} \in \mathcal{X}$.*

We conclude this section by highlighting the differences between our assumptions and those in Scott et al. (2011), in which the consistency of a KG-type policy driven by a Gaussian process is proved. First and foremost, *they impose conditions on both the posterior mean function and the posterior covariance function to regulate their large-sample asymptotic behavior.* Specifically, they assume that uniformly for all n and $\boldsymbol{x}, \boldsymbol{v} \in \mathcal{X}$ with $\boldsymbol{x} \neq \boldsymbol{v}$, (1) $|\mu^n(\boldsymbol{x}) - \mu^n(\boldsymbol{v})|$ is bounded a.s., and (2) $|\mathrm{Corr}^n[\theta(\boldsymbol{x}), \theta(\boldsymbol{v})]|$ is bounded above away from one, where Corr^n means the posterior correlation.[2] *The two assumptions are critical for their analysis but nontrivial to verify in practice.*

---

[2] The subscript i is ignored because there is only one Gaussian process involved in Scott et al. (2011).

By contrast, we do not make such assumptions. Condition (1) is not necessary in our analysis because the "increment in the expected value of the information" is defined as eq. (12) in this paper, whereas in a different form without integration in Scott et al. (2011). There is no need for us to impose Condition (2) in order to regulate the asymptotic behavior of the posterior covariance function, because instead we achieve the same goal by utilizing reproducing kernel Hilbert space (RKHS) theory.

Second, in Scott et al. (2011) the prior covariance function of the underlying Gaussian process is of SE type. We relax it to Assumption 1, which allows a great variety of covariance functions. We also take into account possibly varying sampling costs at different locations.

3. Consistency

It is straightforward to show that $N(B) \to \infty$ if and only if $B \to \infty$, since $c_i(\cdot)$ is bounded both above and below away from zero on $\mathcal{X}$ for each $i = 1, \ldots, M$ under Assumptions 2 and 3. Thus, Theorem 1 is equivalent to Theorem 2 as follows.

Theorem 2. *If Assumptions 1–3 hold, then under the IKG policy,*

(i) $k_i^n(\boldsymbol{x}, \boldsymbol{x}) \to 0$ *a.s. as* $n \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$ *and* $i = 1, \ldots, M$;

(ii) $\mu_i^n(\boldsymbol{x}) \to \theta_i(\boldsymbol{x})$ *a.s. as* $n \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$ *and* $i = 1, \ldots, M$;

(iii) $\operatorname{argmax}_{1 \le i \le M} \mu_i^n(\boldsymbol{x}) \to \operatorname{argmax}_{1 \le i \le M} \theta_i(\boldsymbol{x})$ *a.s. as* $n \to \infty$ *for all* $\boldsymbol{x} \in \mathcal{X}$.

The bulk of the proof of consistency of the IKG policy lies in part (i) of Theorem 2, i.e., to show that $\lim_{n \to \infty} \operatorname{Var}^n[\theta_i(\boldsymbol{x})] = 0$ a.s. for all $\boldsymbol{x} \in \mathcal{X}$ and $i = 1, \ldots, M$. It consists of two steps, which are summarized into the later Propositions 2 and 3. However, both Propositions 2 and 3 critically relies on the asymptotic behavior of the posterior covariance function, which is characterized in the following Proposition 1.

Proposition 1. *Fix* $i = 1, \ldots, M$. *If* k_i^0 *is stationary, then for any* $\boldsymbol{x} \in \mathcal{X}$, $k_i^n(\boldsymbol{x}, \boldsymbol{x}')$ *converges to a limit, denoted by* $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}')$, *uniformly in* $\boldsymbol{x}' \in \mathcal{X}$ *as* $n \to \infty$.

Proposition 1 shows that *irrespective* of the allocation of the design points $\{\boldsymbol{v}^\ell : \ell = 0, \ldots, n-1\}$, $k_i^n(\boldsymbol{x}, \cdot)$ converges *uniformly* as $n \to \infty$ for all $\boldsymbol{x} \in \mathcal{X}$. (Note that this does not mean the limit is necessarily zero.) Not only is this result of interest in its own right, but also is crucial for proving the consistency of IKG policy under assumptions weaker than those imposed for previous related problems (Scott et al. 2011). For example, the uniform convergence preserves the continuity of $k_i^n(\boldsymbol{x}, \cdot)$ in the limit, a property that is crucial for the proof of Proposition 2. A more general version of Proposition 1 is given in Bect et al. (2019, Proposition 2.9), but we present a different proof built on RKHS theory in the Appendix. Proposition 1 sets a foundation for analysing the asymptotic behavior of Bayesian sequential sampling policies based on Gaussian processes with minimal assumptions.

Before we formally state Propositions 2 and 3, the following definitions are required. For each i, let η_i^n denote the (random) number of times that a sample is taken from alternative i regardless of the value of $\boldsymbol{x}$

up to the n-th sample, i.e.,

$$\eta_i^n := \sum_{\ell=0}^{n-1} \mathbb{I}_{\{a^\ell = i\}} .$$

Further, let $\eta_i^\infty := \lim_{n \to \infty} \eta_i^n$, which is well defined since it is a limit of a non-decreasing sequence of random variables.

Proposition 2. *Fix $i = 1, \ldots, M$. If Assumptions 1-3 hold and $\eta_i^\infty = \infty$ a.s., then for any $\boldsymbol{x} \in \mathcal{X}$, $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}) = 0$ a.s. under the IKG policy.*

Proposition 3. *If Assumptions 1-3 hold, then $\eta_i^\infty = \infty$ a.s. for each $i = 1, \ldots, M$ under the IKG policy.*

Part (i) of Theorem 2 is an immediate consequence of Propositions 2 and 3. The proofs of parts (ii) and (iii) of Theorem 2 and Propositions 2 and 3 are all collected in the Appendix.

In practice, the IKG policy (13) can only be solved numerically, as discussed in the next section, in which case the obtained solution $(\tilde{a}^n, \tilde{\boldsymbol{v}}^n)$ is not exactly equal to the true solution $(a^n, \boldsymbol{v}^n)$. Inspired by Bect et al. (2019), we consider the *quasi-IKG sampling policy*, which chooses the sampling decision $(\tilde{a}^n, \tilde{\boldsymbol{v}}^n)$ such that

$$\text{IKG}^n(\tilde{a}^n, \tilde{\boldsymbol{v}}^n) \geq \text{IKG}^n(a^n, \boldsymbol{v}^n) - \varepsilon_n, \tag{15}$$

where $\{\varepsilon_n\}$ is a sequence of non-negative real numbers such that $\varepsilon_n \to 0$ as $n \to \infty$. It is not difficult to see that such quasi-IKG policy is also consistent, as formally stated in the following Theorem 3, whose proof is collected in the Appendix.

Theorem 3. *If Assumptions 1-3 hold, then the quasi-IKG policy as defined in eq. (15) is consistent.*

4. Stochastic Gradient Ascent

We now discuss computation of eq. (13) under Assumptions 1–3. It primarily consists of two steps.

(i) For each $i = 1, \ldots, M$, solve $\max_{\boldsymbol{x} \in \mathcal{X}} \text{IKG}^n(i, \boldsymbol{x})$ to find its maximizer, say $\boldsymbol{v}_i^n$.

(ii) Set $a^n = \operatorname{argmax}_{1 \leq i \leq M} \text{IKG}^n(i, \boldsymbol{v}_i^n)$ and set $\boldsymbol{v}^n = \boldsymbol{v}_{a^n}^n$.

Let $\boldsymbol{\xi}$ denote a $\mathcal{X}$-valued random variable with density $\gamma(\cdot)$, and

$$h_i^n(\boldsymbol{v}, \boldsymbol{x}) := |\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| \phi\left(\left| \frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})} \right| \right) - |\Delta_i^n(\boldsymbol{v})| \Phi\left(-\left| \frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})} \right| \right). \tag{16}$$

Then, we may rewrite eq. (14) as

$$\text{IKG}^n(i, \boldsymbol{x}) = [c_i(\boldsymbol{x})]^{-1} \mathbb{E}[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})], \tag{17}$$

which suggests the following sample average approximation,

$$\widehat{\text{IKG}}^n(i, \boldsymbol{x}) = \frac{1}{c_i(\boldsymbol{x}) J} \sum_{j=1}^{J} h_i^n(\boldsymbol{\xi}_j, \boldsymbol{x}), \tag{18}$$

where $\boldsymbol{\xi}_j$'s are independent copies of $\boldsymbol{\xi}$ and J is the sample size. In particular, we will use eq. (18) in step (ii) above for computing a^n for given $\boldsymbol{v}_i^n$'s. However, the sample average approximation method can easily become computationally prohibitive when applied to solve $\max_{\boldsymbol{x}} \text{IKG}^n(i, \boldsymbol{x})$ in step (i) if the domain $\mathcal{X}$ is high-dimensional. Hence, we consider instead the stochastic gradient ascent method to complete step (i).

Equation (17) means that in step (i) above, we solve the stochastic optimization problem

$$\boldsymbol{v}_i^n \in \operatorname*{argmax}_{\boldsymbol{x} \in \mathcal{X}} [c_i(\boldsymbol{x})]^{-1} \, \mathbb{E}\left[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})\right],$$

for each $i = 1, \ldots, M$. If $g_i^n(\boldsymbol{\xi}, \boldsymbol{x})$ is an unbiased estimator of $\frac{\partial}{\partial \boldsymbol{x}}\{[c_i(\boldsymbol{x})]^{-1} \, \mathbb{E}\left[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})\right]\}$, then $\boldsymbol{v}_i^n$ can be computed approximately using the stochastic gradient ascent (SGA) method; see Kushner and Yin (2003) for a comprehensive treatment and Newton et al. (2018) for a recent survey on the subject. Given an initial solution $\boldsymbol{x}_1 \in \mathcal{X}$ and a maximum iteration limit K, SGA iteratively computes

$$\boldsymbol{x}_{k+1} = \Pi_{\mathcal{X}}[\boldsymbol{x}_k + b_k g_i^n(\boldsymbol{\xi}_k, \boldsymbol{x}_k)], \quad k = 1, \ldots, K, \tag{19}$$

where $\Pi_{\mathcal{X}} : \mathbb{R}^d \mapsto \mathcal{X}$ denotes a projection mapping points outside $\mathcal{X}$ back to $\mathcal{X}$,[3] and b_k is referred as the step size that satisfies $\sum_{k=1}^{\infty} b_k = \infty$ and $\sum_{k=1}^{\infty} b_k^2 < \infty$. In general, the choice of b_k is crucial for the practical performance of SGA, and it is commonly set as $b_k = \alpha/k^{\beta}$ for some constants α and β.

Note that $g_i^n(\boldsymbol{\xi}, \boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}[h_i^n(\boldsymbol{\xi}, \boldsymbol{x})/c_i(\boldsymbol{x})]$ under mild regularity conditions (L'Ecuyer 1995). The explicit forms of $g_i^n(\boldsymbol{\xi}, \boldsymbol{x})$ for several common covariance functions are collected in the Appendix. Besides, in the implementation of SGA algorithm, practical modifications such as mini batch and *Polyak-Ruppert averaging* (Polyak and Juditsky 1992) can be adopted to achieve better performance. Detailed discussion is collected in the Appendix, together with other implementation issues of IKG policy.

5. Numerical Experiments

In this section, we evaluate the performance of the IKG policy via numerical experiments due to two reasons. First, the theoretical analysis, albeit establishing the consistency of the IKG policy in an large-sample asymptotic regime, does not provide a guarantee on the finite-sample performance of the policy. Second, the analysis has implicitly assumed that the sampling decisions of the IKG policy in eq. (13) can be computed exactly, while in practice it needs to be solved numerically via methods such as SGA that we have proposed. Additional numerical experiments on other issues, including the computational cost comparison between SGA versus the sample average approximation and the effect of estimated $\lambda_i(\boldsymbol{x})$, are collected in the Appendix. All the numerical experiments are implemented in MATLAB and the source code is available at `https://github.com/shenhaihui/ikg`.

---

[3]For example, one may set $\Pi_{\mathcal{X}}(\boldsymbol{x})$ to be the point in $\mathcal{X}$ closest to $\boldsymbol{x}$.

5.1. Finite-Sample Performance

The numerical experiments are conducted on synthetic problems, with the number of alternatives $M = 5$ and the dimensionality $d = 1, 3, 5, 7$. For each $i = 1, \ldots, M$, the true performance of alternative i is the revised Griewank function,

$$\theta_i(\boldsymbol{x}) = \sum_{j=1}^{d} \frac{x_j^2}{4000} - 1.5^{d-1} \prod_{j=1}^{d} \cos\left(\frac{x_j}{\sqrt{ij}}\right), \quad \boldsymbol{x} \in \mathcal{X} = [0, 10]^d.$$

Further, we set sampling variance $\lambda_i(\boldsymbol{x}) \equiv 0.01$, and take prior $\mu_i^0(\boldsymbol{x}) = \mu^0(\boldsymbol{x}) \equiv 0$, and $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = k^0(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{1}{d}\|\boldsymbol{x} - \boldsymbol{x}'\|^2\right)$. We set the cost function $c_i(\boldsymbol{x}) \equiv 1$ for each $i = 1, \ldots, M$, but will investigate the impact of a different cost function later.

We consider two density functions for the covariates: (1) uniform distribution on $\mathcal{X}$: $\gamma(\boldsymbol{x}) = 1/|\mathcal{X}|$; (2) multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $4^2\boldsymbol{I}$ truncated on $\mathcal{X}$: $\gamma(\boldsymbol{x}) = \phi(\boldsymbol{x}; \boldsymbol{0}, 4^2\boldsymbol{I})/\int_{\mathcal{X}} \phi(\boldsymbol{v}; \boldsymbol{0}, 4^2\boldsymbol{I})\,\mathrm{d}\boldsymbol{v}$. For convenience, we call the above specifications Problem 1 (P1) and Problem 2 (P2), respectively, depending on the choice of $\gamma(\boldsymbol{x})$.

The parameters involved in the SGA algorithm (see details in the Appendix) are given as follows: $K = 100d$, $K_0 = K/4$, $b_k = 200d/k^{0.7}$, $m = 20d$, and $J = 500d^2$. Moreover, the algorithm is started with a random initial solution. The performance of the IKG policy with respect to the sampling budget B is evaluated via the *opportunity cost* (OC), that is, the integrated difference in performance between the best alternative and the alternative chosen by the IKG policy upon exhausting the sampling budget.

$$\mathrm{OC}(B) := \mathbb{E}\left[\int_{\mathcal{X}} \left(\theta_{i^*(\boldsymbol{x})}(\boldsymbol{x}) - \theta_{\hat{i}^*(\boldsymbol{x};\omega)}(\boldsymbol{x})\right)\gamma(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right],$$

where $\hat{i}^*(\boldsymbol{x};\omega) \in \operatorname{argmax}_{1 \leq i \leq M} \mu_i^{N(B)}(\boldsymbol{x};\omega)$ is the learned decision rule up to the budget B under the IKG policy, ω denotes the samples taken under the policy, and the expectation is with respect to ω. Clearly, $\mathrm{OC}(B) \to 0$ as $B \to \infty$, since the IKG policy is consistent. We estimate $\mathrm{OC}(B)$ via

$$\widehat{\mathrm{OC}}(B) = \frac{1}{L}\sum_{l=1}^{L}\left[\frac{1}{J'}\sum_{j=1}^{J'}\left(\theta_{i^*(\boldsymbol{x}_j)}(\boldsymbol{x}_j) - \theta_{\hat{i}^*(\boldsymbol{x}_j;\omega_l)}(\boldsymbol{x}_j)\right)\right],$$

where $L = 30$ is the number of replications, ω_l denotes the samples for replication $l = 1, \ldots, L$, and $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{J'}\}$ is a random sample of the covariates generated from a given density function $\gamma(\boldsymbol{x})$ with $J' = 1000d^2$ for the purpose of evaluation.

We compare the IKG policy against three other polices:

- *IKG with Random Covariates (IKGwRC).* Recall that in the computation of IKG policy, random solution is used to initiate the SGA algorithm. To check whether such random initialization is a main cause for the effectiveness of IKG, we consider the IKGwRC policy as follows. Let $\boldsymbol{x}_1^n, \ldots, \boldsymbol{x}_M^n$ be the initial solutions for M alternatives used in the SGA algorithm when computing $(\hat{a}^n, \hat{\boldsymbol{v}}^n)$, $n = 0, 1, \ldots$ Then

the IKGwRC policy will sample at $(a^n, \boldsymbol{v}^n)$ given by

$$a^n = \underset{1 \leq i \leq M}{\operatorname{argmax}} \log \widehat{\text{IKG}}^n(i, \boldsymbol{x}_i^n) \quad \text{and} \quad \boldsymbol{v}^n = \boldsymbol{x}_{a^n}^n,$$

where the same samples are used to compute $\widehat{\text{IKG}}^n$ as in the IKG policy.

- *Binned Successive Elimination (BSE).* The BSE policy is proposed by Perchet and Rigollet (2013) for solving nonparametric MAB problems with covariates. In their setting, values of the covaraites arrive randomly, and the policy only determines which alternative to select. To implement BSE in our setting, we randomly generate $\boldsymbol{v}^n$ from uniform distribution on $\mathcal{X}$, and then apply the BSE policy to determine a^n. The BSE policy divides $\mathcal{X}$ into m^d parts, where m is the number of uniformly divided regions on each coordinate. For each problem, m is tuned within the set $\{1, \ldots, 10\}$, while other parameters follows the suggestion in Perchet and Rigollet (2013).

