CONSISTENCY OF SEQUENTIAL BAYESIAN SAMPLING POLICIES

PETER I. FRAZIER* AND WARREN B. POWELL†

Abstract. We consider Bayesian information collection, in which a measurement policy collects information to support a future decision. This framework includes ranking and selection, continuous global optimization, and many other problems in sequential experimental design. We give a sufficient condition under which measurement policies sample each measurement type infinitely often, ensuring consistency, i.e., that a globally optimal future decision is found in the limit. This condition is useful for verifying consistency of adaptive sequential sampling policies that do not do forced random exploration, making consistency difficult to verify by other means. We demonstrate the use of this sufficient condition by showing consistency of two previously proposed ranking and selection policies: OCBA for linear loss, and the knowledge-gradient policy with independent normal priors. Consistency of the knowledge-gradient policy was shown previously, while the consistency result for OCBA is new.

Key words. Bayesian inference, Design of experiments, Sequential design, Ranking and selection

AMS subject classifications. 62F15, 62K99, 62L05, 62F07

1. Introduction. We consider the general class of sequential Bayesian information collection problems. This class of problems is distinguished by: the assumption of a prior distribution on some underlying and unknown truth; the opportunity to adaptively perform a sequence of measurements whose results are observed with noise; a single "implementation decision" made after all measurements are complete; and a loss assessed at the final time depending on this implementation decision and the underlying truth. Our goal when studying such problems is to design a measurement strategy that will best allow making a good implementation decision and, by doing so, minimize the expected loss.

Sequential Bayesian information collection problems appear in a broad collection applications, including ranking and selection for simulation optimization ([10]); global optimization ([11]); combinatorial optimization ([31]); medical diagnosis ([25]); oil exploration ([3]); computer vision ([30]); and drug dosage-response estimation ([13]). Reviews of this class of problems may be found in [15, 14].

Within the framework of sequential Bayesian information collection, a measurement strategy or policy is a rule for selecting which type of measurement to make at each point in time. This selection may depend on the data collected so far. If a measurement policy uses each measurement type infinitely often, then under mild conditions it learns the underlying truth perfectly in the large-sample limit. Moreover, in this large-sample limit, it drives the expected loss of the implementation decision made after all the measuerments are complete to zero. We use the word "consistent' to describe such measurement policies that learn the truth perfectly in the limit.

If a measurement policy is known to use each measurement type infinitely often, then conditions for consistency are well understood. In particular, if the space in which the truth lies has finitely many dimensions, as we assume here, then consistency follows immediately. Previous work on consistency has focused on those problems where it is easy to verify that each measurement type is used infinitely often. For example, when

* School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, 14853 (pf98@cornell.edu).

† Deparment of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey, 08540 (powell@princeton.edu).

there is only a single measurement type, it follows trivially that this lone measurement type is used infinitely often. Consistency in this case is very well-studied in the statistics literature: see, e.g., [12] for a very general result showing consistency on an almost sure set of truths, and a more recent review in [18]. Another example are those problems in which the policy is forced to occasionally sample at random among the measurement types. If this forced exploration is sufficiently frequent, it follows easily that each measurement type is used infinitely often. The reinforcement learning community has studied this case in the context of sampling-based methods for solving Markov decision processes: see, e.g., [2].

For many sequential sampling policies, however, it can be difficult to ascertain whether each measurement type is sampled infinitely often. In particular, when the next measurement type is chosen adaptively as a deterministic function of the (random) data collected so far, then a policy can fall into situations where it becomes stuck using a strict subset of the measurement types, thus failing to be consistent.

In this article, we provide an easy-to-check sufficient condition for consistency, or equivalently, that each measurement type is sampled infinitely often. We demonstrate the use of this sufficient condition by then showing consistency of two measurement policies for ranking and selection proposed previously in the literature: the OCBA policy for linear loss proposed in [22], and the $(R_1, \ldots, R_1)$ policy from [21]. The consistency result for OCBA for linear loss is new, while consistency for the $(R_1, \ldots, R_1)$ policy was shown previously in [16]. The new proof presented here is simpler than the one in [16], and demonstrates how the sufficient conditions can be used to more easily check consistency.

Although the class of sequential Bayesian information collection problems is broad, we stress here the assumed distinctness of measurement and implementation decisions. In particular, we assume that all rewards are collected at the final time as a function only of the implementation decision, and depend only *indirectly* on measurements. This excludes a large class of problems such as multi-armed bandit problems (see, e.g., [19, 1, 27]), in which actions provide both information and a direct cost or reward. It also excludes reinforcement learning in real environments (as opposed to simulated or laboratory environments) in which all actions provide direct costs or rewards.

As stated, most previous work on consistency of sequential sampling policies has focused on either the forced-exploration or single-measurement-type cases. While we believe this work is the first to provide *general* sufficient conditions for verifying whether a sequential sampling policy with finitely many measurement types uses each measurement type infinitely often, there is previous work showing consistency in *specific* information collection problems of specific policies that do not do forced random exploration. Within Bayesian global optimization, consistency has been considered under several prior probability distributions. It is well-known that a sufficient condition for such convergence is that the measurement points are dense in the function's domain ([32]), and much effort has been expended to show that this condition is met by specific algorithms. The algorithms considered are generally one-step optimal Bayesian algorithms such as the expected improvement (EI) algorithm and the P-algorithm (see, e.g., [29]). Consistency of the P-algorithm for Wiener process priors in one dimension has been known since [26], and convergence for a more general class of Gaussian process priors, again in one dimension, was shown in [6]. Consistency of the EI algorithm was shown for a 1-dimensional Wiener process prior in [34], and for a general class of Gaussian process priors in multiple dimensions in [33]. [28] provides

a prior under which the EI algorithm is *not consistent*, and shows how this prior may be altered to guarantee convergence.

Within the context of ranking and selection, [16] and [17] show consistency for the knowledge-gradient policy for the ranking and selection problem with linear loss and independent normally distributed rewards, with [16] showing it for the case of an independent normal prior and [17] showing it for the more general case of a multivariate normal prior. These papers use the term "asymptotic optimality" instead of "consistency" because asymptotically minimal expected loss implies that the suboptimality gap between a policy that converges to the optimum and a Bayes optimal policy shrinks to zero as the measurement budget increases to infinity. This use of the term asymptotic optimality should not be confused with asymptotically optimal *rates* of convergence, as studied for example in [20].

We begin in Sections 2 and 3 with a general description of the sequential Bayesian information collection problem and some preliminary results. Then, in Section 4, we present the main result, which is a sufficient condition for consistency of a measurement policy. In Section 5 we apply this result to a ranking and selection problem with normally distributed rewards and known variance, showing consistency of two sequential sampling policies: OCBA for linear loss from [22], and the $(R_1, \ldots, R_1)$ policy from [21] (analyzed more fully under the name "knowledge-gradient policy" in [16]). We conclude in Section 6. Proofs not included in the text may be found in the appendix.

2. Problem Description. We are interested in whether a particular sequential sampling policy induces consistency. We give an informal description of the problem here, and then give a more formal description in terms of exponential families in the following subsections.

We suppose that there is some unknown parameter θ whose identity we would like to learn by making a sequence of measurements $X_1, X_2, \ldots, X_T$ with corresponding observations $Y_1, \ldots, Y_T$. Here, each X_t is chosen from a finite set of measurement types $\mathcal{X}$, and the observation Y_t has a density $y \mapsto p(y; x, \theta)$ that depends on x and θ. The method by which we choose the next measurement type X_{t+1} given the data collected so far $(X_1, Y_1), \ldots, (X_t, Y_t)$ is called the *measurement policy*. Beginning with a prior distribution on θ, a sequence of posterior distributions Q_t on θ result from these measurements. After all the measurements are complete at time T, an implementation decision i is chosen and a loss $R(\theta; i)$ is realized. We assume that the Bayes optimal implementation decision is chosen, which has expected loss $\min_i \boldsymbol{E}_T[R(\theta; i)]$, where $\boldsymbol{E}_T$ is the expectation under the posterior at time T. We call a measurement policy *consistent* if this achieved expected loss shrinks to a minimal value as $T \to \infty$.