- *Pure Random Search (PRS).* The PRS policy will sample at $(a^n, \boldsymbol{v}^n)$, where a^n is randomly generated from the uniform distribution on $\{1, \ldots, M\}$ and $\boldsymbol{v}^n$ is generated from the uniform distribution on $\mathcal{X}$.

The performances of the four policies for problems P1 and P2 with $d = 1, 3, 5, 7$ are shown in Figures 1 and 2, respectively. Several findings are made as follows.

First, the estimated opportunity cost in all the test problems exhibits a clear trend of convergence to zero. This, from a practical point view, provides an assurance that the IKG policy in conjunction with the SGA algorithm indeed works as intended, that is, the uncertainty about the the performances of the competing alternatives will vanish eventually as the sampling budget grows. Second, the IKG policy can quickly reduce the opportunity cost when the sampling budget is relatively small, but the reduction appears to slow down as the sampling budget increases. This finding is consist with prior research on other KG-type policies such as Frazier et al. (2009), Frazier and Powell (2011), and Xie et al. (2016). Third, the learning task of identifying the best alternative becomes substantially more difficult when the dimensionality of the covariates is large. This can be seen from the growing sampling budget and the slowing reduction in the opportunity cost as d increases.

Overall, IKG outperforms the other three policies. Specific comparisons are as follows. First, IKG has better performance than IKGwRC, especially when the dimensionality is high, which indicates that the SGA algorithm in IKG for solving $\boldsymbol{v}_i^n$ (see Section 4) indeed works well and has a significant effect in IKG. Second, BSE has inferior performance than IKG, which may be caused by the fact that BSE only optimizes a^n given randomly observed $\boldsymbol{v}^n$, while IKG optimizes both a^n and $\boldsymbol{v}^n$ at the same time. Third, PRS overall has the worst performance, which is not surprising since it does not utilize any information gained from previous sampling. Note that PRS is a consistent policy, but the consistency does not guarantee any finite-sample performance. This reflects the value of IKG – it is not only provably consistent, but also takes advantage of information gained from previous samples to yield good finite-sample performance.

Figure 1: Estimated opportunity cost (vertical axis) as a function of the sampling budget (horizontal axis) for P1.
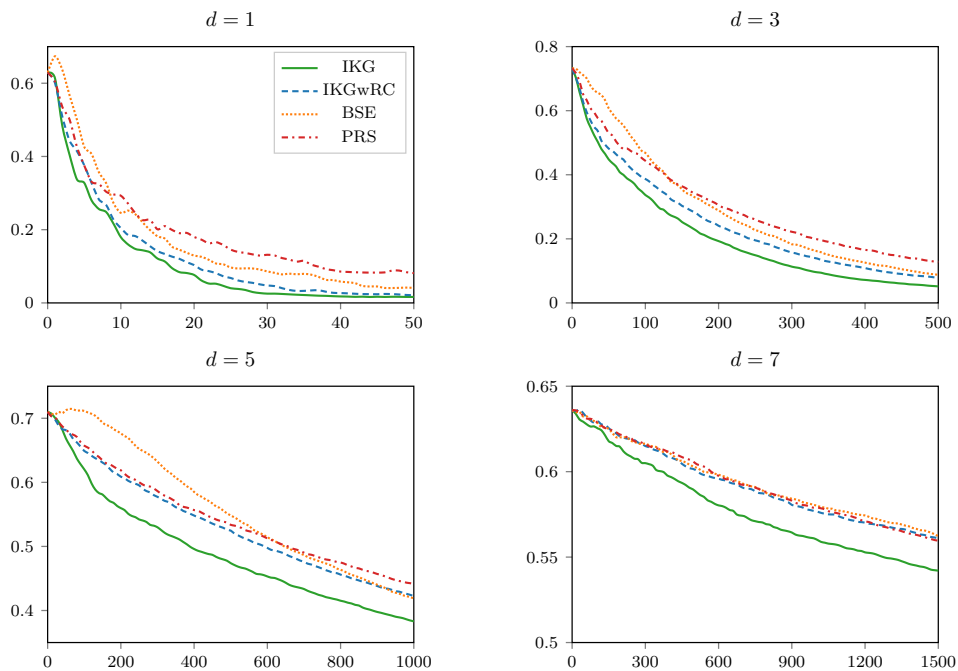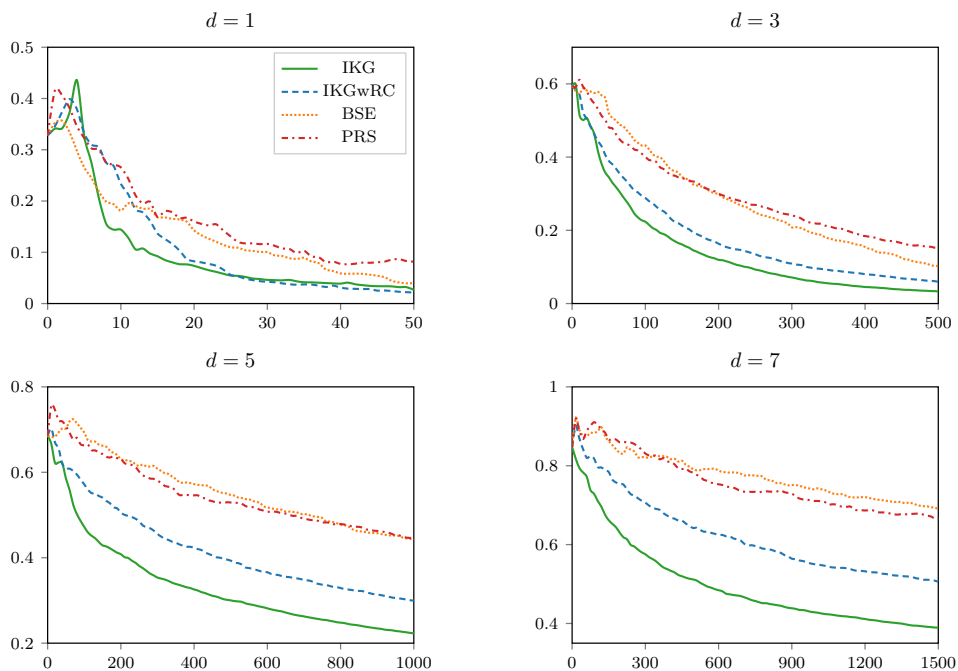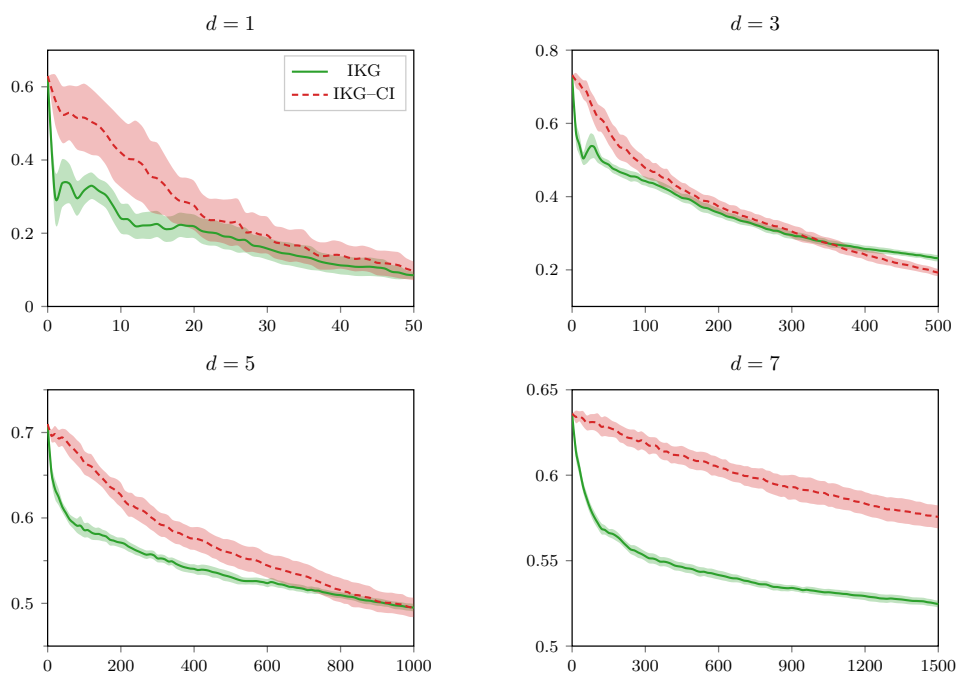

Figure 2: Estimated opportunity cost (vertical axis) as a function of the sampling budget (horizontal axis) for P2.

5.2. Effect of Sampling Cost

We are also interested in the effect of sampling costs on the IKG policy. In particular, we consider a different cost function other than the unit cost function: $c_i(\boldsymbol{x}) = 2^{3-i}\big(1 + \|\boldsymbol{x} - \boldsymbol{5}\|^2/(10d)\big)$, where $\boldsymbol{5}$ is a $d \times 1$ vector of all fives. We set $\gamma(\boldsymbol{x})$ to be the uniform density[4] and call this specification Problem 3 (P3). We compare two scenarios: (i) the sampling cost is incorporated correctly; and (ii) one ignores variations in the sampling cost at different locations and mistakenly uses the unit sampling cost when implementing the IKG policy (but the actual sampling consumption follows $c_i(\boldsymbol{x})$). The comparison is illustrated in Figure 3.

Figure 3: Estimated opportunity cost (vertical axis) as a function of the sampling budget (horizontal axis) for P3.

NOTE. IKG–CI means sampling costs are ignored when implementing the IKG policy. The shaded regions represent the 99% confidence intervals.

There are two observations. On one hand, despite the misspecification in the sampling cost function, the IKG policy is still consistent, with the associated opportunity cost converging to zero. This is not surprising, because using the unit sampling cost function, i.e., $c_i(\boldsymbol{x}) \equiv 1$, is exactly the setup of Theorem 2. On the other hand, however, the finite-sample performance of the IKG policy indeed deteriorates as a result of the misspecification. Further, the deterioration appears to become more significant as the dimensionality of the covariates increases.

---

[4]Setting $\gamma(\boldsymbol{x})$ to be the truncated normal density leads to similar findings.

6. Conclusions

In this paper, we study sequential sampling for the problem of selection with covariates which aims to identify the best alternative as a function of the covariates. Each sampling decision involves choosing an alternative and a value of the covariates, from the pair of which a sample will be taken. We design a sequential sampling policy via a nonparametric Bayesian approach. In particular, following the well-known KG design principle for simulation optimization, we develop the IKG policy that attempts to maximize the "one-step" integrated increment in the expected value of information per unit of sampling cost.

We prove the consistency of the IKG policy under minimal assumptions. Compared to prior work on asymptotic analysis of KG-type sampling policies, our assumptions are simpler and significantly more general, thanks to technical machinery that we develop based on RKHS theory. Nevertheless, to compute the sampling decisions of the IKG policy requires solving a multi-dimensional stochastic optimization problem. To that end, we develop a numerical algorithm based on the SGA method. Numerical experiments illustrate the finite-sample performance of the IKG policy and provide a practical assurance that the developed methodology works as intended.

References

Adler, R. J. and J. E. Taylor (2007). *Random Fields and Geometry*. Springer.

Ankenman, B., B. L. Nelson, and J. Staum (2010). Stochastic kriging for simulation metamodeling. *Oper. Res. 58*(2), 371–382.

Arora, N., X. Dreze, A. Ghose, J. D. Hess, et al. (2008). Putting one-to-one marketing to work: Personalization, customization, and choice. *Market. Lett. 19*(3-4), 305.

Bect, J., F. Bachoc, and D. Discourager (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli 25*(4A), 2883–2919.

Berlinet, A. and C. Thomas-Agnan (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer.

Bubeck, S. and N. Cesa-Bianchi (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Found. Trends Mach. Learn. 5*(1), 1–122.

Chen, C.-H., S. E. Chick, L. H. Lee, and N. A. Pujowidianto (2015). Ranking and selection: Efficient simulation budget allocation. In M. C. Fu (Ed.), *Handbook of Simulation Optimization*, pp. 45–80. Springer.

Choi, S. E., K. E. Perzan, A. C. Tramontano, C. Y. Kong, and C. Hur (2014). Statins and aspirin for chemoprevention in Barrett's esophagus: Results of a cost-effectiveness analysis. *Canc. Prev. Res. 7*(3), 341–350.

Frazier, P., W. Powell, and S. Dayanik (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS J. Comput. 21*(4), 599–613.

Frazier, P. I., W. Powell, and S. Dayanik (2008). A knowledge gradient policy for sequential information collection. *SIAM J. Control Optim. 47*(5), 2410–2439.

Frazier, P. I. and W. B. Powell (2011). Consistency of sequential Bayesian sampling policies. *SIAM J. Control Optim. 49*(2), 712–731.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis* (2nd ed.). Cambridge University Press.

Hu, R. and M. Ludkovski (2017). Sequential design for ranking response surfaces. *SIAM/ASA J. Uncertainty Quantification 5*(1), 212–239.

Hur, C., N. S. Nishioka, and G. S. Gazelle (2004). Cost-effectiveness of aspirin chemoprevention for Barrett's esophagus. *J. Natl. Canc. Inst. 96*(4), 316–325.

Kim, E. S., R. S. Herbst, I. I. Wistuba, J. J. Lee, et al. (2011). The BATTLE trial: Personalizing therapy for lung cancer. *Canc. Discov. 1*(1), 44–53.

Kim, S.-H. and B. L. Nelson (2006). Selecting the best system. In S. G. Henderson and B. L. Nelson (Eds.), *Handbooks in Operations Research and Management Science*, Volume 13, pp. 501–534. Elsevier.

Krause, A. and C. S. Ong (2011). Contextual Gaussian process bandit optimization. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 24*, pp. 2447–2455.

Kushner, H. J. and G. G. Yin (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer-Verlag.

L'Ecuyer, P. (1995). Note: On the interchange of derivative and expectation for likelihood ratio derivative estimators. *Manag. Sci. 41*(4), 738–747.

Mes, M. R., W. B. Powell, and P. I. Frazier (2011). Hierarchical knowledge gradient for sequential sampling. *J. Mach. Learn. Res. 12*, 2931–2974.

Newton, D., F. Yousefian, and R. Pasupathy (2018). Stochastic gradient descent: Recent trends. In E. Gel and D. Lewis (Eds.), *TutORials in Operations Research*, Volume 7, pp. 193–220. INFORMS.

Pearce, M. and J. Branke (2017). Efficient expected improvement estimation for continuous multiple ranking and selection. In *Proc. 2017 Winter Simulation Conf.*, pp. 2161–2172.

Pearce, M. and J. Branke (2018). Continuous multi-task Bayesian optimisation with correlation. *Eur. J. Oper. Res. 270*(3), 1074–1085.

Perchet, V. and P. Rigollet (2013). The multi-armed bandit problem with covariates. *Ann. Stat. 41*(2), 693–721.

Poloczek, M., J. Wang, and P. I. Frazier (2017). Multi-information source optimization. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, pp. 4288–4298.

Polyak, B. T. and A. B. Juditsky (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim. 30*(4), 838–855.

Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Rusmevichientong, P. and J. N. Tsitsiklis (2010). Linearly parameterized bandits. *Math. Oper. Res. 35*(2), 395–411.

Ryzhov, I. O. (2016). On the convergence rates of expected improvement methods. *Oper. Res. 64*(6), 1515–1528.

Scott, W., P. Frazier, and W. Powell (2011). The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression. *SIAM J. Optim. 21*(3), 996–1026.

Shen, H., L. J. Hong, and X. Zhang (2021). Ranking and selection with covariates for personalized decision making. *INFORMS J. Comput.*, forthcoming.

Steinwart, I. and A. Christmann (2008). *Support Vector Machines*. Springer.

Toscano-Palmerin, S. and P. I. Frazier (2018). Bayesian optimization with expensive integrands. *arXiv:1803.08661*.

Wu, J. and P. I. Frazier (2016). The parallel knowledge gradient method for batch Bayesian optimization. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 29*, pp. 3126–3134.

Wu, J., M. Poloczek, A. G. Wilson, and P. I. Frazier (2017). Bayesian optimization with gradients. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, pp. 5267–5278.

Xie, J., P. I. Frazier, and S. E. Chick (2016). Bayesian optimization via simulation with pairwise sampling and correlated prior beliefs. *Oper. Res. 64*(2), 542–559.

Yang, Y. and D. Zhu (2002). Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *Ann. Stat. 30*(1), 100–121.

Appendix

A. Proof of Lemma 1

Before proving Lemma 1, we first establish the following Lemma 2.

Lemma 2. *Let $g(s,t) := t\phi(s/t) - s\Phi(-s/t)$, where Φ is the standard normal distribution function and ϕ is its density function. Then,*

(i) $g(s,t) > 0$ *for all $s \geq 0$ and $t > 0$;*

(ii) $g(s,t)$ *is strictly decreasing in $s \in [0, \infty)$ and strictly increasing in $t \in (0, \infty)$;*

(iii) $g(s,t) \to 0$ *as $s \to \infty$ or as $t \to 0$.*

Proof of Lemma 2. Let $h(u) := \phi(u) - u\Phi(-u)$ for $u \geq 0$, then $g(s,t) = th(s/t)$. Note that

$$h'(u) = \phi'(u) + u\phi(-u) - \Phi(-u) = -u\phi(u) + u\phi(u) - \Phi(-u) = -\Phi(-u) < 0,$$

Hence, $h(u)$ is strictly decreasing in $u \in [0, \infty)$. Note that $\lim_{u\to\infty} \phi(u) = 0$ and

$$\lim_{u\to\infty} u\Phi(-u) = \lim_{u\to\infty} \frac{\Phi(-u)}{u^{-1}} = \lim_{u\to\infty} \frac{\phi(u)}{u^{-2}} = \lim_{u\to\infty} \frac{u^2}{\sqrt{2\pi}e^{u^2/2}} = 0,$$

hence $\lim_{u\to\infty} h(u) = 0$. Then we must have $h(u) > 0$ for all $u \in [0, \infty)$, from which part (i) follows immediately.

For part (ii), the strict decreasing monotonicity of $g(s,t)$ in $s \in [0, \infty)$ and the strict increasing monotonicity of $g(s,t)$ in $t \in (0, \infty)$ follow immediately from the strict decreasing monotonicity of $h(u)$ in u and $g(s,t) = th(s/t)$.

Part (iii) is due to that $\lim_{u\to\infty} h(u) = 0$ and $g(s,t) = th(s/t)$.   □

Now we are ready to prove Lemma 1

Proof of Lemma 1. By eqs. (5) and (7),

$$f(i, \boldsymbol{x}, \boldsymbol{v}) := \mathbb{E}\left[\max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v}) \,\Big|\, \mathcal{F}^n, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right]$$

$$= \mathbb{E}\left[\max_{1 \leq a \leq M} \left(\mu_a^n(\boldsymbol{v}) + \sigma_a^n(\boldsymbol{v}, \boldsymbol{x})Z^{n+1}\right)\right]$$

$$= \mathbb{E}\left[\max\left\{\mu_i^n(\boldsymbol{v}) + \tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})Z^{n+1}, \, \max_{a\neq i}\mu_a^n(\boldsymbol{v})\right\}\right]$$

$$= \mathbb{E}\left[\max\left\{\mu_i^n(\boldsymbol{v}) + |\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})|Z^{n+1}, \, \max_{a\neq i}\mu_a^n(\boldsymbol{v})\right\}\right]. \tag{20}$$

For notational simplicity, let $\alpha := \mu_i^n(\boldsymbol{v})$, $\beta := \tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})$, $\gamma := \max_{a\neq i}\mu_a^n(\boldsymbol{v})$, and $\delta := \alpha - \gamma$.

If $\beta \neq 0$, then

$$f(i, \boldsymbol{x}, \boldsymbol{v}) = \int_{-\infty}^{-\delta/|\beta|} \gamma \phi(z) \, \mathrm{d}z + \int_{-\delta/|\beta|}^{\infty} (\alpha + |\beta|z)\phi(z) \, \mathrm{d}z$$

$$= \gamma \Phi(-\delta/|\beta|) + \alpha[1 - \Phi(-\delta/|\beta|)] + \beta \phi(-\delta/|\beta|),$$

where the second equality follows from the identity $\int_t^\infty z\phi(z) \, \mathrm{d}z = \phi(t)$ for all $t \in \mathbb{R}$. Next, we calculate the integrand in eq. (12). By noting that $\phi(z) = \phi(-z)$ and $\Phi(-z) = 1 - \Phi(z)$ for all $z \in \mathbb{R}$,

$$f(i, \boldsymbol{x}, \boldsymbol{v}) - \max_{1 \leq a \leq M} \mu_a^n(\boldsymbol{v}) = f(i, \boldsymbol{x}, \boldsymbol{v}) - \max(\alpha, \gamma) = \begin{cases} |\beta|\phi(\delta/|\beta|) - \delta\Phi(-\delta/|\beta|), & \text{if } \alpha \geq \gamma \\ |\beta|\phi(\delta/|\beta|) + \delta\Phi(\delta/|\beta|), & \text{if } \alpha < \gamma \end{cases}$$

$$= |\beta|\phi(|\delta/\beta|) - |\delta|\Phi(-|\delta/\beta|). \tag{21}$$

If $\beta = 0$, then it is straightforward from eq. (20) to see that $f(i, \boldsymbol{x}, \boldsymbol{v}) - \max_{1 \leq a \leq M} \mu_a^n(\boldsymbol{v}) = 0$. On the other hand, by Lemma 2 (iii), we can set the right-hand side of eq. (21) to be zero for $\beta = 0$. Hence, eq. (21) holds for $\beta = 0$ as well.

Replacing the integrand in eq. (12) with eq. (21) yields the expression of $\text{IKG}^n(i, \boldsymbol{x})$ in Lemma 1. $\qquad \square$

B. Proof of Proposition 1

To simplify notation, in this subsection we assume $M = 1$ and suppress the subscript i unless otherwise specified, but the results can be generalized to the case of $M > 1$ without essential difficulty. In particular, we use κ to denote a generic covariance function, k^0 the prior covariance function of a Gaussian process, and k^n the posterior covariance function. We will collect below several basic results on reproducing kernel Hilbert space (RKHS) and refer to Berlinet and Thomas-Agnan (2004) for an extensive treatment on the subject.

Definition 2. Let $\mathcal{X}$ be a nonempty set and κ be a covariance function on $\mathcal{X}$. A Hilbert space $\mathcal{H}_\kappa$ of functions on $\mathcal{X}$ equipped with an inner-product $\langle \cdot, \cdot \rangle_{\mathcal{H}_\kappa}$ is called a RKHS with reproducing kernel κ, if (i) $\kappa(\boldsymbol{x}, \cdot) \in \mathcal{H}_\kappa$ for all $\boldsymbol{x} \in \mathcal{X}$, and (ii) $f(\boldsymbol{x}) = \langle f, \kappa(\boldsymbol{x}, \cdot) \rangle_{\mathcal{H}_\kappa}$ for all $\boldsymbol{x} \in \mathcal{X}$ and $f \in \mathcal{H}_\kappa$. Furthermore, the norm of $\mathcal{H}_\kappa$ is induced by the inner-product, i.e., $\|f\|_{\mathcal{H}_\kappa}^2 = \langle f, f \rangle_{\mathcal{H}_\kappa}$ for all $f \in \mathcal{H}_\kappa$.

Remark 4. In Definition 2, for a fixed $\boldsymbol{x}$, $\kappa(\boldsymbol{x}, \cdot)$ is understood as a function mapping $\mathcal{X}$ to $\mathbb{R}$ such that $\boldsymbol{y} \mapsto k(\boldsymbol{x}, \boldsymbol{y})$ for $\boldsymbol{y} \in \mathcal{X}$. Moreover, condition (ii) is called the *reproducing property*. In particular, it implies that $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \kappa(\boldsymbol{x}, \cdot), \kappa(\boldsymbol{x}', \cdot) \rangle_{\mathcal{H}_\kappa}$ and $\kappa(\boldsymbol{x}, \boldsymbol{x}) = \|\kappa(\boldsymbol{x}, \cdot)\|_{\mathcal{H}_\kappa}^2$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$.

Remark 5. By Moore-Aronszajn theorem (Berlinet and Thomas-Agnan 2004, Theorem 3), for each covariance function κ there exists a unique RKHS $\mathcal{H}_\kappa$ for which κ is its reproducing kernel. Specifically,

$$\mathcal{H}_\kappa = \left\{ f = \sum_{i=1}^\infty c_i \kappa(\boldsymbol{x}_i, \cdot) : c_i \in \mathbb{R}, \boldsymbol{x}_i \in \mathcal{X}, i = 1, 2, \ldots, \text{ such that } \|f\|_{\mathcal{H}_\kappa}^2 < \infty \right\},$$

where $\|f\|^2_{\mathcal{H}_\kappa} := \sum_{i,j=1}^{\infty} c_i c_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Moreover, the inner-product is defined by

$$\langle f, g \rangle_{\mathcal{H}_\kappa} = \sum_{i,j=1}^{\infty} a_i b_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}'_j),$$

for any $f = \sum_{i=1}^{\infty} a_i \kappa(\boldsymbol{x}_i, \cdot) \in \mathcal{H}_\kappa$ and $g = \sum_{j=1}^{\infty} b_j \kappa(\boldsymbol{x}'_j, \cdot) \in \mathcal{H}_\kappa$.

The following lemma asserts that convergence in norm in a RKHS implies *uniform* pointwise convergence, provided that the covariance function κ is stationary.

Lemma 3. *Let $\mathcal{X}$ be a nonempty set and κ be a covariance function on $\mathcal{X}$. Suppose that a sequence of functions $\{f_n \in \mathcal{H}_\kappa : n = 1, 2, \ldots\}$ converges in norm $\|\cdot\|_{\mathcal{H}_\kappa}$ as $n \to \infty$. Then the limit, denoted by f, is in $\mathcal{H}_\kappa$. Moreover, if κ is stationary, then $f_n(\boldsymbol{x}) \to f(\boldsymbol{x})$ as $n \to \infty$ uniformly in $\boldsymbol{x} \in \mathcal{X}$.*

Proof of Lemma 3. First of all, $f \in \mathcal{H}_\kappa$ is guaranteed as a Hilbert space is a complete metric space. A basic property of RKHS is that convergence in norm implies pointwise convergence to the same limit; see, e.g., Corollary 1 of Berlinet and Thomas-Agnan (2004, page 10). Namely, $f_n(\boldsymbol{x}) \to f(\boldsymbol{x})$ as $n \to \infty$ for all $\boldsymbol{x} \in \mathcal{X}$.

To show the pointwise convergence is uniform, note that since κ is stationary, there exists a function $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$ such that $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \varphi(\boldsymbol{x} - \boldsymbol{x}')$. Hence, $\|\kappa(\boldsymbol{x}, \cdot)\|^2_{\mathcal{H}_\kappa} = \kappa(\boldsymbol{x}, \boldsymbol{x}) = \varphi(\boldsymbol{0})$. It follows that

$$\begin{aligned}
|f_{n+m}(\boldsymbol{x}) - f_n(\boldsymbol{x})| &= |\langle f_{n+m} - f_n, \kappa(\boldsymbol{x}, \cdot) \rangle_{\mathcal{H}_\kappa}| \\
&\leq \|f_{n+m} - f_n\|_{\mathcal{H}_\kappa} \|\kappa(\boldsymbol{x}, \cdot)\|_{\mathcal{H}_\kappa} = \|f_{n+m} - f_n\|_{\mathcal{H}_\kappa} \sqrt{\varphi(\boldsymbol{0})},
\end{aligned} \tag{22}$$

for all n and m, where the first equality follows from the reproducing property.

Since a Hilbert space is a complete metric space, the $\|\cdot\|_{\mathcal{H}_\kappa}$-converging sequence $\{f_n\}$ is a Cauchy sequence in $\mathcal{H}_\kappa$, meaning that $\|f_{n+m} - f_n\|_{\mathcal{H}_\kappa} \to 0$ as $n \to \infty$ for all m. Since this convergence to zero is independent of $\boldsymbol{x}$, it follows from eq. (22) that $\{f_n\}$ is a uniform Cauchy sequence of functions, thereby converging to f uniformly in $\boldsymbol{x} \in \mathcal{X}$. $\qquad\square$

In the light of Lemma 3, in order to establish the uniform convergence of $k^n(\boldsymbol{x}, \boldsymbol{x}')$ as a function of $\boldsymbol{x}'$, it suffices to prove the norm convergence of $k^n(\boldsymbol{x}, \cdot)$ in the RKHS induced by k^0. We first establish this result for a more general case in the following Lemma 4, where k^0 is not required to be stationary.

Lemma 4. *Let $\mathcal{H}_{k^0}$ be the RKHS induced by k^0. If $k^0(\boldsymbol{x}, \boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{X}$, then for any $\boldsymbol{x} \in \mathcal{X}$, $k^n(\boldsymbol{x}, \cdot)$ converges in norm $\|\cdot\|_{\mathcal{H}_{k^0}}$ as $n \to \infty$.*

Proof of Lemma 4. Fix $\boldsymbol{x} \in \mathcal{X}$. The fact that $k^n(\boldsymbol{x}, \cdot) \in \mathcal{H}_{k^0}$ is due to eq. (4). It follows from eq. (8) that $\{k^n(\boldsymbol{x}, \boldsymbol{x}) : n \geq 1\}$ form a non-increasing sequence bounded below by zero. The monotone convergence theorem implies that $k^n(\boldsymbol{x}, \boldsymbol{x})$ converges as $n \to \infty$. Hence, for all $m \geq 1$,

$$\lim_{n \to \infty} |k^{n+m}(\boldsymbol{x}, \boldsymbol{x}) - k^n(\boldsymbol{x}, \boldsymbol{x})| = 0. \tag{23}$$

Let $\boldsymbol{V}^n := \{\boldsymbol{v}^\ell : \ell = 0, \ldots, n-1\}$ and $\boldsymbol{V}_n^{n+m} := \{\boldsymbol{v}^\ell : \ell = n, \ldots, n+m-1\}$. Then, by eq. (4),

$$k^{n+m}(\boldsymbol{x}, \cdot) - k^n(\boldsymbol{x}, \cdot) = -k^n(\boldsymbol{x}, \boldsymbol{V}_n^{n+m})[k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}_n^{n+m}) + \lambda(\boldsymbol{V}_n^{n+m})]^{-1} k^n(\boldsymbol{V}_n^{n+m}, \cdot). \tag{24}$$

For notational simplicity, let $\boldsymbol{\Sigma}_n^{n+m} := k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}_n^{n+m}) + \lambda(\boldsymbol{V}_n^{n+m})$. Then,

$$\left\| k^{n+m}(\boldsymbol{x}, \cdot) - k^n(\boldsymbol{x}, \cdot) \right\|_{\mathcal{H}_{k^0}}^2$$

$$= \langle k^n(\boldsymbol{x}, \boldsymbol{V}_n^{n+m})[\boldsymbol{\Sigma}_n^{n+m}]^{-1} k^n(\boldsymbol{V}_n^{n+m}, \cdot), k^n(\boldsymbol{x}, \boldsymbol{V}_n^{n+m})[\boldsymbol{\Sigma}_n^{n+m}]^{-1} k^n(\boldsymbol{V}_n^{n+m}, \cdot) \rangle_{\mathcal{H}_{k^0}}$$

$$= k^n(\boldsymbol{x}, \boldsymbol{V}_n^{n+m})[\boldsymbol{\Sigma}_n^{n+m}]^{-1} \langle k^n(\boldsymbol{V}_n^{n+m}, \cdot), k^n(\boldsymbol{V}_n^{n+m}, \cdot) \rangle_{\mathcal{H}_{k^0}} [\boldsymbol{\Sigma}_n^{n+m}]^{-1} k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{x}), \qquad (25)$$

where

$$\langle k^n(\boldsymbol{V}_n^{n+m}, \cdot), k^n(\boldsymbol{V}_n^{n+m}, \cdot) \rangle_{\mathcal{H}_{k^0}} = \left(\langle k^n(\boldsymbol{v}^{n+i}, \cdot), k^n(\boldsymbol{v}^{n+j}, \cdot) \rangle_{\mathcal{H}_{k^0}} \right)_{0 \le i,j \le m-1}.$$

Moreover, note that by eq. (4),

$$k^n(\boldsymbol{V}_n^{n+m}, \cdot) = k^0(\boldsymbol{V}_n^{n+m}, \cdot) - k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}^n)[k^0(\boldsymbol{V}^n, \boldsymbol{V}^n) + \lambda(\boldsymbol{V}^n)]^{-1} k^0(\boldsymbol{V}^n, \cdot). \qquad (26)$$

Let $\boldsymbol{\Sigma}^n := k^0(\boldsymbol{V}^n, \boldsymbol{V}^n) + \lambda(\boldsymbol{V}^n)$. Then, it follows from eq. (26) and the reproducing property that

$$\langle k^n(\boldsymbol{V}_n^{n+m}, \cdot), k^n(\boldsymbol{V}_n^{n+m}, \cdot) \rangle_{\mathcal{H}_{k^0}}$$

$$= k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}_n^{n+m}) - 2k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} k^0(\boldsymbol{V}^n, \boldsymbol{V}_n^{n+m})$$

$$\quad + k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} k^0(\boldsymbol{V}^n, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} k^0(\boldsymbol{V}^n, \boldsymbol{V}_n^{n+m})$$

$$= k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}_n^{n+m}) - k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} k^0(\boldsymbol{V}^n, \boldsymbol{V}_n^{n+m})$$

$$\quad - k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} \{\boldsymbol{I} - k^0(\boldsymbol{V}^n, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1}\} k^0(\boldsymbol{V}^n, \boldsymbol{V}_n^{n+m})$$

$$= k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}_n^{n+m}) - k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} \{\boldsymbol{I} - k^0(\boldsymbol{V}^n, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1}\} k^0(\boldsymbol{V}^n, \boldsymbol{V}_n^{n+m}), \qquad (27)$$

where $\boldsymbol{I}$ denotes the identity matrix of a compatible size. Furthermore, note that

$$\boldsymbol{I} - k^0(\boldsymbol{V}^n, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} = \boldsymbol{I} - [k^0(\boldsymbol{V}^n, \boldsymbol{V}^n) + \lambda(\boldsymbol{V}^n) - \lambda(\boldsymbol{V}^n)][k^0(\boldsymbol{V}^n, \boldsymbol{V}^n) + \lambda(\boldsymbol{V}^n)]^{-1}$$

$$= \boldsymbol{I} - \boldsymbol{I} + \lambda(\boldsymbol{V}^n)[k^0(\boldsymbol{V}^n, \boldsymbol{V}^n) + \lambda(\boldsymbol{V}^n)]^{-1}$$

$$= \lambda(\boldsymbol{V}^n)[\boldsymbol{\Sigma}^n]^{-1}. \qquad (28)$$

We now combine eqs. (27) and (28) to have

$$\langle k^n(\boldsymbol{V}_n^{n+m}, \cdot), k^n(\boldsymbol{V}_n^{n+m}, \cdot) \rangle_{\mathcal{H}_{k^0}}$$

$$= k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}_n^{n+m}) - k^0(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}^n)[\boldsymbol{\Sigma}_n]^{-1} \lambda(\boldsymbol{V}^n)[\boldsymbol{\Sigma}^n]^{-1} k^0(\boldsymbol{V}^n, \boldsymbol{V}_n^{n+m}),$$

which is the difference between two positive semi-definite matrices. Therefore, by eq. (25),

$$\left\| k^{n+m}(\boldsymbol{x}, \cdot) - k^n(\boldsymbol{x}, \cdot) \right\|_{\mathcal{H}_{k^0}}^2 \le k^n(\boldsymbol{x}, \boldsymbol{V}_n^{n+m})[\boldsymbol{\Sigma}_n^{n+m}]^{-1} k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{V}_n^{n+m})[\boldsymbol{\Sigma}_n^{n+m}]^{-1} k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{x})$$

$$\le k^n(\boldsymbol{x}, \boldsymbol{V}_n^{n+m})[\boldsymbol{\Sigma}_n^{n+m}]^{-1} \boldsymbol{\Sigma}_n^{n+m}[\boldsymbol{\Sigma}_n^{n+m}]^{-1} k^n(\boldsymbol{V}_n^{n+m}, \boldsymbol{x})$$

$$= k^n(\boldsymbol{x}, \boldsymbol{x}) - k^{n+m}(\boldsymbol{x}, \boldsymbol{x}),$$

where the second inequality follows from the definition of Σ_n^{n+m} and the equality follows from eq. (24). Then, we apply eq. (23) to conclude that $\|k^{n+m}(\boldsymbol{x}, \cdot) - k^n(\boldsymbol{x}, \cdot)\|_{\mathcal{H}_{k^0}} \to 0$ as $n \to \infty$ for all $m \geq 1$. Therefore, $k^n(\boldsymbol{x}, \cdot)$ converges in norm $\|\cdot\|_{\mathcal{H}_{k^0}}$ as $n \to \infty$.   □

With Lemmas 3 and 4, we are ready to prove Proposition 1.