2.1. Sampling Model. We begin by supposing we have an observation space $\mathcal{Y} \subseteq \boldsymbol{R}^{d'}$, a parameter space $\Theta \subseteq \boldsymbol{R}^d$, and a finite set $\mathcal{X}$ of measurement types. We also adopt a probability measure Q_0 on $(\Theta, \mathcal{B}(\Theta))$, which is our Bayesian prior, quantifying our beliefs on which underlying truths are most likely. Here $\mathcal{B}(\Theta)$ denotes the Borel σ-algebra on Θ, and similarly for sets other than Θ. We use the notation $M(\Theta)$ to denote the space of probability measures on Θ.

Corresponding to each $\vartheta \in \Theta$ and $x \in \mathcal{X}$ is a probability measure $P(\,\cdot\,; x, \vartheta)$ on the observation space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ governing the likelihoods of our observations. We assume that this probability measure has a density $p(\,\cdot\,; x, \vartheta) : \mathcal{Y} \mapsto \boldsymbol{R}_+$ with respect to some σ-finite measure $\nu(\,\cdot\,; x)$ on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ given by

$$p(y; x, \vartheta) = \exp\left(\alpha(\vartheta; x)^T \gamma(y; x) - \zeta(\vartheta; x)\right), \tag{2.1}$$

where ^{T} denotes matrix transposition, $\alpha : \boldsymbol{R}^d \times \mathcal{X} \mapsto \boldsymbol{R}^{l'}$ and $\gamma : \mathcal{Y} \times \mathcal{X} \mapsto \boldsymbol{R}^{l'}$ are given functions with l' an integer, and a normalizing function $\zeta : \boldsymbol{R}^d \times \mathcal{X} \mapsto \boldsymbol{R}$ is defined by

$$\zeta(\vartheta; x) = \log \int_{\mathcal{Y}} \exp(\alpha(\vartheta; x)^T \gamma(y; x)) \nu(dy; x).$$

We further assume $\Theta = \left\{ \vartheta \in \boldsymbol{R}^d : |\zeta(\vartheta; x)| < \infty \ \forall x \right\}$. This assumption is not restrictive, since we may place all our prior's probability mass on a proper subset of Θ.

We now construct the over-arching probability space $(\Omega, \mathcal{F}, \boldsymbol{P})$ on which we define the random variables θ and $(X_t, Y_t)_{t \geq 1}$. We define a filtration $(\mathcal{F}_t)$ by letting $\mathcal{F}_t$ be the σ-algebra generated by $\{(X_{t'}, Y_{t'}) \mid t' \leq t\}$, and we define $\boldsymbol{P}$ by requiring

$$\begin{aligned} &\boldsymbol{P}\{\theta \in A\} = Q_0(A) \quad \text{for } A \in \mathcal{B}(\Theta), \\ &\boldsymbol{P}\{Y_t \in C \mid X_t, \theta\} = P(C; X_t, \theta) \quad \text{for } C \in \mathcal{B}(\mathcal{Y}). \end{aligned}$$

We also require the Y_t to be conditionally independent of the other random variables given θ and X_t, and the X_t to be conditionally independent of θ given $\mathcal{F}_t$. We let $\boldsymbol{P}_t = \boldsymbol{P}\{\cdot \mid \mathcal{F}_t\}$ and $Q_t := \boldsymbol{P}_t\{\theta \in \cdot\}$, and we define $\boldsymbol{E}_t$ to be the expectation with respect to $\boldsymbol{P}_t$. While θ is a random variable taking values in Θ, we write ϑ throughout to indicate a generic element of Θ. Similarly, X_t and Y_t are random variables while x and y are generic elements of $\mathcal{X}$ and $\mathcal{Y}$ respectively.

The distributions $\boldsymbol{P}_t\{X_{t+1} \in \cdot\}$ are chosen by the experimenter and may either be chosen to be concentrated on a single point in $\mathcal{X}$, or may be more general, allowing randomized measurement decisions. We assume that the sequential sampling policy depends only upon t and the information collected up to t on θ. That is, we assume the existence of some measurable function $\Pi : \boldsymbol{N} \times M(\Theta) \times \mathcal{B}(\mathcal{X}) \mapsto [0, 1]$ giving

$$\boldsymbol{P}_t\{X_{t+1} \in C\} = \Pi(t, Q_t, C) \text{ a.s.}$$

This choice of Π is the choice of sequential experimental design or sampling policy, and should be chosen to make the inference about θ as accurate as possible. This construction may be understood by supposing that nature chooses and fixes θ according to Q_0 sight-unseen. The experimenter then conducts a sequence of experiments indexed by t, choosing for each experiment an experiment type X_t based only on the previous experiment types $X_{t'}$ and results $Y_{t'}$ (where $t' < t$) and possibly on some external source of randomization, and then observing the result Y_t whose distribution is determined by X_t and θ. From these experiments we then infer the value of θ.

With these definitions, we have specified that the sampling distribution for each measurement type in $\mathcal{X}$ comes from an exponential family, with the functions α and γ and their dependence on the common parameter θ allowing the possibility for dependence between the sampling distributions at each measurement type, with information from samples of one measurement type allowing inference about the sampling distributions for other measurement types.

2.2. Loss Function and Consistency. The sampling model and posterior distribution from Sections 2.1 and 3.1, and the inference on θ that they support, exist together with a loss function. The loss function quantifies a sampling policy's performance, and defines an optimization problem over the space of potential sampling policies.

Let $\mathcal{I}$ be a finite set of terminal or "implementation" decisions that could be employed, and $R : \Theta \times \mathcal{I} \mapsto \boldsymbol{R}_+$ a function that specifies the (non-negative) loss $R(\theta; i)$ incurred when taking implementation decision i while the true sampling distribution is given by θ. We assume that $\vartheta \mapsto R(\vartheta; i)$ is a measurable function for each i, and that $R(\theta; i)$ is integrable under Q_0 for all $i \in \mathcal{I}$.

Given this loss function and a fixed and finite number of samples T that we may take, we define the expected loss of the sampling policy as $\boldsymbol{E}\left[\min_{i\in\mathcal{I}} \boldsymbol{E}_T\left[R(\theta; i)\right]\right]$. Our concern is with the asymptotic expected loss of a policy, and so we define the asymptotic risk of a policy as

$$\lim_{T\to\infty} \boldsymbol{E}\left[\min_{i\in\mathcal{I}} \boldsymbol{E}_T\left[R(\theta; i)\right]\right].$$

This limit exists because R is non-negative, and Jensen's inequality together with the tower property of conditional expectation show that $\boldsymbol{E}\left[\min_{i\in\mathcal{I}} \boldsymbol{E}_T\left[R(\theta; i)\right]\right]$ is nondecreasing in T. We seek conditions on the sampling policy under which this asymptotic loss is as small as possible.

The best asymptotic loss that we can achieve through sampling is to perfectly learn the sampling distribution of each measurement type $x \in \mathcal{X}$. Define $G_x := P(\,\cdot\,; x, \theta)$, which is this sampling distribution under measurement type x. With this in mind, we call the sampling policy *consistent* if

$$\lim_{T\to\infty} \boldsymbol{E}\left[\min_{i\in\mathcal{I}} \boldsymbol{E}_T\left[R(\theta; i)\right]\right] = \boldsymbol{E}\left[\min_{i} \boldsymbol{E}\left[R(\theta; i) \mid G_x, x \in \mathcal{X}\right]\right]. \tag{2.2}$$

If knowing the sampling distribution for all $x \in \mathcal{X}$ completely determines θ, i.e., if θ is identifiable, then the right hand side of (2.2) is simply $\boldsymbol{E}\left[\min_i R(\theta; i)\right]$, which is the minimal loss achievable given perfect information about θ. Thus, we may think of the consistency condition as indicating convergence to perfect knowledge and a perfect implementation decision in the limit as our sampling policy is allowed infinitely many measurements.

3. Preliminary Results. In this section, we give a few additional definitions and preliminary results based on the theory of exponential families, the value of information, and the law of large numbers. These definitions and results will be used in the main development in Section 4.

3.1. Mean-Value Paramaterization of the Posterior. In the Bayesian setting, our belief about θ adjusted by our observations up to time t determines the posterior distribution Q_t. In our exponential family setting, this posterior distribution takes a convenient form. First, for each $t \in \boldsymbol{N}$ and $x \in \mathcal{X}$ define

$$S_{tx} = \sum_{t'\le t} \mathbf{1}_{\{X_{t'}=x\}} \gamma(Y_{t'}; x), \qquad N_{tx} = \sum_{t'\le t} \mathbf{1}_{\{X_{t'}=x\}}.$$

Also define larger random vectors containing them, $S_t = [S_{tx}]_{x\in\mathcal{X}}$ and $N_t = [N_{tx}]_{x\in\mathcal{X}}$. Here, S_{tx} takes values in $\boldsymbol{R}^{l'}$, S_t in $\boldsymbol{R}^{l'|\mathcal{X}|}$, N_{tx} in $\boldsymbol{R}$, and N_t in $\boldsymbol{R}^{|\mathcal{X}|}$. It will also be useful to define for each $x \in \mathcal{X}$ a random variable $N_{\infty,x} := \sum t' \mathbf{1}_{\{X_{t'}=x\}}$ taking values in $\boldsymbol{N} \cup \{\infty\}$. With these definitions, we write the posterior density on θ as

$$\frac{dQ_t}{dQ_0}(\vartheta) = \exp\left(\sum_{x\in\mathcal{X}} \left[\alpha(\vartheta; x)^T S_{tx} - \zeta(\vartheta; x) N_{tx}\right] - \Gamma(S_t, N_t)\right),$$

where the normalizer $\Gamma : \boldsymbol{R}^{l'|\mathcal{X}|} \times \boldsymbol{R}^{|\mathcal{X}|} \mapsto \boldsymbol{R}$ is defined by

$$\Gamma(s,n) := \log \int_\Theta \exp\left(\sum_{x\in\mathcal{X}} \alpha(\vartheta;x)^T s_x - \zeta(\vartheta;x) n_x\right) Q_0(d\vartheta),$$

and has domain $\mathrm{dom}(\Gamma) = \left\{(s,n) \in \boldsymbol{R}^{l'|\mathcal{X}|} \times \boldsymbol{R}^{|\mathcal{X}|} : |\Gamma(s,n)| < \infty\right\}$. We assume that $\mathrm{dom}(\Gamma)$ is open.

This is an exponential family for the posterior distribution on θ parameterized by S_t and N_t. As written, the family has $(l'+1)|\mathcal{X}|$ parameters, but its rank l may be smaller. (See [4] for a definition of the rank of an exponential family.) We now re-parameterize the family into its minimal-rank parameterization, and from there into its mean-value parameterization. The mean-value parameterization is then used in Section 4.

To write the family in minimal-rank form, we define linear functions $\beta : \Theta \mapsto \boldsymbol{R}^l$ and $\tau : \boldsymbol{R}^{l'|\mathcal{X}|} \times \boldsymbol{R}^{|\mathcal{X}|} \mapsto \boldsymbol{R}^l$ so that

$$\sum_{x\in\mathcal{X}} \alpha(\vartheta;x)^T s_x - \zeta(\vartheta;x) n_x = \beta(\vartheta)^T \tau(s,n) \quad \text{for all } \vartheta \in \Theta, s \in \boldsymbol{R}^{l'|\mathcal{X}|}, n \in \boldsymbol{R}^{|\mathcal{X}|}.$$

If the original family was of full rank then $l = (l'+1)|\mathcal{X}|$ and $\beta(\vartheta)$ is formed by concatenating the vectors $[\alpha(\vartheta;x), -\zeta(\vartheta;x)]$, $x \in \mathcal{X}$, and $\tau(s,n)$ is formed by concatenating $[s_x, n_x]$, $x \in \mathcal{X}$. If not, and $l < (l'+1)|\mathcal{X}|$, then $\beta(\vartheta)$ contains a subset of l linearly independent rows from $[\alpha(\vartheta;x), -\zeta(\vartheta;x)]_{x\in\mathcal{X}}$, and $\tau(s,n)$ is a linear combination of $[s_x, n_x]_{x\in\mathcal{X}}$ in which the linearly dependent rows have been removed. With β and τ defined in this way,

$$\frac{dQ_t}{dQ_0}(\vartheta) = \exp\left(\beta(\vartheta)^T \tau(S_t, N_t) - \Lambda(\tau(S_t, N_t))\right), \tag{3.1}$$

where Λ is defined by

$$\Lambda(\tau) := \log \int_\Theta \exp\left(\beta(\vartheta)^T \tau\right) Q_0(d\vartheta),$$

so that $\Lambda(\tau(S_t,N_t)) = \Gamma(S_t,N_t)$ and Λ has domain $\mathrm{dom}(\Lambda) = \left\{\tau \in \boldsymbol{R}^l : |\Lambda(\tau)| < \infty\right\} = \tau(\mathrm{dom}(\Gamma))$. Here, when we write $\tau(\mathrm{dom}(\Gamma))$, we mean the set constructed by applying the function τ to each point in $\mathrm{dom}(\Gamma)$. We use similar notation below.

Our assumption that $\mathrm{dom}(\Gamma)$ is open, together with the linearity of τ, implies that the natural parameter space $\mathrm{dom}(\Lambda) = \tau(\mathrm{dom}(\Gamma))$ is open. We then recall a fundamental result on exponential families (see, e.g., [4] Theorem 1.6.4). The openness of $\mathrm{dom}(\Lambda)$ implies that $\nabla\Lambda(\tau(s,n)) = \boldsymbol{E}\left[\beta(\theta) \mid S_t = s, N_t = n\right]$ for all $s, n \in \mathrm{dom}(\Gamma)$, and $\nabla\Lambda$ is a bijection between the natural parameter space $\mathrm{dom}(\Lambda)$ and $\mathcal{K} := \nabla\Lambda(\mathrm{dom}(\Lambda))$. Thus we are free to parameterize our posterior at time t in terms of

$$K_t := \nabla\Lambda(\tau(S_t, N_t)),$$

which is a random variable taking values in $\mathcal{K}$. This is called the mean-value parameterization. Since K_t completely determines the posterior distribution Q_t on θ, our policy's sampling decision can then be written entirely in terms of K_t, and we write $\Pi(t,k,C)$ to indicate $\boldsymbol{P}\left\{X_{t+1} \in C \mid K_t = k\right\}$.

We also use the notation $\boldsymbol{P}^{(k)}$ for $k \in \mathcal{K}$ to denote the measure on $(\Theta, \mathcal{B}(\Theta))$ uniquely determined by k. In particular, we then have $Q_t = \boldsymbol{P}^{(K_t)}$. We also write $\boldsymbol{E}^{(k)}$ to indicate the expectation taken under $\boldsymbol{P}^{(k)}$.

The following lemma is needed later when we work with the asymptotic behavior of the sampling policy, and concerns the limit of the stochastic process $(K_t)_{t\in\boldsymbol{N}}$. It uses the notation $\mathrm{cl}(\mathcal{K})$ to denote the closure of $\mathcal{K}$ in $\boldsymbol{R}^l$.