Proof of Proposition 1. Fix $i = 1, \ldots, M$. Since k_i^0 is stationary, $k_i^0(\boldsymbol{x}, \boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{X}$. Then by Lemma 4, for any $\boldsymbol{x} \in \mathcal{X}$, $k_i^n(\boldsymbol{x}, \cdot)$ converges in norm $\|\cdot\|_{\mathcal{H}_{k^0}}$ as $n \to \infty$. Then by Lemma 3, for the $\|\cdot\|_{\mathcal{H}_{k^0}}$-converging limit $k_i^\infty(\boldsymbol{x}, \cdot)$, $k_i^n(\boldsymbol{x}, \boldsymbol{x}') \to k_i^\infty(\boldsymbol{x}, \boldsymbol{x}')$ uniformly in $\boldsymbol{x}' \in \mathcal{X}$ as $n \to \infty$.   □

C. Proof of Proposition 2

Notice that if $\eta_i^\infty = \infty$ under a sampling policy π, then due to the compactness of $\mathcal{X}$, $\{\boldsymbol{v}^n : a^n = i, n = 0, 1, \ldots\}$ (i.e., the sampling locations associated with alternative i) must have an accumulation point $\boldsymbol{x}_i^{\mathrm{acc}} \in \mathcal{X}$. Namely, there exists a subsequence of $\{n : a^n = i, n = 0, 1, \ldots\}$, say $\{\ell_{i,n}\}_{n=0}^\infty$, such that $\ell_{i,n} \to \infty$ and $\boldsymbol{v}^{\ell_{i,n}} \to \boldsymbol{x}_i^{\mathrm{acc}}$ as $n \to \infty$. For any $\epsilon > 0$, let $\mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon) := \{\boldsymbol{x} : \|\boldsymbol{x}_i^{\mathrm{acc}} - \boldsymbol{x}\| \leq \epsilon\}$ be the closed ball centered at $\boldsymbol{x}_i^{\mathrm{acc}}$ with radius ϵ. Let $\mathrm{Var}^{\pi,n}[\cdot]$ denote the posterior variance conditioned on $\mathcal{F}^n$ that is induced by π.

The proof of Proposition 2 is preceded by four technical results, i.e., Lemmas 5–8. In Lemmas 5 and 6, we establish an upper bound on $\mathrm{Var}^{\pi,n}[\theta_i(\boldsymbol{x})]$ for $\boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$. This result does not rely on the IKG policy *per se*, but is implied by the existence of the accumulation point $\boldsymbol{x}_i^{\mathrm{acc}}$ instead. In particular, the upper bound which depends on ϵ can be made arbitrarily small as $\epsilon \to 0$. This basically means that in the light of an unlimited number of samples of alternative i that are taken in proximity to $\boldsymbol{x}_i^{\mathrm{acc}}$, the uncertainty about $\theta_i(\boldsymbol{x}_i^{\mathrm{acc}})$ will eventually be eliminated, thanks to the correlation between $\theta_i(\boldsymbol{x}_i^{\mathrm{acc}})$ and $\theta_i(\boldsymbol{x})$ for $\boldsymbol{x}$ near $\boldsymbol{x}_i^{\mathrm{acc}}$.

Lemma 7 asserts that $\mathrm{IKG}^n(i, \boldsymbol{x})$ is bounded by a multiple of the posterior standard deviation of $\theta_i(\boldsymbol{x})$. This implies that when the posterior variance approaches to zero, the IKG factor does too.

Following the last three lemmas, Lemma 8 asserts that the limit inferior of the IKG factor is zero. The reasoning is as follows. By Lemmas 5 and 6, the posterior variance at those sampling locations that fall inside $\mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$ is small. Then, by Lemma 7, the IKG factor at these locations are also small, so does the limit superior. Since the sampling locations inside $\mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$ is a subsequence of the entire sampling locations, the limit inferior of the IKG factor over all sampling locations is even smaller.

At last, Proposition 2 is proved by contradiction — if there is a location that the posterior variance does not approach zero, then the limit inferior of the IKG factor at the same location must be positive as well.

Lemma 5. *Fix $i = 1, \ldots, M$, $n \geq 1$, and a compact set $\mathcal{S} \subseteq \mathcal{X}$. Suppose that the sampling decisions satisfy $a^0 = \cdots = a^{n-1} = i$ and $\boldsymbol{v}^0, \ldots, \boldsymbol{v}^{n-1} \in \mathcal{S}$. If Assumptions 1-3 hold, then for all $\boldsymbol{x} \in \mathcal{S}$,*

$$\mathrm{Var}^n[\theta_i(\boldsymbol{x})] \leq \tau_i^2 - \frac{n \min_{\boldsymbol{x}' \in \mathcal{S}} \left[k_i^0(\boldsymbol{x}, \boldsymbol{x}') \right]^2}{n\tau_i^2 + \lambda_i^{\max}},$$

where $\lambda_i^{\max} := \max_{\boldsymbol{x} \in \mathcal{X}} \lambda_i(\boldsymbol{x}) \in (0, \infty)$.

Proof of Lemma 5. Fix $\boldsymbol{x} \in \mathcal{S}$. First note that $\lambda_i^{\max}$ is well defined under Assumptions 2 and 3. Let $\boldsymbol{V}_i^n$ be the set of the locations of the samples taken from θ_i up to time n. Under Assumption 1, eq. (4) reads

$$\mathrm{Var}^n[\theta_i(\boldsymbol{x})] = \tau_i^2 - k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i(\boldsymbol{V}_i^n)]^{-1} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}),$$

where $\boldsymbol{V}_i^n = \{\boldsymbol{v}^0, \ldots, \boldsymbol{v}^{n-1}\}$ due to the assumption that $a^0 = \cdots = a^{n-1} = i$.

For notational simplicity, let $\boldsymbol{A} := k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i(\boldsymbol{V}_i^n)$ and $\boldsymbol{B} := k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i^{\max}\boldsymbol{I}$. Note that $\boldsymbol{B} - \boldsymbol{A} = \lambda_i^{\max}\boldsymbol{I} - \lambda_i(\boldsymbol{V}_i^n)$ is a diagonal matrix with nonnegative elements, so it is positive semi-definite. Since $\boldsymbol{A}$ and $\boldsymbol{B}$ are both positive definite, by Horn and Johnson (2012, Corollary 7.7.4), $\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1}$ is positive semi-definite. Therefore,

$$k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i(\boldsymbol{V}_i^n)]^{-1} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}) - k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i^{\max}\boldsymbol{I}]^{-1} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x})$$
$$= k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)(\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1}) k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}) \geq 0. \tag{29}$$

It then follows from eqs. (4) and (29) that

$$\mathrm{Var}^n[\theta_i(\boldsymbol{x})] \leq \tau_i^2 - k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i^{\max}\boldsymbol{I}]^{-1} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}).$$

Thus, it suffices to prove that

$$f(\boldsymbol{v}^0, \ldots, \boldsymbol{v}^{n-1}) := k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i^{\max}\boldsymbol{I}]^{-1} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}) \geq \frac{n \min_{\boldsymbol{x}' \in \mathcal{S}} \left[k_i^0(\boldsymbol{x}, \boldsymbol{x}')\right]^2}{n\tau_i^2 + \lambda_i^{\max}}, \tag{30}$$

for all $\boldsymbol{v}^0, \ldots, \boldsymbol{v}^{n-1} \in \mathcal{S}$.

Since $k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n)$ is symmetric, we can always write $k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) = \boldsymbol{Q}\mathrm{diag}\{\alpha_1, \ldots, \alpha_n\}\boldsymbol{Q}^\mathsf{T}$, where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0$ are the eigenvalues of $k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n)$, and $\boldsymbol{Q}$ is an orthogonal matrix, i.e., $\boldsymbol{Q}\boldsymbol{Q}^\mathsf{T} = \boldsymbol{I}$. Therefore,

$$k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i^{\max}\boldsymbol{I} = \boldsymbol{Q}\mathrm{diag}\{(\alpha_1 + \lambda_i^{\max}), \ldots, (\alpha_n + \lambda_i^{\max})\}\boldsymbol{Q}^\mathsf{T}.$$

and

$$[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i^{\max}\boldsymbol{I}]^{-1} = \boldsymbol{Q}\mathrm{diag}\{(\alpha_1 + \lambda_i^{\max})^{-1}, \ldots, (\alpha_n + \lambda_i^{\max})^{-1}\}\boldsymbol{Q}^\mathsf{T}.$$

If we let β_j be the j-th element of the row vector $k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)\boldsymbol{Q}$, i.e., $k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n)\boldsymbol{Q} = [\beta_1, \ldots, \beta_n]$, then

$$f(\boldsymbol{v}^0, \ldots, \boldsymbol{v}^{n-1}) = \frac{\beta_1^2}{\alpha_1 + \lambda_i^{\max}} + \cdots + \frac{\beta_n^2}{\alpha_n + \lambda_i^{\max}}.$$

Here, α_j and β_j clearly both depend on $\boldsymbol{v}^0, \ldots, \boldsymbol{v}^{n-1}$, for $j = 1, \ldots, n$. Moreover, they satisfy the following two conditions. First, $\sum_{j=1}^n \alpha_j = \mathrm{tr}(k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n)) = n\tau_i^2$, where the first equality is a straightforward fact that the trace of a matrix equals the sum of its eigenvalues, and the second equality is from Assumption 1.

Second,

$$\sum_{j=1}^{n} \beta_j^2 = k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n) \boldsymbol{Q}\boldsymbol{Q}^{\mathsf{T}} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}) = k_i^0(\boldsymbol{x}, \boldsymbol{V}_i^n) k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}) = \sum_{\ell=0}^{n-1} [k_i^0(\boldsymbol{x}, \boldsymbol{v}^\ell)]^2 \geq n \min_{\boldsymbol{x}' \in \mathcal{S}} [k_i^0(\boldsymbol{x}, \boldsymbol{x}')]^2.$$

If we define $g : \mathbb{R}^{2n} \mapsto \mathbb{R}$ as follows

$$g(a_1, \dots, a_n, b_1, \dots, b_n) := \frac{b_1}{a_1 + \lambda_i^{\max}} + \cdots + \frac{b_n}{a_n + \lambda_i^{\max}},$$

then $f(\boldsymbol{v}^0, \dots, \boldsymbol{v}^{n-1}) = g(\alpha_1, \dots, \alpha_n, \beta_1^2, \dots, \beta_n^2)$. It follows that

$$\min_{\boldsymbol{v}^0, \dots, \boldsymbol{v}^{n-1} \in \mathcal{S}} f(\boldsymbol{v}^0, \dots, \boldsymbol{v}^{n-1}) \geq \min_{\substack{(a_1, \dots, a_n) \in \mathcal{C}_1 \\ (b_1, \dots, b_n) \in \mathcal{C}_2}} g(a_1, \dots, a_n, b_1, \dots, b_n), \tag{31}$$

where

$$\mathcal{C}_1 := \left\{ (a_1, \dots, a_n) \in \mathbb{R}^n : a_1 \geq \cdots \geq a_n \geq 0 \quad \text{and} \quad \sum_{j=1}^{n} a_j = n\tau_i^2 \right\},$$

$$\mathcal{C}_2 := \left\{ (b_1, \dots, b_n) \in \mathbb{R}^n : b_1 \geq 0, \dots, b_n \geq 0 \quad \text{and} \quad \sum_{j=1}^{n} b_j \geq n \min_{\boldsymbol{x}' \in \mathcal{S}} [k_i^0(\boldsymbol{x}, \boldsymbol{x}')]^2 \right\}.$$

The reason for the inequality in eq. (31) is that the two minimization problems have the same objective function while the one in left-hand side has smaller feasible region.

We now solve the minimization problem on the right-hand side of eq. (31). Note that for any $(a_1, \dots, a_n) \in \mathcal{C}_1$, $\min_{(b_1, \dots, b_n) \in \mathcal{C}_2} g(a_1, \dots, a_n, b_1, \dots, b_n)$ is a linear programming problem, and it is easy to see that its optimal solution is $b_1^* = n \min_{\boldsymbol{x}' \in \mathcal{S}} [k_i^0(\boldsymbol{x}, \boldsymbol{x}')]^2$ and $b_2^* = \cdots = b_n^* = 0$. Hence,

$$\min_{\substack{(a_1, \dots, a_n) \in \mathcal{C}_1 \\ (b_1, \dots, b_n) \in \mathcal{C}_2}} g(a_1, \dots, a_n, b_1, \dots, b_n) = \min_{(a_1, \dots, a_n) \in \mathcal{C}_1} \frac{n \min_{\boldsymbol{x}' \in \mathcal{S}} [k_i^0(\boldsymbol{x}, \boldsymbol{x}')]^2}{a_1 + \lambda_i^{\max}} = \frac{n \min_{\boldsymbol{x}' \in \mathcal{S}} [k_i^0(\boldsymbol{x}, \boldsymbol{x}')]^2}{n\tau_i^2 + \lambda_i^{\max}}.$$

$$\tag{32}$$

Then, we can apply eqs. (31) and (32) to show eq. (30), completing the proof of Lemma 5.   □

Lemma 6. *Fix $i = 1, \dots, M$ and a sampling policy π. Suppose that the sequence of sampling locations $\{\boldsymbol{v}^n : a^n = i, n = 0, 1, \dots\}$ under π has an accumulation point $\boldsymbol{x}_i^{\mathrm{acc}}$. If Assumptions 1-3 hold, then for any $\epsilon > 0$,*

$$\limsup_{n \to \infty} \max_{\boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)} \mathrm{Var}^{\pi, n}[\theta_i(\boldsymbol{x})] \leq \tau_i^2 [1 - \rho_i^2(2\epsilon \mathbf{1})],$$

where $\mathbf{1}$ is the vector of all ones with size $d \times 1$.

Proof of Lemma 6. It follows from eq. (8) that $\{\mathrm{Var}^{\pi, n}[\theta_i(\boldsymbol{x})]\}_{n=0}^{\infty}$ is a non-increasing sequence bounded below by zero. Hence, $\mathrm{Var}^{\pi, n}[\theta_i(\boldsymbol{x})]$ converges as $n \to \infty$ and its limit is well defined.

Fix $\epsilon > 0$. Let $s_{i,n} := \left| \{ \boldsymbol{v}^\ell \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon) : a^\ell = i, \ell = 0, \dots, n-1 \} \right|$ be the number of times that al-

ternative i is sampled at a point in $\mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$ under π among the total n samples. Then, we must have $s_{i,n} \to \infty$ since $\boldsymbol{x}_i^{\mathrm{acc}}$ is an accumulation point. Note that reordering the sampling decision-observation pairs $((a^0, \boldsymbol{v}^0), y^1), \ldots, ((a^{n-1}, \boldsymbol{v}^{n-1}), y^n)$ does not alter the conditional variance of $\theta_i(\boldsymbol{x})$. Hence, we may assume without loss of generality that the first $s_{i,n}$ samples are all taken from alternative i at locations that belong to $\mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$. Since the posterior variance decreases in the number of samples by eq. (8), we conclude that for all $\boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$,

$$\mathrm{Var}^{\pi,n}[\theta_i(\boldsymbol{x})] \leq \mathrm{Var}^{\pi,s_{i,n}}[\theta_i(\boldsymbol{x})] \leq \tau_i^2 - \frac{s_{i,n} \min_{\boldsymbol{x}' \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)}[k_i^0(\boldsymbol{x}, \boldsymbol{x}')]^2}{s_{i,n}\tau_i^2 + \lambda_i^{\max}}, \tag{33}$$

where the second inequality follows from Lemma 5.

Note that $[k^0(\boldsymbol{x}, \boldsymbol{x}')]^2 = \tau_i^4 \rho_i^2(|\boldsymbol{x} - \boldsymbol{x}'|)$, and that $\|\boldsymbol{x} - \boldsymbol{x}'\| \leq \|\boldsymbol{x} - \boldsymbol{x}_i^{\mathrm{acc}}\| + \|\boldsymbol{x}_i^{\mathrm{acc}} - \boldsymbol{x}'\| \leq 2\epsilon$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$. Hence, each component of $|\boldsymbol{x} - \boldsymbol{x}'|$ is no greater than 2ϵ. Since $\rho_i(\boldsymbol{\delta})$ is decreasing in $\boldsymbol{\delta}$ component-wise for $\boldsymbol{\delta} > 0$ (see Assumption 1), $\rho_i(|\boldsymbol{x} - \boldsymbol{x}'|) \geq \rho_i(2\epsilon\mathbf{1})$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)$. It then follows from eq. (33) that

$$\max_{\boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}_i^{\mathrm{acc}}, \epsilon)} \mathrm{Var}^{\pi,n}[\theta_i(\boldsymbol{x})] \leq \tau_i^2 - \frac{s_{i,n}\tau_i^4 \rho_i^2(2\epsilon\mathbf{1})}{s_{i,n}\tau_i^2 + \lambda_i^{\max}}.$$

Sending $n \to \infty$ completes the proof.   □

Lemma 7. *Fix $i = 1, \ldots, M$. If Assumptions 1–3 hold, then for all $\boldsymbol{x} \in \mathcal{X}$,*

$$\mathrm{IKG}^n(i, \boldsymbol{x}) \leq \sqrt{\frac{\tau_i^2 \mathrm{Var}^n[\theta_i(\boldsymbol{x})]}{2\pi \lambda_i(\boldsymbol{x}) c_i^2(\boldsymbol{x})}}.$$

Proof of Lemma 7. Notice that

$$\int_{\mathcal{X}} \mathbb{E}\left[\max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v}) \,\middle|\, \mathcal{F}^n, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \gamma(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}$$

$$= \int_{\mathcal{X}} \mathbb{E}^n\left[\max_{1 \leq a \leq M} \left(\mu_a^n(\boldsymbol{v}) + \sigma_a^n(\boldsymbol{v}, \boldsymbol{v}^n) Z^{n+1}\right) \,\middle|\, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \gamma(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}$$

$$\leq \int_{\mathcal{X}} \max_{1 \leq a \leq M} \mu_a^n(\boldsymbol{v}) \gamma(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} + \int_{\mathcal{X}} \mathbb{E}^n\left[\max_{1 \leq a \leq M} \left(\sigma_a^n(\boldsymbol{v}, \boldsymbol{v}^n) Z^{n+1}\right) \,\middle|\, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \gamma(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}. \tag{34}$$

Since $k_i^0(\boldsymbol{x}, \boldsymbol{x}')$ is a continuous function by Assumption 1, it follows from Assumption 3 and the updating eq. (3) that $\mu_i^n(\boldsymbol{x})$ is a continuous function for any n. Hence, $\mu_i^n(\boldsymbol{x})$ is bounded on $\mathcal{X}$ by Assumption 2. This implies that the first integral in eq. (34) is finite and can be subtracted from both sides of the inequality. Then, by the definition eq. (12),

$$\mathrm{IKG}^n(i, \boldsymbol{x}) \leq \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \mathbb{E}^n\left[\max_{1 \leq a \leq M} \left(\sigma_a^n(\boldsymbol{v}, \boldsymbol{v}^n) Z^{n+1}\right) \,\middle|\, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \gamma(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} := I. \tag{35}$$

It follows from eq. (7) that

$$I = \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \mathbb{E}^n\left[\max\left\{\sigma_{a^n}^n(\boldsymbol{v}, \boldsymbol{v}^n) Z^{n+1}, 0\right\} \,\middle|\, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \gamma(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}$$

$$= \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \mathbb{E}^n \left[\max \left\{ |\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| Z^{n+1}, 0 \right\} \right] \gamma(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}$$

$$= \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \mathbb{E}^n \left[|\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| Z^{n+1} \mathbb{I}_{\{|\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| Z^{n+1} > 0\}} \right] \gamma(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}$$

$$= \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} |\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| \left[\int_0^{\infty} z \phi(z) \, \mathrm{d}z \right] \gamma(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}$$

$$= \frac{1}{\sqrt{2\pi} c_i(\boldsymbol{x})} \int_{\mathcal{X}} |\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| \gamma(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}. \tag{36}$$

Moreover, by eq. (7),

$$|\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})| = \left| \frac{\mathrm{Cov}^n[\theta_i(\boldsymbol{v}), \theta_i(\boldsymbol{x})]}{\sqrt{\mathrm{Var}^n[\theta_i(\boldsymbol{x})] + \lambda_i(\boldsymbol{x})}} \right| \leq \sqrt{\frac{\mathrm{Var}^n[\theta_i(\boldsymbol{v})] \, \mathrm{Var}^n[\theta_i(\boldsymbol{x})]}{\mathrm{Var}^n[\theta_i(\boldsymbol{x})] + \lambda_i(\boldsymbol{x})}} \leq \sqrt{\frac{\tau_i^2 \, \mathrm{Var}^n[\theta_i(\boldsymbol{x})]}{\lambda_i(\boldsymbol{x})}}, \tag{37}$$

where the last inequality follows because $0 \leq \mathrm{Var}^n[\theta_i(\boldsymbol{v})] \leq \mathrm{Var}[\theta_i(\boldsymbol{v})] = \tau_i^2$ for all $\boldsymbol{v} \in \mathcal{X}$ by eq. (8). The proof is completed by combining eqs. (35)–(37). □

Lemma 8. *Fix $i = 1, \ldots, M$. If Assumptions 1–3 hold, and $\eta_i^{\infty} = \infty$ under the IKG policy, then for any $\boldsymbol{x} \in \mathcal{X}$,*

$$\liminf_{n \to \infty} \mathrm{IKG}^n(i, \boldsymbol{x}) = 0.$$

Proof of Lemma 8. Since $\mathcal{X}$ is compact by Assumption 2, the sequence $\{\boldsymbol{v}^n \in \mathcal{X} : a^n = i, n = 0, 1, \ldots, \}$ is bounded, and it is of length $\eta_i^{\infty} = \infty$. Hence, it has an accumulation point $\boldsymbol{x}_i^{\mathrm{acc}}$. Let $\{\ell_{i,n}\}_{n=0}^{\infty}$ be the subsequence of $\{n : a^n = i, n = 0, 1, \ldots\}$ such that $\ell_{i,n} \to \infty$ and $\boldsymbol{v}^{\ell_{i,n}} \to \boldsymbol{x}_i^{\mathrm{acc}}$ as $n \to \infty$. Fix $\epsilon > 0$. Then, by Lemma 6,

$$\limsup_{n \to \infty} \mathrm{Var}^n[\theta_i(\boldsymbol{v}^{\ell_{i,n}})] \leq \tau_i^2 [1 - \rho_i^2(2\epsilon \mathbf{1})].$$

It then follows from Lemma 7 that

$$\limsup_{n \to \infty} \mathrm{IKG}^{\ell_{i,n}}(i, \boldsymbol{v}^{\ell_{i,n}}) \leq \limsup_{n \to \infty} \sqrt{\frac{\tau_i^2 \, \mathrm{Var}^{\ell_{i,n}}[\theta_i(\boldsymbol{v}^{\ell_{i,n}})]}{2\pi \lambda_i(\boldsymbol{x}) c_i^2(\boldsymbol{x})}} \leq \sqrt{\frac{\tau_i^4 [1 - \rho_i^2(2\epsilon \mathbf{1})]}{2\pi \lambda_i(\boldsymbol{x}) c_i^2(\boldsymbol{x})}}.$$

By sending $\epsilon \to 0$, we have $\rho_i(2\epsilon \mathbf{1}) \to 1$ and thus, $\limsup_{n \to \infty} \mathrm{IKG}^{\ell_{i,n}}(i, \boldsymbol{v}^{\ell_{i,n}}) \leq 0$. Since the limit inferior of a sequence is no greater than that of its subsequence,

$$\liminf_{n \to \infty} \mathrm{IKG}^n(i, \boldsymbol{v}^n) \leq \liminf_{n \to \infty} \mathrm{IKG}^{\ell_{i,n}}(i, \boldsymbol{v}^{\ell_{i,n}}) \leq \limsup_{n \to \infty} \mathrm{IKG}^{\ell_{i,n}}(i, \boldsymbol{v}^{\ell_{i,n}}) \leq 0. \tag{38}$$

Moreover, by the definition of IKG eq. (12) and Jensen's inequality,

$$\mathrm{IKG}^n(i, \boldsymbol{x}) \geq \frac{1}{c_i(\boldsymbol{x})} \int_{\mathcal{X}} \left\{ \max_{1 \leq a \leq M} \mathbb{E} \left[\mu_a^{n+1}(\boldsymbol{v}) \, \Big| \, \mathcal{F}^n, a^n = i, \boldsymbol{v}^n = \boldsymbol{x} \right] - \max_{1 \leq a \leq M} \mu_a^n(\boldsymbol{v}) \right\} \gamma(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} = 0,$$

for each $i = 1, \ldots, M$ and $\boldsymbol{x} \in \mathcal{X}$, where the equality follows immediately from the updating eq. (5). This, in conjunction with eq. (38), implies that $\liminf_{n \to \infty} \mathrm{IKG}^n(i, \boldsymbol{v}^n) = 0$. By the definition of the sampling

location $\boldsymbol{v}^n$ in eq. (13), $\mathrm{IKG}^n(i, \boldsymbol{v}^n) = \max_{\boldsymbol{x} \in \mathcal{X}} \mathrm{IKG}^n(i, \boldsymbol{x})$. Hence, for any $\boldsymbol{x} \in \mathcal{X}$,

$$0 \leq \liminf_{n \to \infty} \mathrm{IKG}^n(i, \boldsymbol{x}) \leq \liminf_{n \to \infty} \mathrm{IKG}^n(i, \boldsymbol{v}^n) = 0, \tag{39}$$

which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Before we prove Proposition 2, we need one more technical result about the almost sure convergence of μ_i^n, which is stated in the following Lemma 9. A similar result is also given in Bect et al. (2019, Proposition 2.9), and the proof there directly applies here.

Lemma 9. *If Assumptions 1-3 hold, then for all $i = 1, \ldots, M$, $\mu_i^n(\boldsymbol{x})$ converges to $\mu_i^\infty(\boldsymbol{x}) := \mathbb{E}[\theta_i(\boldsymbol{x})|\mathcal{F}^\infty]$ uniformly in $\boldsymbol{x} \in \mathcal{X}$ a.s. as $n \to \infty$. That is,*

$$\mathbb{P}\left\{\omega : \sup_{\boldsymbol{x} \in \mathcal{X}} |\mu_i^n(\boldsymbol{x}; \omega) - \mu_i^\infty(\boldsymbol{x}; \omega)| \to 0\right\} = 1.$$

Proof of Lemma 9. Fix $i = 1, \ldots, M$. Note that θ_i is a Gaussian process under the prior. It follows from Assumptions 1 and 3 and Theorem 1.4.1 of Adler and Taylor (2007) that the sample paths of θ_i are continuous a.s.. The proof is then completed by directly following the arguments in the proof of Proposition 2.9 in Bect et al. (2019). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We are now ready to prove Proposition 2.

Proof of Proposition 2. Let $\mu^n := (\mu_1^n, \ldots, \mu_M^n)$ denote the posterior mean of $(\theta_1, \ldots, \theta_M)$ conditioned on $\mathcal{F}^n$. Let ω denote a generic sample path. Fix $i = 1, \ldots, M$. Define

$$\Omega_0 = \{\omega : \eta_i^\infty(\omega) = \infty, \ \mu^n(\boldsymbol{x}; \omega) \to \mu^\infty(\boldsymbol{x}; \omega) \text{ uniformly in } \boldsymbol{x} \in \mathcal{X} \text{ as } n \to \infty\}.$$

Then, $\mathbb{P}(\Omega_0) = 1$ by the assumption of Proposition 2 and Lemma 9. Fix an arbitrary $\boldsymbol{x} \in \mathcal{X}$. We now prove that, under the IKG policy,

$$k_i^\infty(\boldsymbol{x}, \boldsymbol{x}; \omega) = 0, \text{ for any } \omega \in \Omega_0, \tag{40}$$

which establishes Proposition 2. We prove eq. (40) by contraction and assume that there exists some $\omega_0 \in \Omega_0$ such that $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}; \omega) > 0$. In the remaining proof, we suppress the sample path ω_0 to simplify notation.

It follows from the continuity of $k_i^0(\boldsymbol{x}, \cdot)$ assumed in Assumption 1 and the updating eq. (4) that $k_i^n(\boldsymbol{x}, \cdot)$ is continuous. The uniform convergence of $k_i^n(\boldsymbol{x}, \cdot)$ by Proposition 1 then implies that $k_i^\infty(\boldsymbol{x}, \cdot)$ is also continuous. Hence, there exist $\epsilon > 0$ such that $\min_{\boldsymbol{v} \in \mathcal{B}(\boldsymbol{x}, \epsilon)} k_i^\infty(\boldsymbol{x}, \boldsymbol{v}) > 0$. The uniform convergence of $k_i^n(\boldsymbol{x}, \cdot)$ further implies that there exists $\delta > 0$ such that $k_i^n(\boldsymbol{x}, \boldsymbol{v}) \geq \delta$ for all $\boldsymbol{v} \in \mathcal{B}(\boldsymbol{x}, \epsilon)$ and $n \geq 1$. By eq. (7),

$$\inf_{\boldsymbol{v} \in \mathcal{B}(\boldsymbol{x}, \epsilon), n \geq 1} \tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x}) = [k_i^n(\boldsymbol{x}, \boldsymbol{x}) + \lambda_i(\boldsymbol{x})]^{-1/2} \inf_{\boldsymbol{v} \in \mathcal{B}(\boldsymbol{x}, \epsilon), n \geq 1} k_i^n(\boldsymbol{v}, \boldsymbol{x})$$

$$\geq \delta[k_i^0(\boldsymbol{x}, \boldsymbol{x}) + \lambda_i(\boldsymbol{x})]^{-1/2} := \alpha_1 > 0.$$

Let $g(s,t) := t\phi(s/t) - s\Phi(-s/t)$; see Lemma 2 for properties of $g(s,t)$, including positivity and monotonicity. Then,

$$|\tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})|\phi\left(\left|\frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})}\right|\right) - |\Delta_i^n(\boldsymbol{v})|\Phi\left(-\left|\frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})}\right|\right) = g(|\Delta_i^n(\boldsymbol{v})|, |\tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})|) \geq 0,$$

for all $\boldsymbol{v} \in \mathcal{X}$. Consequently, Lemma 1 implies that

$$\text{IKG}^n(i,\boldsymbol{x}) \geq \frac{1}{c_i(\boldsymbol{x})}\int_{\mathcal{B}(\boldsymbol{x},\epsilon)} g(|\Delta_i^n(\boldsymbol{v})|, |\tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})|)\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} \geq \frac{1}{c_i(\boldsymbol{x})}\int_{\mathcal{B}(\boldsymbol{x},\epsilon)} g(|\Delta_i^n(\boldsymbol{v})|, \alpha_1)\gamma(\boldsymbol{v})\mathrm{d}\boldsymbol{v},$$

for all $n \geq 1$, where the second inequality holds because $g(s,t)$ is strictly increasing in $t \in (0,\infty)$. Note that $\liminf_{n\to\infty} \text{IKG}^n(i,\boldsymbol{x}) = 0$ by Lemma 8. Hence,

$$0 \geq \liminf_{n\to\infty} \frac{1}{c_i(\boldsymbol{x})}\int_{\mathcal{B}(\boldsymbol{x},\epsilon)} g(|\Delta_i^n(\boldsymbol{v})|, \alpha_1)\gamma(\boldsymbol{v})\mathrm{d}\boldsymbol{v} \geq \frac{1}{c_i(\boldsymbol{x})}\int_{\mathcal{B}(\boldsymbol{x},\epsilon)} \liminf_{n\to\infty} g(|\Delta_i^n(\boldsymbol{v})|, \alpha_1)\gamma(\boldsymbol{v})\mathrm{d}\boldsymbol{v}, \quad (41)$$

where the second inequality holds due to Fatou's lemma. Furthermore, since for any $\boldsymbol{v} \in \mathcal{X}$, $\Delta_i^n(\boldsymbol{v}) = \mu_i^n(\boldsymbol{v}) - \max_{a\neq i}\mu_a^n(\boldsymbol{v})$, and $\mu_a^n(\boldsymbol{v}) \to \mu_a^\infty(\boldsymbol{v})$ for $a = 1,\ldots,M$, where $|\mu_a^\infty(\boldsymbol{v})| < \infty$. then

$$\limsup_{n\to\infty}|\Delta_i^n(\boldsymbol{v})| \leq \limsup_{n\to\infty}[2\max_a|\mu_a^n(\boldsymbol{v})|] = 2\max_a|\mu_a^\infty(\boldsymbol{v})| := \alpha_2(\boldsymbol{v}) < \infty,$$

for all $\boldsymbol{v} \in \mathcal{B}(\boldsymbol{x},\epsilon)$. Then, in the light of eq. (41) and the fact that $g(s,t)$ is strictly decreasing in $s \in [0,\infty)$,

$$0 \geq \frac{1}{c_i(\boldsymbol{x})}\int_{\mathcal{B}(\boldsymbol{x},\epsilon)} \liminf_{n\to\infty} g(|\Delta_i^n(\boldsymbol{v})|, \alpha_1)\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} \geq \frac{1}{c_i(\boldsymbol{x})}\int_{\mathcal{B}(\boldsymbol{x},\epsilon)} g(\alpha_2(\boldsymbol{v}), \alpha_1)\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}.$$

This contracts the fact that $g(s,t) > 0$ for all $s \in [0,\infty)$ and $t \in (0,\infty)$. Therefore, eq. (40) is proved. □

D. Proof of Proposition 3

Let $S^n := (\mu_1^n,\ldots,\mu_M^n, k_1^n,\ldots,k_M^n)$ denote the state at time n, which fully determines the posterior distribution of $(\theta_1,\ldots,\theta_M)$ conditioned on $\mathcal{F}^n$. The state transition $S^n \to S^{n+1}$ is governed by eqs. (5) and (6), which is determined by the sampling decision $(a^n, \boldsymbol{v}^n)$.