LEMMA 3.1. *The limit* $K_\infty := \lim_{t\to\infty} K_t$ *exists almost surely, is integrable, and takes values in* $\mathrm{cl}(\mathcal{K})$.

3.2. The Value of Information. The sufficient conditions for consistency, introduced in Section 4, use a function g that quantifies the value of information that could be obtained by learning one or more sampling distributions. The value of informamtion is a quantity that was first introduced by [23] and has been applied frequently within decision theory.

The function g is defined for $k \in \mathcal{K}$ and $C \subseteq \mathcal{X}$ by

$$g(k; C) := \min_i \boldsymbol{E}^{(k)}\left[R(\theta; i)\right] - \boldsymbol{E}^{(k)}\left[\min_i \boldsymbol{E}^{(k)}\left[R(\theta; i) \mid G_x, x \in C\right]\right].$$

The quantity $g(k; C)$ gives the expected incremental value of learning the true sampling distribution of all measurement types $x \in C$ when we already have the posterior distribution given by k. Seen another way, $g(k; C)$ is the incremental value of measuring all $x \in C$ an infinite number of times.

The non-negativity of R implies that both outer expectations are non-negative and possibly equal to $+\infty$, and that g is well-defined when both expectations are finite. To ensure g is well-defined, we take its domain to be

$$\mathrm{dom}(g) := \left\{k \in \mathcal{K} : \boldsymbol{E}^{(k)}\left[R(\theta; i)\right] < \infty \ \forall i \in \mathcal{I}\right\}.$$

Then $g(k; C)$ is well-defined and finite for all $k \in \mathrm{dom}(g)$ because Jensen's inequality and the tower property of conditional expectation imply $\boldsymbol{E}^{(k)}\left[\min_i \boldsymbol{E}^{(k)}\left[R(\theta; i) \mid G_x, x \in C\right]\right]$ is bounded above by $\min_i \boldsymbol{E}^{(k)}\left[R(\theta; i)\right]$. This also shows that $g(k; C)$ is non-negative.

Additionally, fixing any $i \in \mathcal{I}$ and any $t \in \boldsymbol{N}$, our assumption that $R(\theta; i)$ is integrable under Q_0 implies that $\infty > \boldsymbol{E}\left[R(\theta; i)\right] = \boldsymbol{E}\left[\boldsymbol{E}\left[R(\theta; i) \mid K_t\right]\right]$, and hence $R(\theta; i)$ is almost-surely integrable under Q_t. Since this is true for each $i \in \mathcal{I}$, we have that $K_t \in \mathrm{dom}(g)$ almost surely.

The following lemma gives a relationship between the function g and consistency.

LEMMA 3.2. *The sampling policy* Π *is consistent iff* $\lim_{t\to\infty} g(K_t; \mathcal{X}) = 0$ *almost surely.*

We now make the following assumption on the structure of the sampling policy, which may be interpreted in this way: if the expected loss cannot be reduced by learning perfectly the result of any *single* measurement type, then there is also nothing to be gained from learning perfectly the results of *all* the measurement types.

ASSUMPTION 1. *If* $(k_t)_{t\in\boldsymbol{N}}$ *is a sequence converging in* $\mathrm{cl}(\mathcal{K})$ *and satisfying* $\lim_{t\to\infty} g(k_t; \{x\}) = 0$ *for all* $x \in \mathcal{X}$, *then* $\lim_{t\to\infty} g(k_t; \mathcal{X}) = 0$.

If this assumption holds, consistency is equivalent to the condition $\lim_{t\to\infty} g(K_t; \{x\}) = 0$ almost surely for all $x \in \mathcal{X}$. This assumption is not restrictive, since in cases for which it does not hold we may expand the set $\mathcal{X}$ of allowed measurement types to a larger set $\mathcal{X}' \subseteq 2^{\mathcal{X}}$ for which it does hold. This is done by considering a block

of measurements of formerly distinct types as a new composite measurement type. Nevertheless, for the discussion that follows, we must check that our sampling model satisfies Assumption 1 and, if it does not, expand $\mathcal{X}$ until it does.

3.3. Law of Large Numbers. If we are able to establish that each measurement type is sampled infinitely often, then the following pair of lemmas are sufficient to prove consistency. They will allow us to concentrate in Section 4 on showing that a measurement policy does indeed sample each measurement type infinitely often.

The first lemma tells us that no further information is to be obtained about $R(\theta;i)$ from G_x if we have already measured x infinitely often. It is essentially a restatement of the strong law of large numbers, and shares much with the Glivenko-Cantelli theorem (see, e.g., [24]). The second lemma builds upon the first and tells us that the incremental value $g(K_t;\{x\})$ of learning the sampling distribution of x goes to zero if we measure x infinitely often.

LEMMA 3.3. *Let $x \in \mathcal{X}$, $i \in \mathcal{I}$. Then $\boldsymbol{E}_\infty [R(\theta;i) \mid G_x] = \boldsymbol{E}_\infty [R(\theta;i)]$ almost surely on $\{N_{\infty,x} = \infty\}$.*

LEMMA 3.4. *Let $x \in \mathcal{X}$. Then $\lim_{t\to\infty} g(K_t;\{x\}) = 0$ almost surely on $\{N_{\infty,x} < \infty\}$.*

4. Sufficient Conditions for Consistency. In this section we present our main result, which is a sufficient condition for consistency that can be applied to policies for which it is difficult to verify whether they sample each measurement type infinitely often. This sufficient condition is phrased in terms of sets M_x and M_*, which partition posteriors according to which types of measurement have value, and sets A_k, which partition measurements according to those that have value and those that do not.

We define M_* and M_x, $x \in \mathcal{X}$, as

$$M_x := \left\{ k \in \mathrm{cl}(\mathcal{K}) : \exists (k_t) \subseteq \mathrm{dom}(g) \text{ converging to } k \text{ with } \lim_{t\to\infty} g(k_t;\{x\}) = 0 \right\},$$
$$M_* := \left\{ k \in \mathrm{cl}(\mathcal{K}) : \forall (k_t) \subseteq \mathrm{dom}(g) \text{ converging to } k, \lim_{t\to\infty} g(k_t;\{x\}) = 0 \ \forall x \in \mathcal{X} \right\}.$$

If, for each $x \in \mathcal{X}$, the function $k \mapsto g(k;\{x\})$ is continuous and can be extended continuously onto the closure $\mathrm{cl}(\mathrm{dom}(g))$ of its domain with a new range $\boldsymbol{R}_+ \cup \{\infty\}$, then these definitions may be simplified to

$$M_x = \{k \in \mathrm{cl}(\mathrm{dom}(g)) : g(k;\{x\}) = 0\},$$
$$M_* = \{k \in \mathrm{cl}(\mathrm{dom}(g)) : g(k;\{x\}) = 0, \ \forall x \in \mathcal{X}\}.$$

If Assumption 1 also holds then we have $M_* = \{k \in \mathrm{cl}(\mathrm{dom}(g)) : g(k;\mathcal{X}) = 0\}$.

In this case, the set M_x is the set of knowledge states in which sampling x will not improve the expected loss. This includes knowledge states in which the sampling distribution of x is already known. The set M_* is the set of knowledge states in which further sampling (of any type) has no value. This is the set to which we would like our sampling policy to push us, since it is the set in which we have learned as much as possible. The sets $M_x \setminus M_*$ are the "x-sticking sets" in which measurements of type x do not have value, but other measurements do.

For each $k \in \mathrm{cl}(\mathcal{K})$, we define $A_k := \{x \in \mathcal{X} : k \in M_x\}$ to be the set of measurements that are stuck when we are in knowledge state k. To avoid sticking, a sampling policy should avoid measuring x when in such states. In fact, to guarantee consistency, a policy need only do a little more. It need only maintain an open region around these sticking sets in which it avoids measuring the stuck alternatives. This is the essential content of Theorem 4.1 below.

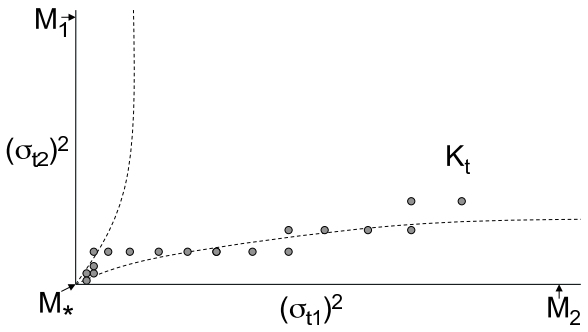

FIG. 4.1. *Two-dimensional projection of $\mathcal{K}$ for the ranking and selection problem. The dimensions pictured are the posterior variances σ^2_{t1} and σ^2_{t2}. The vertical axis is M_1, where $\sigma^2_{t1} = 0$ and the first alternative's sampling mean is known. Similarly, the horizontal axis is M_2. Their intersection is M_*. Dashed lines give open regions around M_1 and M_2 in which stuck alternatives are not measured, and the dots K_t are a sequence of knowledge states converging to M_*.*

The open regions required by Theorem 4.1 are illustrated in Figure 4.1, where we picture the knowledge state that results from a ranking and selection problem with two alternatives, samples from which are normally distributed with unknown mean and known variance. This problem is discussed in greater detail in Section 5. In it, if we take an independent normally distributed prior distribution on these two unknown means then the posterior distribution on each alternative is parameterized by its mean and variance, causing knowledge states to have four dimensions.

In the figure, we picture two of these four dimensions: the variances of our belief about each of the two unknown means. When the variance of belief about an unknown mean is zero then that mean is known perfectly, and so the sticking set for an alternative is this set of points. These are pictured as thick lines along the vertical axis for alternative 2 and along the horizontal axis for alternative 1. The point at the origin is M_*, and is where both alternatives are known perfectly. To ensure consistency by Theorem 4.1, a sampling policy needs an open region around the vertical axis, excluding the origin, in which it only measures alternative 2, and an open region around the horizontal axis, excluding the origin, in which it only measures alternative 1. If this condition is met, then the sequence of knowledge states K_t will converge to the origin, attaining consistency.

In this example, we see the necessity of the open regions around the sticking sets. The variance of our belief about an alternative's unknown mean is always strictly positive after finitely many measurements, and can only approach 0 in the limit. Thus, a policy can never reach M_1 or M_2 with finitely many measurements but can only come arbitrarily close. If a policy merely guarantees that it never measures alternative 1 when in M_1, but measures 1 in all other knowledge states, then its knowledge state may converge to a point outside M_*, and the policy may not be consistent.

We now state the theorem.

THEOREM 4.1. *Suppose that the sampling model satisfies Assumption 1, and let $\widetilde{\mathcal{K}}$ be a measurable subset of the closure* $\mathrm{cl}(\mathcal{K})$ *of $\mathcal{K}$ such that the event* $\left\{K_t \in \widetilde{\mathcal{K}},\ \forall t \in \boldsymbol{N} \cup \{\infty\}\right\}$

is almost sure. If each $k \in \widetilde{\mathcal{K}} \setminus M_*$ *has an open neighborhood* U *in* $\mathrm{cl}(\mathcal{K})$ *satisfying*

$$\limsup_{t\to\infty} \sup_{k'\in U\cap\widetilde{\mathcal{K}}\cap\mathcal{K}} \Pi(t,k',A_k) < 1, \tag{4.1}$$

then the sampling policy given by Π *is consistent.*

In the next sections we apply this theorem to the ranking and selection problem, showing consistency of two previously proposed policies. Both of these policies, as well as many other policies for which it is difficult to verify that they measure each measurement type infinitely often, are *stationary* and *non-randomized*. By this, we mean that Π does not depend on t and that $\Pi(t,k,\cdot)$ is a point mass on one particular measurement type, call it $X^\Pi(k)$, for each k. $X^\Pi(k)$ is the measurement type used when the current belief is k. Under these conditions, Theorem 4.1 simplifies to the following corollary.

COROLLARY 4.2. *Suppose the sampling policy given by* Π *is stationary and non-randomized, Assumption 1 is met, and* $\widetilde{\mathcal{K}}$ *is defined as in Theorem 4.1. If each* $k \in \widetilde{\mathcal{K}} \setminus M_*$ *has an open neighborhood* U *in* $\mathrm{cl}(\mathcal{K})$ *such that* $X^\Pi(k') \notin A_k$ *for each* $k' \in U \cap \widetilde{\mathcal{K}} \cap \mathcal{K}$*, then the policy is consistent.*

5. Application to Ranking and Selection. We apply Theorem 4.1 from the previous section to the normal ranking and selection problem, using it to show convergence of a policy previously proposed in the literature. In the ranking and selection problem, we have a collection of alternatives from which we would like to choose the one that is best in some sense. One may make a series of measurements before making the final selection. We suppose that measurements of an alternative are normally distributed, and the best alternative is the one with the largest mean.

This problem is formulated as follows. We call the set of alternatives $\mathcal{I}$ and its cardinality $|\mathcal{I}| = d > 1$. Then we take $\Theta = \boldsymbol{R}^d$, and samples of alternative $i \in \mathcal{I}$ are normally distributed with mean θ_i and known precision $\lambda_i > 0$. We suppose that measurements are taken in batches of some fixed integer size $B \geq 1$. We allow $B = 1$, which corresponds to taking the measurements one-at-a-time. The possible measurement types are then $\mathcal{X} = \mathcal{I}^B$, with $\sum_{b=1}^B \mathbf{1}_{\{x_b=i\}}$ being the number of samples measurement type x takes from alternative i. We also have $\mathcal{Y} = \boldsymbol{R}^B$, with y_b being an observation of alternative x_b. We take the *linear loss function* given by

$$R(\vartheta;i) = \max_{j\in\mathcal{I}} \vartheta_j - \vartheta_i. \tag{5.1}$$

The sampling density with respect to the Lebesgue measure is

$$p(y;x,\vartheta) = \prod_{b=1}^B \sqrt{\frac{\lambda_{x_b}}{2\pi}} \exp\left(-\frac{1}{2}\lambda_{x_b}(y_b - \vartheta_{x_b})^2\right).$$

To put this in the form of (2.1), we take $\alpha(\vartheta;x) = (\vartheta_{x_1}\lambda_{x_1},\ldots,\vartheta_{x_b}\lambda_{x_b})$, $\gamma(y;x) = y$, $\zeta(\vartheta;x) = \sum_{b=1}^B \vartheta_{x_b}^2\lambda_{x_b}/2$, and $\nu(dy;x) = \prod_{b=1}^B \sqrt{\frac{\lambda_{x_b}}{2\pi}} \exp(-\frac{1}{2}\lambda_{x_b}y_{x_b}^2)dy$. Then the density of the posterior is given by (3.1) with

$$\beta(\vartheta) = \left[\vartheta_i\lambda_i, -\frac{1}{2}\vartheta_i^2\lambda_i\right]_{i\in\mathcal{I}}, \qquad \tau(S_t,N_t) = \left[\tilde{S}_{ti}, \widetilde{N}_{ti}\right]_{i\in\mathcal{I}},$$

where $\tilde{S}_{ti} = \sum_{t'\leq t}\sum_{b=1}^B \mathbf{1}_{\{X_{t'b}=i\}}Y_{t'b} = \sum_{x\in\mathcal{X}} S_{tx}\sum_{b=1}^B \mathbf{1}_{\{x_b=i\}}$ is the sum of all observations of alternative i, and $\widetilde{N}_{ti} = \sum_{t'\leq t}\sum_{b=1}^B \mathbf{1}_{\{X_{t'b}=i\}} = \sum_{x\in\mathcal{X}} N_{tx}\sum_{b=1}^B \mathbf{1}_{\{x_b=i\}}$ is the number of times we have observed alternative i.