Let $s := (\mu_1,\ldots,\mu_M, k_1,\ldots,k_M) \in \mathbb{S}$ be a generic state and $\mathbb{S}$ denote the set of states for which μ_i is a continuous function and k_i is a continuous covariance function for each $i = 1,\ldots,M$. For $s \in \mathbb{S}$, define

$$V(s) := \int_{\mathcal{X}} \max_{1\leq a\leq M} \mu_a(\boldsymbol{v})\gamma(\boldsymbol{v})\mathrm{d}\boldsymbol{v},$$

and

$$Q(s,i,\boldsymbol{x}) := \mathbb{E}[V(S^{n+1}) \mid S^n = s, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}].$$

By the following Lemma 10, it is easy to see that at time n, the IKG policy (13) chooses

$$\underset{1 \leq i \leq M, \boldsymbol{x} \in \mathcal{X}}{\operatorname{argmax}} \ [c_i(\boldsymbol{x})]^{-1}[Q(S^n, i, \boldsymbol{x}) - V(S^n)]. \tag{42}$$

Lemma 10. *Fix $s \in \mathbb{S}$, $i = 1, \ldots, M$, and $\boldsymbol{x} \in \mathcal{X}$ where $\mathcal{X}$ is compact,*

$$Q(s, i, \boldsymbol{x}) = \int_{\mathcal{X}} \mathbb{E}\left[\max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v}) \,\Big|\, S^n = s, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \gamma(\boldsymbol{v}) \mathrm{d}\boldsymbol{v}.$$

Proof of Lemma 10. Notice that by the updating eq. (5), given $S^n = s$, $a^n = i$, $\boldsymbol{v}^n = \boldsymbol{x}$ and Z^{n+1},

$$\max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v}) = \max\left\{\mu_i(\boldsymbol{v}) + \tilde{\sigma}_i(\boldsymbol{v}, \boldsymbol{x})Z^{n+1}, \max_{a \neq i} \mu_a(\boldsymbol{v})\right\}.$$

Let $f\left(s, i, \boldsymbol{x}, \boldsymbol{v}, Z^{n+1}\right) = \max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v}) - \max_{a \neq i} \mu_a(\boldsymbol{v})$. Then $f\left(s, i, \boldsymbol{x}, \boldsymbol{v}, Z^{n+1}\right) \geq 0$, for all $\boldsymbol{v} \in \mathcal{X}$ and Z^{n+1}. Hence,

$$\begin{aligned} Q(s, i, \boldsymbol{x}) &= \mathbb{E}\left[\int_{\mathcal{X}} \max_{1 \leq a \leq M} \mu_a^{n+1}(\boldsymbol{v})\gamma(\boldsymbol{v})\mathrm{d}\boldsymbol{v} \,\Big|\, S^n = s, a^n = i, \boldsymbol{v}^n = \boldsymbol{x}\right] \\ &= \mathbb{E}\left[\int_{\mathcal{X}} \left(f\left(s, i, \boldsymbol{x}, \boldsymbol{v}, Z^{n+1}\right) + \max_{a \neq i} \mu_a(\boldsymbol{v})\right)\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}\right] \\ &= \mathbb{E}\left[\int_{\mathcal{X}} f\left(s, i, \boldsymbol{x}, \boldsymbol{v}, Z^{n+1}\right)\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}\right] + \int_{\mathcal{X}} \max_{a \neq i} \mu_a(\boldsymbol{v})\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} \\ &= \int_{\mathcal{X}} \mathbb{E}\left[f\left(s, i, \boldsymbol{x}, \boldsymbol{v}, Z^{n+1}\right)\right]\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} + \int_{\mathcal{X}} \max_{a \neq i} \mu_a(\boldsymbol{v})\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}\,, \end{aligned}$$

where the interchange of integral and expectation is justified by Tonelli's theorem for nonnegative functions, and $\int_{\mathcal{X}} \max_{a \neq i} \mu_a(\boldsymbol{v})\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}$ is finite since $\mu_i(\boldsymbol{v})$ is continuous on the compact set $\mathcal{X}$ for $i = 1, \ldots, M$. Thus the result in Lemma 10 follows immediately. $\square$

Lemma 11. *Fix $s \in \mathbb{S}$, $i = 1, \ldots, M$, and $\boldsymbol{x} \in \mathcal{X}$ where $\mathcal{X}$ is compact and $\alpha(\cdot) > 0$ on $\mathcal{X}$. Then, $Q(s, i, \boldsymbol{x}) \geq V(s)$ and the equality holds if and only if $k_i(\boldsymbol{x}, \boldsymbol{x}) = 0$.*

Proof of Lemma 11. Applying Lemma 10 and the updating eq. (5),

$$\begin{aligned} Q(s, i, \boldsymbol{x}) &= \int_{\mathcal{X}} \mathbb{E}\left[\max\left\{\mu_i(\boldsymbol{v}) + \tilde{\sigma}_i(\boldsymbol{v}, \boldsymbol{x})Z^{n+1}, \max_{a \neq i} \mu_a(\boldsymbol{v})\right\}\right]\gamma(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} \\ &\geq \int_{\mathcal{X}} \max\left\{\mathbb{E}\left[\mu_i(\boldsymbol{v}) + \tilde{\sigma}_i(\boldsymbol{v}, \boldsymbol{x})Z^{n+1}\right], \max_{a \neq i} \mu_a(\boldsymbol{v})\right\}\gamma(\boldsymbol{v})\mathrm{d}\boldsymbol{v} \tag{43} \\ &= \int_{\mathcal{X}} \max_{1 \leq a \leq M} \mu_a(\boldsymbol{v})\gamma(\boldsymbol{v})\mathrm{d}\boldsymbol{v} = V(s), \end{aligned}$$

where eq. (43) follows from Jensen's inequality since $\max(\cdot, \cdot)$ is a strictly convex function.

If $k_i(\boldsymbol{x}, \boldsymbol{x}) = 0$, then in the light of the fact that k_i is a covariance function, we must have that

$$|k_i(\boldsymbol{v}, \boldsymbol{x})| \leq \sqrt{k_i(\boldsymbol{v}, \boldsymbol{v})k_i(\boldsymbol{x}, \boldsymbol{x})} = 0,$$

so $k_i(\boldsymbol{v}, \boldsymbol{x}) = 0$ for all $\boldsymbol{v} \in \mathcal{X}$. Hence, $\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x}) = 0$ by eq. (7), so $\mu_a^{n+1}(\boldsymbol{v}) = \mu_a^n(\boldsymbol{v})$ for all $a = 1, \ldots, M$ and $\boldsymbol{v} \in \mathcal{X}$. Hence, $\mu_a^{n+1}(\boldsymbol{v})$ is deterministic given S^n for all $a = 1, \ldots, M$ and $\boldsymbol{v} \in \mathcal{X}$. Thus, the inequality eq. (43) holds with equality.

Next, assume conversely that $Q(s, i, \boldsymbol{x}) = V(s)$. If $k_i(\boldsymbol{x}, \boldsymbol{x}) \neq 0$, then the continuity of k_i implies that $k_i(\boldsymbol{v}, \boldsymbol{x}) \neq 0$ for all $\boldsymbol{v} \in \tilde{\mathcal{X}}$, where $\tilde{\mathcal{X}} \subset \mathcal{X}$ is an open neighborhood of $\boldsymbol{x}$. Without loss of generality, we assume that for all $\boldsymbol{v} \in \tilde{\mathcal{X}}$, $k_i(\boldsymbol{v}, \boldsymbol{x}) > 0$ and thus, $\tilde{\sigma}_i(\boldsymbol{v}, \boldsymbol{x}) = k_i(\boldsymbol{v}, \boldsymbol{x})/\sqrt{k_i(\boldsymbol{x}, \boldsymbol{x}) + \lambda_i(\boldsymbol{x})} > 0$. By the strict convexity of $\max(\cdot, \cdot)$ and Jensen's inequality,

$$\mathbb{E}\left[\max\left\{\mu_i(\boldsymbol{v}) + \tilde{\sigma}_i(\boldsymbol{v}, \boldsymbol{x})Z^{n+1}, \max_{a \neq i} \mu_a(\boldsymbol{v})\right\}\right] > \max\left\{\mathbb{E}\left[\mu_i(\boldsymbol{v}) + \tilde{\sigma}_i(\boldsymbol{v}, \boldsymbol{x})Z^{n+1}\right], \max_{a \neq i} \mu_a(\boldsymbol{v})\right\},$$

for $\boldsymbol{v} \in \tilde{\mathcal{X}}$. Hence, eq. (43) becomes a strict inequality since $\gamma(\boldsymbol{v}) > 0$ for all $\boldsymbol{v} \in \mathcal{X}$. This contradicts $Q(s, i, \boldsymbol{x}) = V(s)$, so $k_i(\boldsymbol{x}, \boldsymbol{x}) = 0$. $\qquad\square$

Lemma 12. *Fix $i = 1, \ldots, M$. If Assumptions 1 and 3 hold, and $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}) = 0$ for some $\boldsymbol{x} \in \mathcal{X}$, then $\eta_i^\infty = \infty$.*

Proof of Lemma 12. We prove by contradiction and assume that $\eta_i^\infty < \infty$. Then, $N_i := \min\{n : \eta_i^n = \eta_i^\infty\} < \infty$ and $a^n \neq i$ for all $n \geq N_i$. Due to the mutual independence between the alternatives, it follows that the posterior distribution of θ_i remains the same for $n \geq N_i$. In particular, $k_i^n(\boldsymbol{x}, \boldsymbol{x}) = k_i^{N_i}(\boldsymbol{x}, \boldsymbol{x})$ for all $n > N_i$. Hence, $k_i^{N_i}(\boldsymbol{x}, \boldsymbol{x}) = k_i^\infty(\boldsymbol{x}, \boldsymbol{x}) = 0$. It follows from eq. (6) that

$$k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{x}) = k_i^{N_i}(\boldsymbol{x}, \boldsymbol{x}) + \left[\sigma_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{v}^{N_i-1})\right]^2 = \left[\sigma_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{v}^{N_i-1})\right]^2.$$

By the definition of N_i, $a^{N_i-1} = i$. Then by eq. (7),

$$k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{x}) = \frac{\left[k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{v}^{N_i-1})\right]^2}{k_i^{N_i-1}(\boldsymbol{v}^{N_i-1}, \boldsymbol{v}^{N_i-1}) + \lambda_i(\boldsymbol{v}^{N_i-1})}. \tag{44}$$

Notice that

$$\begin{aligned}
[k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{v}^{N_i-1})]^2 &= \left\{\text{Cov}^{N_i-1}\left[\theta_i(\boldsymbol{x}), \theta_i(\boldsymbol{v}^{N_i-1})\right]\right\}^2 \\
&\leq \text{Var}^{N_i-1}[\theta_i(\boldsymbol{x})]\,\text{Var}^{N_i-1}[\theta_i(\boldsymbol{v}^{N_i-1})] \\
&= k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{x})k_i^{N_i-1}(\boldsymbol{v}^{N_i-1}, \boldsymbol{v}^{N_i-1}). \tag{45}
\end{aligned}$$

It follows from eqs. (44) and (45) that $\lambda_i(\boldsymbol{v}^{N_i-1})k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{x}) \leq 0$. Thus, $k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{x}) = 0$, since $k_i^{N_i-1}(\boldsymbol{x}, \boldsymbol{x}) \geq 0$ and $\lambda_i(\boldsymbol{v}^{N_i-1}) > 0$ in Assumption 3. By induction, we can conclude that $k_i^0(\boldsymbol{x}, \boldsymbol{x}) = 0$, which contracts the fact that $k_i^0(\boldsymbol{x}, \boldsymbol{x}) = \tau_i^2 > 0$ in Assumption 1. Therefore, we must have $\eta_i^\infty = \infty$. $\qquad\square$

We are now ready to prove Proposition 3.

Proof of Proposition 3. Define $\Omega_1 := \{\omega : S^n(\omega) \to S^\infty(\omega) \text{ pointwise as } n \to \infty\}$. By Lemma 9 and Proposition 1, $\mathbb{P}(\Omega_1) = 1$. For any $i = 1, \ldots, M$, define the event $\mathsf{H}_i := \{\omega : \eta_i^\infty(\omega) < \infty\}$. Then, $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}; \omega) > 0$ for all $\boldsymbol{x} \in \mathcal{X}$ and $\omega \in \mathsf{H}_i$ by Lemma 12. On the other hand, Proposition 2 implies that $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}; \omega) = 0$ for all $\boldsymbol{x} \in \mathcal{X}$ and $\omega \in \mathsf{H}_i^{\mathsf{c}} \cap \Omega_1$, where $\mathsf{H}_i^{\mathsf{c}}$ is the complement of H_i. Thus, by Lemma 11,

$$Q(S^\infty(\omega), i, \boldsymbol{x}) > V(S^\infty(\omega)), \quad \text{for all } \omega \in \mathsf{H}_i \cap \Omega_1,$$
$$Q(S^\infty(\omega), i, \boldsymbol{x}) = V(S^\infty(\omega)), \quad \text{for all } \omega \in \mathsf{H}_i^{\mathsf{c}} \cap \Omega_1. \tag{46}$$

Further, for any subset $A \subseteq \{1, \ldots, M\}$, define the event

$$\mathsf{H}_A := \{\cap_{i \in A} \mathsf{H}_i\} \cap \{\cap_{i \notin A} \mathsf{H}_i^{\mathsf{c}}\}.$$

Choose any $A \neq \emptyset$. When $A = \{1, \ldots, M\}$, $\mathsf{H}_A = \emptyset$, because it is impossible that all alternative have finite samples while $n \to \infty$. So $\mathsf{H}_A \cap \Omega_1 = \emptyset$. When $A \neq \{1, \ldots, M\}$, we prove $\mathsf{H}_A \cap \Omega_1 = \emptyset$ by contradiction. Suppose that $\mathsf{H}_A \cap \Omega_1 \neq \emptyset$ so that we can choose and fix a sample path $\omega_0 \in \mathsf{H}_A \cap \Omega_1$. Then, $\eta_i^\infty(\omega_0) < \infty$ for all $i \in A$. Hence, there exists $T_i(\omega_0) < \infty$ for all $i \in A$ such that the IKG policy does not choose alternative i for $n > T_i(\omega_0)$. Let $T(\omega_0) := \max_{i \in A} T_i(\omega_0)$. Then, $T(\omega_0) < \infty$ and the IKG policy does not choose $i \in A$ for $n > T(\omega_0)$. On the other hand, it follows from eq. (46) that for all $i \in A$, $i' \notin A$, and $\boldsymbol{x} \in \mathcal{X}$,

$$Q(S^\infty(\omega_0), i, \boldsymbol{x}) - V(S^\infty(\omega_0)) > Q(S^\infty(\omega_0), i', \boldsymbol{x}) - V(S^\infty(\omega_0)) = 0.$$

Let $Q^\dagger(s, i, \boldsymbol{x}) := Q(s, i, \boldsymbol{x}) - V(s)$ for simplicity. Then, by virtue of the compactness of $\mathcal{X}$ and the positivity of $c_i(\boldsymbol{x})$,

$$\max_{\boldsymbol{x} \in \mathcal{X}} [c_i(\boldsymbol{x})]^{-1} Q^\dagger(S^\infty(\omega_0), i, \boldsymbol{x}) > \max_{\boldsymbol{x} \in \mathcal{X}} [c_{i'}(\boldsymbol{x})]^{-1} Q^\dagger(S^\infty(\omega_0), i', \boldsymbol{x}) = 0,$$

for all $i \in A$ and $i' \notin A$. Hence,

$$\min_{i \in A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_i(\boldsymbol{x})]^{-1} Q^\dagger(S^\infty(\omega_0), i, \boldsymbol{x}) > \max_{i' \notin A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_{i'}(\boldsymbol{x})]^{-1} Q^\dagger(S^\infty(\omega_0), i', \boldsymbol{x}) = 0. \tag{47}$$

Notice that $S^n(\omega_0) \to S^\infty(\omega_0)$ pointwise as $n \to \infty$ since $\omega_0 \in \Omega_1$. Hence, there exists a finite number $\tilde{n}(\omega_0) > T(\omega_0)$ such that

$$\min_{i \in A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_i(\boldsymbol{x})]^{-1} Q^\dagger(S^{\tilde{n}(\omega_0)}(\omega_0), i, \boldsymbol{x}) > \max_{i' \notin A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_{i'}(\boldsymbol{x})]^{-1} Q^\dagger(S^{\tilde{n}(\omega_0)}(\omega_0), i', \boldsymbol{x}),$$

which implies that IKG policy must choose alternative $i \in A$ at time $\tilde{n}(\omega_0)$ by eq. (42). This contradicts the definition of $T(\omega_0)$. Therefore, the event $\mathsf{H}_A \cap \Omega_1$ must be empty for any nonempty $A \subseteq \{1, \ldots, M\}$.