We suppose that the prior Q_0 on θ is a multivariate normal distribution with independent components and with θ_i having mean $\hat{\mu}_{0i}$ and variance $\sigma_{0i}^2 > 0$. With this assumption, the posterior Q_t is again normal with independent components, but now the mean and variance of θ_i, which we denote $\hat{\mu}_{ti}$ and σ_{ti}^2 respectively, are given by

$$\hat{\mu}_{ti} = \left(\frac{\hat{\mu}_{0i}}{\lambda_i \sigma_{0i}^2} + \tilde{S}_{ti}\right) \Big/ \left(\frac{1}{\lambda_i \sigma_{0i}^2} + \widetilde{N}_{ti}\right), \qquad \sigma_{ti}^2 = \left(\frac{1}{\lambda_i}\right) \Big/ \left(\frac{1}{\lambda_i \sigma_{0i}^2} + \widetilde{N}_{ti}\right).$$

With this, we compute K_t as $K_t = \boldsymbol{E}_t\left[\beta(\theta)\right] = \left[\hat{\mu}_{ti}\lambda_i, -\lambda_i(\hat{\mu}_{ti}^2 + \sigma_{ti}^2)/2\right]_{i\in\mathcal{I}}$, and $\mathcal{K}$ and its closure are given by

$$\mathcal{K} = \left\{ \left[u_i\lambda_i, -\lambda_i(u_i^2 + v_i)/2\right]_{i\in\mathcal{I}} : u \in \boldsymbol{R}^d, v \in \boldsymbol{R}_{++}^d \right\},$$
$$\mathrm{cl}(\mathcal{K}) = \left\{ \left[u_i\lambda_i, -\lambda_i(u_i^2 + v_i)/2\right]_{i\in\mathcal{I}} : u \in \boldsymbol{R}^d, v \in \boldsymbol{R}_{+}^d \right\},$$

with $\boldsymbol{R}_{++}$ being the set of strictly positive real numbers.

We may recover $\hat{\mu}_{ti}$ and σ_{ti}^2 from K_t, and, more generally, recover the mean and variance of θ_i under $\boldsymbol{P}^{(k)}$ from any $k \in \mathcal{K}$. To do so, we define functions $\hat{\mu}_i : \mathrm{cl}(\mathcal{K}) \mapsto \boldsymbol{R}$ and $\sigma_i^2 : \mathrm{cl}(\mathcal{K}) \mapsto \boldsymbol{R}_+$ by

$$\hat{\mu}_i(k) = k_{i1}/\lambda_i, \qquad \sigma_i^2(k) = -(k_{i1}/\lambda_i)^2 - 2k_{i2}/\lambda_i.$$

Then $\hat{\mu}_{ti} = \hat{\mu}_i(K_t)$, and $\sigma_{ti}^2 = \sigma_i^2(K_t)$. Note that we define $\hat{\mu}_i$ and σ_i^2 on $\mathrm{cl}(\mathcal{K})$, rather than on just $\mathcal{K}$.

5.1. Analysis of the Sampling Model. We now discuss the continuity of the function g in this sampling model, and describe the sets M_x and M_*.

For $C \subseteq \mathcal{X}$, define $\mathcal{I}_C = \left\{ i \in \mathcal{I} : \sum_{x\in C}\sum_{b=1}^{B} \mathbf{1}_{\{x_b = i\}} > 0 \right\}$ and, for $k \in \mathrm{cl}(\mathcal{K})$, define $\mathcal{I}(k) = \left\{ i \in \mathcal{I} : \sigma_i^2(k) = 0 \right\}$. For $k \in \mathrm{cl}(\mathcal{K}) \setminus \mathcal{K}$, let $\boldsymbol{P}^{(k)}$ denote the measure on Θ under which μ_i is independent of $(\mu_j)_{j\neq i}$, and is distributed according to $\mu_i = \hat{\mu}_i(k)$ almost surely if $\sigma_i^2(k) = 0$, or $\mu_i \sim \mathrm{Normal}(\hat{\mu}_i(k), \sigma_i^2(k))$ if $\sigma_i^2(k) > 0$, This definition of $\boldsymbol{P}^{(k)}$ for $k \notin \mathcal{K}$ is a natural extension to the earlier definition for $k \in \mathcal{K}$ because we can see by checking convergence on elementary sets that if $(k_t) \subseteq \mathrm{cl}(\mathcal{K})$ converges to k_* in $\mathrm{cl}(\mathcal{K})$, then $\boldsymbol{P}^{(k_t)}$ converges weakly to $\boldsymbol{P}^{(k_*)}$. Let $\boldsymbol{E}^{(k)}$ denote the expectation under $\boldsymbol{P}^{(k)}$.

We now have the following pair of lemmas. In the first lemma, and elsewhere throughout this work, max over the empty set is understood to be $-\infty$, and min over the empty set is understood to by $+\infty$.

LEMMA 5.1. *$\mathrm{dom}(g) = \mathcal{K}$, and, for each $C \subseteq \mathcal{X}$, $k \mapsto g(k; C)$ is continuous on $\mathrm{dom}(g)$ and can be extended continuously onto $\mathrm{cl}(\mathrm{dom}(g)) = \mathrm{cl}(\mathcal{K})$ by*

$$g(k; C) = \boldsymbol{E}^{(k)}\left[\max\left(\max_{i\in\mathcal{I}_C}\mu_i, \max_{i\notin\mathcal{I}_C}\hat{\mu}_i(k)\right)\right] - \max_{i\in\mathcal{I}}\hat{\mu}_i(k). \tag{5.2}$$

LEMMA 5.2. *For $k \in \mathrm{cl}(\mathrm{dom}(g))$ and $C \subseteq \mathcal{X}$, $g(k; C) = 0$ iff $\mathcal{I}_C \subseteq \mathcal{I}(k)$.*

These lemmas allow us to write the sets M_x and M_* as

$$\begin{aligned} M_x &= \{k \in \mathrm{cl}(\mathrm{dom}(g)) : g(k; \{x\}) = 0\} = \left\{k \in \mathrm{cl}(\mathcal{K}) : \mathcal{I}_{\{x\}} \subseteq \mathcal{I}(k)\right\} \\ &= \left\{k \in \mathrm{cl}(\mathcal{K}) : \sigma_i^2(k) = 0 \ \forall i \in \mathcal{I}_{\{x\}}\right\}, \\ M_* &= \left\{k \in \mathrm{cl}(\mathcal{K}) : \sigma_i^2(k) = 0 \ \forall i \in \mathcal{I}\right\}. \end{aligned}$$

We can also conclude that the sampling model satisfies Assumption 1, since if $k \in \mathrm{cl}(\mathcal{K})$ has $g(k;\{x\}) = 0$ for all $x \in \mathcal{X}$, then $\sigma_i^2(k) = 0$ for all $i \in \mathcal{I}$, and $\mathcal{I}(k) = \mathcal{I} = \mathcal{I}_{\mathcal{X}}$ implying through Lemma 5.2 that $g(k;\mathcal{X}) = 0$.

5.2. The Knowledge-Gradient Policy. In this section we prove consistency of the so-called $(R1,\ldots,R1)$ policy first introduced in [21] as one of a large number of policies for ranking and selection, and analyzed more fully under the name "Knowledge-Gradient (KG) Policy" in [16]. Throughout this section we will call this policy the KG policy. The consistency of this policy was established in [16] in a framework specific to that policy. The new proof presented here is simpler than that in [16], because it uses Theorem 4.1, and helps to demonstrate how Theorem 4.1 can be applied to prove consistency of a variety of measurement policies.

The KG policy is defined by considering what the incremental value would be of taking a single sample from alternative x, and then stopping. This myopic value of information is called the KG-factor and written $\nu_x^{KG}(k)$ when the current knowledge state is k, and in the normal ranking and selection problem considered here can be explicitly calculated for $k \in \mathrm{cl}(\mathcal{K})$ as

$$\nu_x^{KG}(k) = \begin{cases} \tilde{\sigma}_x(k) f\left(-\left|\hat{\mu}_x(k) - \max_{i\neq x} \hat{\mu}_i(k)\right| \Big/ \tilde{\sigma}_x(k)\right), & \text{if } \sigma_x^2(k) > 0, \\ 0, & \text{if } \sigma_x^2(k) = 0, \end{cases}$$

where $\tilde{\sigma}_x(k) = \sqrt{\lambda_x \sigma_x^2(k) / \left(\frac{1}{\sigma_x^2(k)} + \lambda_x\right)}$, and $f(z) = z\Phi(z) + \varphi(z)$, with Φ the standard normal cumulative distribution function and φ the standard normal probability density function. The KG policy calculates the KG factor $\nu_x^{KG}(K_t)$ for each alternative x, and then samples the alternative with the largest KG factor, thus satisfying

$$X_t \in \arg\max_x \nu_x^{KG}(K_t). \tag{5.3}$$

Ties may be broken at random, or according to some fixed ordering on the alternatives (e.g., smallest index). In the second case, the KG policy is stationary and nonrandomized. The way in which ties are broken does not affect consistency.

THEOREM 5.3. *The KG policy defined by* (5.3) *is consistent.*

Proof. Let $\widetilde{\mathcal{K}} = \mathrm{cl}(\mathcal{K})$, and choose any $k \in \widetilde{\mathcal{K}} \setminus M_*$. Define the set U for this k as

$$U := \left\{ k' \in \mathrm{cl}(\mathcal{K}) : \min_{x \notin A_k} \nu_x^{KG}(k') > \max_{x \in A_k} \nu_x^{KG}(k') \right\}.$$

We now check the conditions of Theorem 4.1. When A_k is empty, $U = \mathrm{cl}(\mathcal{K})$, and the conditions of the theorem (U is open; $k \in U$; and the condition in (4.1)) are met trivially. Consider the case when A_k is not empty.

First, the function $k' \mapsto \nu_x^{KG}(k')$ is continuous with domain $\mathrm{cl}(\mathcal{K})$, which implies that U is open in $\mathrm{cl}(\mathcal{K})$. Second, $A_k = \left\{x : \sigma_x^2(k) = 0\right\}$, and so the two facts $\nu_x(k) = 0$ when $\sigma_x^2(k) = 0$ and $\nu_x(k) > 0$ when $\sigma_x^2(k) > 0$ together imply that $k \in U$. Third, for any $k' \in U$, $\min_{x\notin A_k} \nu_x^{KG}(k') > \max_{x\in A_k} \nu_x^{KG}(k')$. Since $k \notin M_*$ implies $A_k \neq \mathcal{X}$, there is at least one $x \notin A_k$ whose KG factor $\nu_x^{KG}(k')$ is strictly larger than the KG factor of each alternative in A_k. This implies that the KG policy does not measure an alternative in A_k, and $\Pi(t,k',A_k) = 0$ for each t. This, together with $U \subseteq \mathrm{cl}(\mathcal{K}) = \mathcal{K}$, implies that the condition in (4.1) is met. This shows that the KG policy meets the conditions of Theorem 4.1, and thus is consistent. □

5.3. The OCBA for Linear Loss Policy. We now show consistency of the optimal computing budget allocation (OCBA) for linear loss, which is a ranking and selection policy that operates within this known-variance normal sampling model, and was first proposed by [22]. This policy is the product of a line of research on OCBA policies beginning with [7, 8, 9].

OCBA for linear loss is derived by first supposing that the current batch of B samples to be taken will be the last, and then choosing a measurement allocation that approximates the measurement allocation that would be optimal were this single-batch assumption true. The policy bases its measurement decisions upon an approximation to the reduction in expected linear loss achieved by a batch of measurements. It is described fully in Algorithm 1.

Algorithm 1 OCBA for linear loss

Require: Input knowledge state $k \in \mathrm{cl}(\mathcal{K})$ satisfying $|\arg\max_i \hat{\mu}_i(k)| = 1$, the batch size B, and an integer parameter m dividing B.

1: Choose $i^* \in \arg\max_i \hat{\mu}_i(k)$. This choice is unique by assumption.

2: For $i \in \mathcal{I} \setminus \{i^*\}$, define

$$\delta_i := \hat{\mu}_{i^*}(k) - \hat{\mu}_i(k), \qquad \tilde{\sigma}_{i,i^*} := \sqrt{\sigma_i^2(k) + (B/m + 1/\sigma_{i^*}^2(k))^{-1}},$$
$$\tilde{\sigma}_i := \sqrt{\sigma_i^2(k) + \sigma_{i^*}^2(k)}, \qquad \tilde{\sigma}_{i^*,i} := \sqrt{\sigma_{i^*}^2(k) + (B/m + 1/\sigma_i^2(k))^{-1}},$$

with $\tilde{\sigma}_{i,i^*} = \sigma_i(k)$ if $\sigma_{i^*}(k) = 0$, and $\tilde{\sigma}_{i^*,i} = \sigma_{i^*}(k)$ if $\sigma_i(k) = 0$.

3: Define

$$D_i(k) := \begin{cases} \tilde{\sigma}_{i^*,i} f(\delta_i/\tilde{\sigma}_{i^*,i}) - \tilde{\sigma}_i f(\delta_i/\tilde{\sigma}_i), & \text{if } i \neq i^*, \\ \sum_{i \neq i^*} \tilde{\sigma}_{i,i^*} f(\delta_i/\tilde{\sigma}_{i,i^*}) - \tilde{\sigma}_i f(\delta_i/\tilde{\sigma}_i), & \text{if } i = i^*, \end{cases}$$

where $f(z) := \varphi(z) + z\Phi(z)$, φ is the normal pdf, and Φ is the normal cdf. For any $\delta \in \boldsymbol{R}$, we take $0 \cdot f(\delta/0) = \lim_{\sigma \to 0} \sigma f(\delta/\sigma) = 0$.

4: Define $S(m) := \{i \in \mathcal{I} : D_i(k) \text{ is among the } m \text{ lowest values}\}$.

5: Allocate B/m samples to each alternative in $S(m)$, and no samples to other alternatives.

Algorithm 1 assumes of its input that $|\arg\max_i \hat{\mu}_i(k)| = 1$. When using the OCBA for linear loss policy as described in [22], one usually begins with a noninformative prior, then takes a fixed number of measurements from each alternative to obtain an informative posterior belief, and only afterward uses the OCBA policy. The belief after the fixed first stage will almost surely satisfy this assumption $|\arg\max_i \hat{\mu}_i(k)| = 1$, as will all subsequent beliefs. In such a setting, our consistency result may then be understood as beginning where sampling under OCBA for linear loss begins – immediately after the first stage completes – and Theorem 5.4 below to show that OCBA for linear loss is consistent under the belief held at this time.

THEOREM 5.4. *Suppose that* $|\arg\max_i \hat{\mu}_{0i}| = 1$. *Then the OCBA for linear loss defined in Algorithm 1 is consistent for the sampling model and loss function in Section 5.*

Proof. Let $\widetilde{\mathcal{K}} = \{k \in \mathrm{cl}(\mathcal{K}) : |\arg\max_i \hat{\mu}_i(k)| = 1\}$. The assumed uniqueness of $\arg\max_i \hat{\mu}_{0i}$ implies $K_t \in \widetilde{\mathcal{K}}$ almost surely for each $t \in \boldsymbol{N}$, and this together with the fact that $\boldsymbol{P}\{\mu_i = \mu_j,\ i \neq j\} = 0$ implies $K_\infty \in \widetilde{\mathcal{K}}$ almost surely.