It then follows immediately that $\mathbb{P}(\mathsf{H}_A) = 0$ for any nonempty $A \subseteq \{1, \ldots, M\}$, since $\mathbb{P}(\Omega_1) = 1$. Notice that the whole sample space $\Omega = \cup_{A \subseteq \{1, \ldots, M\}} \mathsf{H}_A$. Hence,

$$1 = \mathbb{P}(\mathsf{H}_\emptyset) = \mathbb{P}\left(\cap_{i=1}^M \mathsf{H}_i^{\mathsf{c}}\right) = \mathbb{P}\left(\{\omega : \eta_i^\infty = \infty \text{ for all } i = 1, \ldots, M\}\right),$$

which completes the proof.   □

E. Proof of Theorem 2

Part (i) is an immediate consequence of Propositions 2 and 3. The other two parts follow closely the proof
of similar results in Theorem 1 of Xie et al. (2016).

For part (ii), fix an arbitrary $\boldsymbol{x} \in \mathcal{X}$. Note that for each $i = 1, \ldots, M$,

$$\mathbb{E}[(\mu_i^n(\boldsymbol{x}) - \theta_i(\boldsymbol{x}))^2] = \mathbb{E}[\mathbb{E}^n[(\mu_i^n(\boldsymbol{x}) - \theta_i(\boldsymbol{x}))^2]] = \mathbb{E}[k_i^n(\boldsymbol{x}, \boldsymbol{x})] \to \mathbb{E}[k_i^\infty(\boldsymbol{x}, \boldsymbol{x})] = 0,$$

as $n \to \infty$, where the convergence holds due to the fact that $0 \le k_i^n(\boldsymbol{x}, \boldsymbol{x}) \le k_i^0(\boldsymbol{x}, \boldsymbol{x})$ from eq. (8) and the
dominated convergence theorem. This asserts that $\mu_i^n(\boldsymbol{x}) \to \theta_i(\boldsymbol{x})$ in L^2. By Lemma 9, $\mu_i^n(\boldsymbol{x}) \to \mu_i^\infty(\boldsymbol{x})$
a.s., which implies that $\theta_i(\boldsymbol{x}) = \mu_i^\infty(\boldsymbol{x})$ a.s., due to the a.s. uniqueness of convergence in probability. Thus,
$\mu_i^n(\boldsymbol{x}) \to \theta_i(\boldsymbol{x})$ a.s. as $n \to \infty$.

For part (iii), let us again fix $\boldsymbol{x} \in \mathcal{X}$. Let $i^*(\boldsymbol{x}) \in \operatorname{argmax}_i \theta_i(\boldsymbol{x})$. We now show that $\operatorname{argmax}_i \mu_i^n(\boldsymbol{x}) \to$
$i^*(\boldsymbol{x})$ a.s. as $n \to \infty$. Again, we let ω denote a generic sample path and use notations like $i^*(\boldsymbol{x}; \omega)$ to
emphasize the dependence on ω. Let $\epsilon(\boldsymbol{x}; \omega) \coloneqq \theta_{i^*(\boldsymbol{x},\omega)}(\boldsymbol{x}; \omega) - \max_{a \ne i^*(\boldsymbol{x},\omega)} \theta_a(\boldsymbol{x}; \omega)$. Then, $\mathbb{P}(\{\omega :$
$\epsilon(\boldsymbol{x}; \omega) > 0\}) = 1$ because $(\theta_1(\boldsymbol{x}; \omega), \ldots, \theta_M(\boldsymbol{x}; \omega))$ is a realization of a multivariate normal random vari-
able under the prior distribution. Hence, the event $\tilde{\Omega} \coloneqq \{\omega : \epsilon(\boldsymbol{x}; \omega) > 0 \text{ and } \mu_i^n(\boldsymbol{x}; \omega) \to \theta_i(\boldsymbol{x}; \omega) \text{ for all}$
$i = 1, \ldots, M\}$ occurs with probability 1. Fix an arbitrary $\tilde{\omega} \in \tilde{\Omega}$. To complete the proof, it suffices to show
that $\operatorname{argmax}_i \mu_i^n(\boldsymbol{x}; \tilde{\omega}) \to i^*(\boldsymbol{x}; \tilde{\omega})$ as $n \to \infty$.

Clearly, there exists $N(\tilde{\omega}) < \infty$ such that $|\mu_i^n(\boldsymbol{x}; \tilde{\omega}) - \theta_i(\boldsymbol{x}; \tilde{\omega})| < \epsilon(\boldsymbol{x}; \tilde{\omega})/2$ for all $n > N(\tilde{\omega})$ and
$i = 1, \ldots, M$. Hence, for all $i \ne i^*(\boldsymbol{x}; \tilde{\omega})$ and $n \ge N(\tilde{\omega})$,

$$\mu_{i^*(\boldsymbol{x};\tilde{\omega})}^n(\boldsymbol{x}; \tilde{\omega}) > \theta_{i^*(\boldsymbol{x};\tilde{\omega})}(\boldsymbol{x}; \tilde{\omega}) - \frac{\epsilon(\boldsymbol{x}; \tilde{\omega})}{2} \ge \theta_i(\boldsymbol{x}; \tilde{\omega}) + \frac{\epsilon(\boldsymbol{x}; \tilde{\omega})}{2} > \mu_i^n(\boldsymbol{x}; \tilde{\omega}).$$

This implies that $i^*(\boldsymbol{x}; \tilde{\omega}) = \operatorname{argmax}_i \mu_i^n(\boldsymbol{x}; \tilde{\omega})$ for all $n > N(\tilde{\omega})$, and thus $\operatorname{argmax}_i \mu_i^n(\boldsymbol{x}; \tilde{\omega}) \to i^*(\boldsymbol{x}; \tilde{\omega})$
as $n \to \infty$.   □

F. Proof of Theorem 3

The steps to prove Theorem 3 are exactly the same to those for Theorem 2, and we only need to modify the
arguments related to the actual sampling decision (which is $(a^n, \boldsymbol{v}^n)$ satisfies eq. (13) in the IKG policy, and
$(\tilde{a}^n, \tilde{\boldsymbol{v}}^n)$ satisfies eq. (15) in the quasi-IKG policy). So, we will not repeat the entire proofs, but only point
out the modification briefly. Specifically, the two main steps to prove Theorem 3 are summarized as the
following Propositions 4 and 5, which are parallel to Propositions 2 and 3.

Proposition 4. *Fix $i = 1, \ldots, M$. If Assumptions 1-3 hold and $\eta_i^\infty = \infty$ a.s., then for any $\boldsymbol{x} \in \mathcal{X}$, $k_i^\infty(\boldsymbol{x}, \boldsymbol{x}) = 0$
a.s. under the quasi-IKG policy.*

Proof of Proposition 4. All the intermediate lemmas for Proposition 2 directly apply to Proposition 4, ex-
cept for Lemma 8. Instead, we now need to show that under the quasi-IKG policy, for any $\boldsymbol{x} \in \mathcal{X}$,

$\liminf_{n\to\infty} \mathrm{IKG}^n(i, \boldsymbol{x}) = 0$. We first observe that, for the sequence $\{\tilde{\boldsymbol{v}}^n \in \mathcal{X} : \tilde{a}^n = i, n = 0, 1, \ldots, \}$, we can still show $\liminf_{n\to\infty} \mathrm{IKG}^n(i, \tilde{\boldsymbol{v}}^n) = 0$ with the same arguments. Then, by the definition of the quasi-IKG policy (15),

$$\mathrm{IKG}^n(i, \tilde{\boldsymbol{v}}^n) = \mathrm{IKG}^n(\tilde{a}^n, \tilde{\boldsymbol{v}}^n) \geq \mathrm{IKG}^n(a^n, \boldsymbol{v}^n) - \varepsilon_n \geq \mathrm{IKG}^n(i, \boldsymbol{v}^n) - \varepsilon_n.$$

On the other hand, since $\mathrm{IKG}^n(i, \boldsymbol{v}^n) = \max_{\boldsymbol{x} \in \mathcal{X}} \mathrm{IKG}^n(i, \boldsymbol{x})$, then for any $\boldsymbol{x} \in \mathcal{X}$, eq. (39) is replaced by

$$0 \leq \liminf_{n\to\infty} \mathrm{IKG}^n(i, \boldsymbol{x}) \leq \liminf_{n\to\infty} \mathrm{IKG}^n(i, \boldsymbol{v}^n) \leq \liminf_{n\to\infty} [\mathrm{IKG}^n(i, \tilde{\boldsymbol{v}}^n) + \varepsilon_n] = 0,$$

where the equality is due to $\liminf_{n\to\infty} \mathrm{IKG}^n(i, \tilde{\boldsymbol{v}}^n) = 0$ and $\varepsilon_n \to 0$. Finally, the proof of Proposition 4 follows the similar arguments as in the proof of Proposition 2. $\qquad \square$

Proposition 5. *If Assumptions 1–3 hold, then $\eta_i^\infty = \infty$ a.s. for each $i = 1, \ldots, M$ under the quasi-IKG policy.*

Proof of Proposition 5. All the intermediate lemmas for Proposition 3 directly apply to Proposition 5. We then proceed by following the same arguments as in the proof Proposition 3, with $T_i(\omega_0)$ meaning that the quasi-IKG policy does not choose alternative i for $n > T_i(\omega_0)$. After we obtain eq. (47), we now want to show that quasi-IKG policy must choose alternative $i \in A$ at some time $\tilde{n}(\omega_0) > T(\omega_0)$, which leads to the contradiction.

Due to eq. (47), there exists some small $\Delta > 0$ such that

$$\min_{i \in A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_i(\boldsymbol{x})]^{-1} Q^\dagger(S^\infty(\omega_0), i, \boldsymbol{x}) - \Delta > \max_{i' \notin A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_{i'}(\boldsymbol{x})]^{-1} Q^\dagger(S^\infty(\omega_0), i', \boldsymbol{x}) = 0.$$

Notice that $S^n(\omega_0) \to S^\infty(\omega_0)$ pointwise as $n \to \infty$ since $\omega_0 \in \Omega_1$. Hence, there exists a finite number $\tilde{n}_1(\omega_0) > T(\omega_0)$ such that

$$\min_{i \in A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_i(\boldsymbol{x})]^{-1} Q^\dagger(S^n, i, \boldsymbol{x}) - \Delta/2 > \max_{i' \notin A} \max_{\boldsymbol{x} \in \mathcal{X}} [c_{i'}(\boldsymbol{x})]^{-1} Q^\dagger(S^n, i', \boldsymbol{x}), \tag{48}$$

for all $n \geq \tilde{n}_1(\omega_0)$. Since $\varepsilon_n \to 0$ as $n \to \infty$, there exists a finite number $\tilde{n}_2(\omega_0) > T(\omega_0)$ such that $\varepsilon_n < \Delta/2$ for all $n \geq \tilde{n}_2(\omega_0)$. Then, we can conclude that at time $\tilde{n}(\omega_0) := \max(\tilde{n}_1(\omega_0), \tilde{n}_2(\omega_0)) > T(\omega_0)$, quasi-IKG policy must choose alternative $i \in A$. Otherwise, for $n = \tilde{n}(\omega_0)$, if $\tilde{a}^n \notin A$, then by eqs. (42) and (48),

$$\mathrm{IKG}^n(\tilde{a}^n, \tilde{\boldsymbol{v}}^n) < \mathrm{IKG}^n(a^n, \boldsymbol{v}^n) - \Delta/2 < \mathrm{IKG}^n(a^n, \boldsymbol{v}^n) - \varepsilon_n,$$

which violates the definition of quasi-IKG policy defined in eq. (15). $\qquad \square$

G. Gradient Calculation

It is easy to see that

$$g_i^n(\boldsymbol{v}, \boldsymbol{x}) = \frac{\partial[h_i^n(\boldsymbol{v}, \boldsymbol{x})/c_i(\boldsymbol{x})]}{\partial \boldsymbol{x}} = \frac{\frac{\partial h_i^n(\boldsymbol{v}, \boldsymbol{x})}{\partial \boldsymbol{x}} c_i(\boldsymbol{x}) - h_i^n(\boldsymbol{v}, \boldsymbol{x}) \frac{\mathrm{d}c_i(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}}{[c_i(\boldsymbol{x})]^2},$$

$$
\frac{\partial h_i^n(\boldsymbol{v}, \boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{cases} \phi\left(\left|\frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})}\right|\right) \frac{\partial \tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})}{\partial \boldsymbol{x}}, & \text{if } \tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x}) > 0, \\ 0, & \text{if } \tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x}) = 0, \\ -\phi\left(\left|\frac{\Delta_i^n(\boldsymbol{v})}{\tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})}\right|\right) \frac{\partial \tilde{\sigma}_i^n(\boldsymbol{v},\boldsymbol{x})}{\partial \boldsymbol{x}}, & \text{if } \tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x}) < 0, \end{cases}
$$

and that, by the definition of $\tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})$ in eq. (7) and Assumptions 1 and 3,

$$
\frac{\partial \tilde{\sigma}_i^n(\boldsymbol{v}, \boldsymbol{x})}{\partial \boldsymbol{x}} = [k_i^n(\boldsymbol{x}, \boldsymbol{x}) + \lambda_i(\boldsymbol{x})]^{-\frac{1}{2}} \frac{\partial k_i^n(\boldsymbol{v}, \boldsymbol{x})}{\partial \boldsymbol{x}} - \frac{[k_i^n(\boldsymbol{x}, \boldsymbol{x}) + \lambda_i(\boldsymbol{x})]^{-\frac{3}{2}} k_i^n(\boldsymbol{v}, \boldsymbol{x})}{2} \left[\frac{\mathrm{d}k_i^n(\boldsymbol{x}, \boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} + \frac{\mathrm{d}\lambda_i(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\right],
$$
(49)

provided that the prior correlation function ρ_i and the cost function c_i are both differentiable. Assuming ρ_i to be differentiable excludes some covariance functions that satisfy Assumption 1 such as the Matérn(ν) type with $\nu = 1/2$, but many others including both the Matérn(ν) type with $\nu > 1$ and the SE type do have the desired differentiability. We next calculate analytically the derivatives of $k_i^n(\cdot, \cdot)$ in eq. (49) for several common covariance functions. The calculation is a routine exercise so we omit the details.

Throughout the subsequent Examples 1–3, we use the following notation. For $i = 1, \dots, M$,

$$
\boldsymbol{\alpha}_i := (\alpha_{i,1}, \dots, \alpha_{i,d})^\mathsf{T} \quad \text{and} \quad r_i(\boldsymbol{x}, \boldsymbol{x}') := \sqrt{\sum_{j=1}^d \alpha_{i,j}(x_j - x'_j)^2}.
$$

Recall that $\boldsymbol{V}_i^n$ denotes the set of locations of the samples taken from θ_i up to time n. With slight abuse of notation, here we treat $\boldsymbol{V}_i^n$ as a matrix wherein the columns are corresponding to the points in the set and arranged in the order of appearance. Moreover, for notational simplicity, let $\boldsymbol{V}_i^n := (\boldsymbol{v}_1, \dots, \boldsymbol{v}_{m_i^n})$, where m_i^n is the number of columns of $\boldsymbol{V}_i^n$. Let $\boldsymbol{X}$ be a matrix with the same dimension as $\boldsymbol{V}_i^n$ and all columns are identically $\boldsymbol{x}$. We adopt the denominator layout for matrix calculus. For the following Examples 1–3, it can be shown that

$$
\frac{\partial k_i^n(\boldsymbol{v}, \boldsymbol{x})}{\partial \boldsymbol{x}} = \mathrm{diag}(\boldsymbol{\alpha}_i)(\boldsymbol{x} - \boldsymbol{v})a_0 - \boldsymbol{A} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{v}),
$$
(50)

$$
\frac{\mathrm{d}k_i^n(\boldsymbol{x}, \boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = -2\boldsymbol{A} k_i^0(\boldsymbol{V}_i^n, \boldsymbol{x}),
$$
(51)

where

$$
\boldsymbol{A} := \mathrm{diag}(\boldsymbol{\alpha}_i)(\boldsymbol{X} - \boldsymbol{V}_i^n)\mathrm{diag}\{a_1, \dots, a_{m_i^n}\}[k_i^0(\boldsymbol{V}_i^n, \boldsymbol{V}_i^n) + \lambda_i \boldsymbol{I}]^{-1},
$$

while the values of $a_0, a_1, \dots, a_{m_i^n}$ depend on the choice of the covariance function.

Example 1 (SE). Let $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = \tau_i^2 \exp\left(-r_i^2(\boldsymbol{x}, \boldsymbol{x}')\right)$. Then, in eqs. (50) and (51), $a_0 := -2k_i^0(\boldsymbol{v}, \boldsymbol{x})$ and $a_\ell := -2k_i^0(\boldsymbol{v}_\ell, \boldsymbol{x})$, for $\ell = 1, \dots, m_i^n$.

Example 2 (Matérn(3/2)). Let $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = \tau_i^2\left(1 + \sqrt{3}r_i(\boldsymbol{x}, \boldsymbol{x}')\right) \exp\left(-\sqrt{3}r_i(\boldsymbol{x}, \boldsymbol{x}')\right)$. Then, in eqs. (50) and (51),

$$
a_\ell := \sqrt{3}r_i^{-1}(\boldsymbol{v}_\ell, \boldsymbol{x})\left[\tau_i^2 \exp\left(-\sqrt{3}r_i(\boldsymbol{v}_\ell, \boldsymbol{x})\right) - k_i^0(\boldsymbol{v}_\ell, \boldsymbol{x})\right],
$$

for $\ell = 0, 1, \dots, m_i^n$ and $\boldsymbol{v}_0 = \boldsymbol{v}$.

Example 3 (Matérn(5/2))**.** Let $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = \tau_i^2 \big(1 + \sqrt{5}r_i(\boldsymbol{x}, \boldsymbol{x}') + \frac{5}{3}r_i^2(\boldsymbol{x}, \boldsymbol{x}')\big) \exp\big(-\sqrt{5}r_i(\boldsymbol{x}, \boldsymbol{x}')\big)$. Then, in eqs. (50) and (51),

$$a_\ell := \Big(\sqrt{5}r_i^{-1}(\boldsymbol{v}_\ell, \boldsymbol{x}) + \tfrac{10}{3}\Big)\tau_i^2 \exp\Big(-\sqrt{5}r_i(\boldsymbol{v}_\ell, \boldsymbol{x})\Big) - \sqrt{5}r_i^{-1}(\boldsymbol{v}_\ell, \boldsymbol{x})k_i^0(\boldsymbol{v}_\ell, \boldsymbol{x}),$$

for $\ell = 0, 1, \ldots, m_i^n$ and $\boldsymbol{v}_0 = \boldsymbol{v}$.

H. Implementation Issues

A plain-vanilla implementation of $\widehat{\mathrm{IKG}}^n(i, \boldsymbol{x})$ in eq. (18) may encounter rounding errors, since $h_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})$ may be rounded to zero when evaluated via eq. (16); see Frazier et al. (2009) for discussion on a similar issue. To enhance numerical stability, we first evaluate the logarithm of the summand and then do exponentiation. For notational simplicity, we set

$$u_j := |\Delta_i^n(\boldsymbol{\xi}_j)/\tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})| \quad \text{and} \quad h_i^n(\boldsymbol{\xi}_j, \boldsymbol{x}) = |\tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})|[\phi(u_j) - u_j\Phi(-u_j)].$$

If $\tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x}) \neq 0$, we compute

$$g_j := \log\left(\frac{h_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})}{J}\right) = \log\left(\frac{|\tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})|}{\sqrt{2\pi}J}\right) - \frac{1}{2}u_j^2 + \log\left(1 - u_j\frac{\Phi(-u_j)}{\phi(u_j)}\right),$$

where $\Phi(-u_j)/\phi(u_j)$ is known as the Mills ratio, and can be asymptotically approximated by $u_j/(u_j^2+1)$ for large u_j. Moreover, $\log(1 + x)$ can be accurately computed by `log1p` function available in most numerical software packages. At last, we compute

$$\log\widehat{\mathrm{IKG}}^n(i, \boldsymbol{x}) = \log\sum_{j \in \mathcal{J}} e^{g_j} - \log(c_i(\boldsymbol{x})) = g^* + \log\sum_{j \in \mathcal{J}} e^{g_j - g^*} - \log(c_i(\boldsymbol{x})),$$

where $\mathcal{J} := \{j : \tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x}) \neq 0, j = 1, \ldots, J\}$, and $g^* = \max_{j \in \mathcal{J}} g_j$; we set $\log\widehat{\mathrm{IKG}}_\gamma^n(i, \boldsymbol{x}) = -\infty$ if $\mathcal{J}$ is empty. The above procedure is summarized in Algorithm 1.

In the implementation of SGA, we adopt two well-known modifications.

(i) We use *mini-batch* SGA to have more productive iterations. Specifically, in each iteration eq. (19), instead of using a single $g_i^n(\boldsymbol{\xi}_k, \boldsymbol{x}_k)$ as the gradient estimate, we use the average of m independent estimates $g_i^n(\boldsymbol{\xi}_{k1}, \boldsymbol{x}_k), \ldots, g_i^n(\boldsymbol{\xi}_{km}, \boldsymbol{x}_k)$, which is denoted as $\bar{g}_i^n(\boldsymbol{\xi}_{k1}, \ldots, \boldsymbol{\xi}_{km}, \boldsymbol{x}_k)$.

(ii) We adopt the *Polyak-Ruppert averaging* (Polyak and Juditsky 1992) to mitigate of the algorithm's sensitivity on the choice of the step size. Specifically, when K iterations are completed, we report $\frac{1}{K+2-K_0}\sum_{k=K_0}^{K+1} \boldsymbol{x}_k$, instead of $\boldsymbol{x}_{K+1}$, as the approximated solution of $\boldsymbol{v}_i^n$, where $1 \leq K_0 \leq K$ is a pre-specified integer.

Upon computing $\hat{\boldsymbol{v}}_i^n \approx \mathrm{argmax}_{\boldsymbol{x}} \mathrm{IKG}^n(i, \boldsymbol{x})$ with SGA for each i, we set

$$\hat{a}^n = \underset{1 \leq i \leq M}{\mathrm{argmax}} \log\widehat{\mathrm{IKG}}^n(i, \hat{\boldsymbol{v}}_i^n) \quad \text{and} \quad \hat{\boldsymbol{v}}^n = \hat{\boldsymbol{v}}_{\hat{a}^n}^n,$$

Algorithm 1 Computing $\log \widehat{\mathrm{IKG}}^n(i, \boldsymbol{x})$.

Inputs: $\mu_1^n, \ldots, \mu_M^n, k_1^n, \ldots, k_M^n, \lambda_1, \ldots, \lambda_M, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_J, i, \boldsymbol{x}, c_i(\boldsymbol{x})$
Outputs: log_IKG
1:   $\mathcal{J} \leftarrow \emptyset, log\_IKG \leftarrow -\infty$
2:   **for** $j = 1$ to J **do**
3:      **if** $|\tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})| > 0$ **then**
4:         $u \leftarrow |\Delta_i^n(\boldsymbol{\xi}_j)| / |\tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})|$
5:         **if** $u < 20$ **then**
6:            $r \leftarrow \Phi(-u)/\phi(u)$
7:         **else**
8:            $r \leftarrow u/(u^2 + 1)$
9:         **end if**
10:        $g_j \leftarrow \log\left(\frac{|\tilde{\sigma}_i^n(\boldsymbol{\xi}_j, \boldsymbol{x})|}{\sqrt{2\pi J}} \right) - \frac{1}{2}u^2 + \texttt{log1p}(-ur)$ $\triangleright$ $\texttt{log1p}(x) = \log(1 + x)$.
11:        $\mathcal{J} \leftarrow \{\mathcal{J}, j\}$
12:      **end if**
13: **end for**
14: **if** $\mathcal{J} \neq \emptyset$ **then**
15:      $g^* \leftarrow \max_{j \in \mathcal{J}} g_j$
16:      $log\_IKG \leftarrow g^* + \log \sum_{j \in \mathcal{J}} e^{g_j - g^*} - \log(c_i(\boldsymbol{x}))$
17: **end if**

Algorithm 2 Approximately Computing $(a^n, \boldsymbol{v}^n)$ Using SGA.

Inputs: $\mu_1^n, \ldots, \mu_M^n, k_1^n, \ldots, k_M^n, \lambda_1, \ldots, \lambda_M, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_J, c_1, \ldots, c_M$
Outputs: $\hat{a}^n, \hat{\boldsymbol{v}}^n$
1:   **for** $i = 1$ to M **do**
2:      $\boldsymbol{x}_1 \leftarrow$ initial value
3:      **for** $k = 1$ to K **do**
4:         Generate independent sample $\{\boldsymbol{\xi}_{k1}, \ldots, \boldsymbol{\xi}_{km}\}$ from density $\gamma(\cdot)$
5:         $\boldsymbol{x}_{k+1} \leftarrow \Pi_{\mathcal{X}}[\boldsymbol{x}_k + b_k \bar{g}_i^n(\boldsymbol{\xi}_{k1}, \ldots, \boldsymbol{\xi}_{km}, \boldsymbol{x}_k)]$ $\triangleright$ Mini-batch SGA.
6:      **end for**
7:      $\hat{\boldsymbol{v}}_i^n \leftarrow \frac{1}{K+2-K_0} \sum_{k=K_0}^{K+1} \boldsymbol{x}_k$ $\triangleright$ Polyak-Ruppert averaging.
8:      $log\_IKG_i \leftarrow \log \widehat{\mathrm{IKG}}^n(i, \hat{\boldsymbol{v}}_i^n)$ $\triangleright$ Call Algorithm 1.
9:   **end for**
10: $\hat{a}^n \leftarrow \operatorname{argmax}_i log\_IKG_i$
11: $\hat{\boldsymbol{v}}^n \leftarrow \hat{\boldsymbol{v}}_{\hat{a}^n}^n$

---

to be the sampling decision at time n, i.e., let $(\hat{a}^n, \hat{\boldsymbol{v}}^n)$ be the computed solution for $(a^n, \boldsymbol{v}^n)$ under the IKG policy. The complete procedure is summarized in Algorithm 2.

I. Additional Numerical Experiments
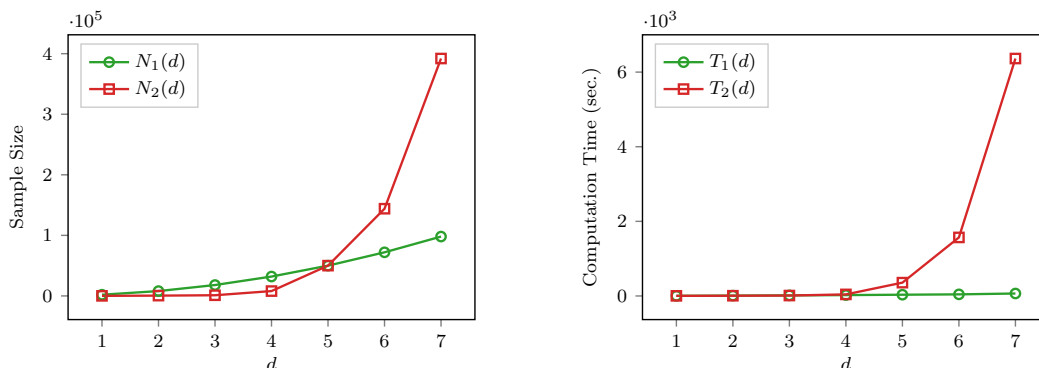
Computational cost comparison

We conduct simple experiment to compare the computational cost when the IKG sampling policy defined in eq. (13), i.e., $(a^n, \boldsymbol{v}^n) \in \operatorname{argmax}_{1 \leq i \leq M, \boldsymbol{x} \in \mathcal{X}} \mathrm{IKG}^n(i, \boldsymbol{x})$, is solved purely using sample average approximation (SAA) method or our proposed SGA (together with SAA). For IKG with SGA, here refer to as method 1, as described in Section 5, the computation of eq. (13) consists of two steps. Step (i) is to solve $\boldsymbol{v}_i^n = \max_{\boldsymbol{x} \in \mathcal{X}} \mathrm{IKG}^n(i, \boldsymbol{x})$ for all $i = 1, \ldots, M$ with SGA, and step (ii) is to solve $a^n = \operatorname{argmax}_{1 \leq i \leq M} \mathrm{IKG}^n(i, \boldsymbol{v}_i^n)$ with SAA. For IKG with pure SAA, here refer to as method 2, the problems in the above two steps are both solved with SAA. In particular, the problem in step (i) is converted into a continuous deterministic optimization after applying SAA, which is solved directly using the `fmincon` solver in MATLAB. It is expected that for either method, the computational cost will increase as the dimensionality d increases. But to evaluate the

exact value of the computational cost, one needs to know the true optimal solution of eq. (13) and control the optimality gap when a specific method is used. Here we simply compare the relative computational cost of two methods by roughly controlling the resulting OC at the same level.

The same problem in Section 5 is considered and we let the dimensionality d increase from 1 to 7. The density function of covariates is the uniform distribution and the cost function is constantly 1. All the parameters for the problem, the IKG with SGA (i.e., method 1) and the evaluation of OC are the same as before. For each d, the sampling policy under the two methods is carried out respectively until the budget $B = 100$ is exhausted, and the $\widehat{\text{OC}}(B)$ curves are obtained. For fair comparison, we let the step (ii) of method 2 be exactly the same (i.e., same sample used) as that of method 1, and tune the sample size in step (i) of method 2 as follows. We gradually increase the sample size of random covariates used in SAA (not the number of samples from alternatives), until the resulting $\widehat{\text{OC}}(B)$ curve from method 2 is roughly comparable to that from method 1. Note that for either method the computation time needs to solve eq. (13) depends on the number of samples allocated to each alternative so far, and will increase as the samples accumulate. So, we report the total computation time spent on solving eq. (13) during the entire sampling process (until the budget $B = 100$ is exhausted, which means eq. (13) is solved for 100 times), averaged on $L = 30$ replications, for the two methods, which are denoted as $T_1(d)$ and $T_2(d)$ respectively.

The following Figure 4 shows the comparison between methods 1 and 2. Left panel of Figure 4 shows the sample sizes of random covariates used to approximately solve $\boldsymbol{v}_i^n = \max_{\boldsymbol{x} \in \mathcal{X}} \text{IKG}^n(i, \boldsymbol{x})$ for each i in step (i) by the two methods, which are denoted as $N_1(d)$ and $N_2(d)$ respectively. Note that $N_1(d) = mK = 20d \times 100d = 2000d^2$ as specified, and $N_2(d)$ is tuned so that the performance of method 2 matches that of method 1. Right panel of Figure 4 shows the computation times $T_1(d)$ and $T_2(d)$ (in MATLAB, Windows 10 OS, 3.60 GHz CPU, 16 GB RAM). It can be seen that IKG with SGA (i.e., method 1) scales much better in dimensionality d than IKG with pure SAA (i.e., method 2). Recall that for each method, the sample size of covariates in step (ii) is set as $500d^2$. So, even consider the fact that in method 2 the function approximation in step (i) can be directly used in step (ii), which saves the sample of covariates and the relevant computation in step (ii), the entire sample size and the computation time of method 2 still grows much faster than method 1.

Figure 4: Computational cost comparison between SGA and SAA.

Estimated sampling variance

In practice, the sampling variance $\lambda_i(\boldsymbol{x})$ is usually unknown and needs to be estimated. We suggest to follow the approach in Ankenman et al. (2010). Specifically, for each alternative i, at some predetermined design points $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^m$, multiple simulations are run and the sample variances are computed, which are denoted as $s_i^2(\boldsymbol{x}^1), \ldots, s_i^2(\boldsymbol{x}^m)$. Then ordinary kriging (i.e., Gaussian process interpolation) is used to fit the entire surface of $\lambda_i(\boldsymbol{x})$. Under the Bayesian viewpoint, it is equivalent to impose a Gaussian process with constant mean function $\mu_i^0(\boldsymbol{x}) \equiv \mu_i^0$ and covariance function $k_i^0(\boldsymbol{x}, \boldsymbol{x}')$ as prior of $\lambda_i(\boldsymbol{x})$, and compute the posterior mean function by ignoring the sampling variance at the design points, i.e.,

$$\mu_i^m(\boldsymbol{x}) = \mu_i^0 + k_i^0(\boldsymbol{x}, \boldsymbol{X}_i) k_i^0(\boldsymbol{X}_i, \boldsymbol{X}_i)^{-1}[\boldsymbol{y}_i - \mu_i^0 \boldsymbol{I}],$$

where $\boldsymbol{X}_i := (\boldsymbol{x}^1, \ldots, \boldsymbol{x}^m)$ and $\boldsymbol{y}_i := (s_i^2(\boldsymbol{x}^1), \ldots, s_i^2(\boldsymbol{x}^m))^\mathsf{T}$. Then, $\mu_i^m(\boldsymbol{x})$ is used as estimate of $\lambda_i(\boldsymbol{x})$, and the IKG policy is applied as if $\lambda_i(\boldsymbol{x})$ was known. In ordinary kriging, μ_i^0 and the parameters in $k_i^0(\boldsymbol{x}, \boldsymbol{x}')$ are usually optimized via maximum likelihood estimation (MLE).

We again consider the problem in Section 5. To better investigate the effect of estimating $\lambda_i(\boldsymbol{x})$, we consider two sampling variance: (1) $\lambda_i(\boldsymbol{x}) \equiv 0.01$, as before; (2) $\lambda_i(\boldsymbol{x}) = 0.01 \times (1.5^{d-1} + \theta_i(\boldsymbol{x}))$. The density function of covariates is the uniform distribution and the cost function is constantly 1. All the other parameters for the problem are the same as before. The prior for estimating $\lambda_i(\boldsymbol{x})$ is directly set as $\mu_i^0 = 0$, and $k_i^0(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{1}{d}\|\boldsymbol{x} - \boldsymbol{x}'\|^2\right)$, without invoking the MLE. The design points are generated by Latin hypercube sampling and the same design points are used for each i. All the parameters for the IKG policy and the evaluation of OC are the same as before. The following Figure 5 shows the estimated opportunity cost when the sampling variance is known or estimated using the above approach, for the case of $d = 1$ or 3 and sampling variance (1) or (2). Numerical results show that the effect of estimating the sampling variance is minor for this problem, which agrees with the observation in Ankenman et al. (2010).

Figure 5: Estimated opportunity cost (vertical axis) as a function of the sampling budget (horizontal axis).