Now choose $\tilde{k} \in \widetilde{\mathcal{K}} \setminus M_*$, let $i^* \in \arg\max_i \hat{\mu}_i(\tilde{k})$, and define U by

$$U := \left\{ k \in \mathrm{cl}(\mathcal{K}) : \hat{\mu}_{i^*}(k) > \max_{i \neq i^*} \hat{\mu}_i(k), \min_{i \in \mathcal{I}(\tilde{k})} D_i(k) > \max_{i \notin \mathcal{I}(\tilde{k})} D_i(k) \right\}.$$

We first note that if $i \in \mathcal{I}(\tilde{k})$ then $\sigma_i^2(\tilde{k}) = 0$, implying $D_i(\tilde{k}) = 0$. Also, if $i \notin \mathcal{I}(\tilde{k})$ then $\sigma_i^2(\tilde{k}) > 0$, implying $D_i(\tilde{k}) < 0$. Thus $\min_{i \in \mathcal{I}(\tilde{k})} D_i(\tilde{k}) > \max_{i \notin \mathcal{I}(\tilde{k})} D_i(\tilde{k})$, including in the cases $\mathcal{I}(\tilde{k}) = \emptyset$ and $\mathcal{I}(\tilde{k}) = \mathcal{X}$. Thus $\tilde{k} \in U$.

Second, for each $k \in U \cap \widetilde{\mathcal{K}} \cap \mathcal{K}$, $\min_{i \in \mathcal{I}(\tilde{k})} D_i(k) > \max_{i \notin \mathcal{I}(\tilde{k})} D_i(k)$ implies that the OCBA for linear loss policy assigns at least B/m measurements to an alternative outside $\mathcal{I}(\tilde{k})$. Since $A_{\tilde{k}}$ consists of those measurement types allocating all B samples within $\mathcal{I}(\tilde{k})$, we have $X^{\Pi}(k) \notin A_{\tilde{k}}$.

Third, since both D_i and $\hat{\mu}_i$ are continuous for each $i \in \mathcal{I}$, U is open in $\mathrm{cl}(\mathcal{K})$. Thus the conditions of Corollary 4.2 are met and consistency follows. □

6. Conclusion. We have presented a powerful and general sufficient condition for consistency of a sequential sampling policy. We have demonstrated the applicability of this sufficient condition by using it to show consistency of two policies for ranking and selection: OCBA for linear loss and $(R_1, \ldots, R_1)$. Although consistency by itself is no guarantee that a policy performs well in practice, lack of consistency is a dangerous signal suggesting that a policy may perform extremely badly in some cases. For this reason, it is our hope that the work we present here may be used to more easily check whether a policy under consideration is consistent, offering a warning to those using inconsistent policies, and reassurance those using consistent policies.

7. Appendix. Proof of Lemma 3.1: The random variable $\beta(\theta)$ is integrable under Q_0 and hence under $\boldsymbol{P}$ because the measure Q_0 is in the exponential family with natural parameter $\tau(S_0, N_0) = 0 \in \mathrm{dom}(\Lambda)$, and the openness of $\mathrm{dom}(\Lambda)$ implies the moment generating function of $\beta(\theta)$ under Q_0, which is $\exp(\Psi(\cdot))$, is finite in an open ball around 0 in $\boldsymbol{R}^l$.

The integrability of $\beta(\theta)$, together with the fact $K_t = \boldsymbol{E}_t[\beta(\theta)]$, implies that $(K_t)_{t \in \boldsymbol{N}}$ is a uniformly integrable martingale. By [24] Theorem 6.21, $(K_t)_{t \in \boldsymbol{N}}$ converges almost surely to an integrable random variable. This random variable is K_∞. It takes values in $\mathrm{cl}(\mathcal{K})$ because each K_t takes values in $\mathcal{K}$.

Proof of Lemma 3.2: For each $T \in \boldsymbol{N}$, the inner expectation on the right-hand side of (2.2) satisfies $\boldsymbol{E}[R(\theta; i) \mid G_x, x \in \mathcal{X}] = \boldsymbol{E}_T[R(\theta; i) \mid G_x, x \in \mathcal{X}]$ since $(X_t, Y_t)_{t \leq T}$ is conditionally independent of $R(\theta; i)$ given G_x, $x \in \mathcal{X}$. Using this and the tower property of conditional expectation, the right-hand side of (2.2) can be rewritten as $\boldsymbol{E}[\boldsymbol{E}_T[\min_{i \in \mathcal{I}} \boldsymbol{E}_T[R(\theta; i) \mid G_x, x \in \mathcal{X}]]]$. Thus, consistency holds iff

$$\lim_{T \to \infty} \boldsymbol{E}\left[\min_i \boldsymbol{E}_T[R(\theta; i)] - \boldsymbol{E}_T\left[\min_i \boldsymbol{E}_T[R(\theta; i) \mid G_x, x \in \mathcal{X}]\right]\right] = 0,$$

which is true iff $g(K_T; \mathcal{X}) = \min_i \boldsymbol{E}_T[R(\theta; i)] - \boldsymbol{E}_T[\min_i \boldsymbol{E}_T[R(\theta; i) \mid G_x, x \in \mathcal{X}]]$ converges almost surely to 0, where we use the fact that applying Jensen's inequality and the tower property to $\boldsymbol{E}_T[\min_i \boldsymbol{E}_T[R(\theta; i) \mid G_x, x \in \mathcal{X}]]$ shows $g(K_T; \mathcal{X})$ is nonnegative.

Proof of Lemma 3.3: Fix $x \in \mathcal{X}$ and $i \in \mathcal{I}$, and let Ω_x denote the event $\{N_{\infty,x} = \infty\}$. Let $\mathcal{C}$ denote a countable separating class for $\mathcal{Y}$. Then, any measure on $\mathcal{Y}$ is completely

determined by the values it takes on the elements of $\mathcal{C}$, and $\sigma\{G_x\} = \bigvee_{C\in\mathcal{C}} \sigma\{G_x(C)\}$, where $G_x(C) = P(C; x, \theta)$.

Thus, $\boldsymbol{E}_\infty[R(\theta;i) \mid G_x]$ may be written as $f((X_t,Y_t)_{t\in\boldsymbol{N}}, (G_x(C))_{C\in\mathcal{C}})$ for some measurable function $f : (\mathcal{X}\times\mathcal{Y})^{\boldsymbol{N}} \times [0,1]^{\mathcal{C}} \mapsto \boldsymbol{R}$. For each $C \in \mathcal{C}$ define

$$\widehat{G}_C := \begin{cases} 0, & \text{if } \sum_{t\in\boldsymbol{N}} \mathbf{1}_{\{X_t=x\}} = 0, \\ \lim_{t\to\infty} \left(\sum_{t'\le t} \mathbf{1}_{\{X_{t'}=x\}}\mathbf{1}_{\{Y_{t'}\in C\}}\right)/N_{tx}, & \text{otherwise.} \end{cases}$$

This random variable is $\mathcal{F}_\infty$-measurable, and the event $\left\{\widehat{G}_C = G_x(C)\ \forall C\in\mathcal{C}\right\}$ is almost sure on Ω_x by the strong law of large numbers and the countability of $\mathcal{C}$. Thus,

$$\boldsymbol{E}_\infty[R(\theta;i) \mid G_x] = f((X_t,Y_t)_{t\in\boldsymbol{N}}, (G_x(C))_{C\in\mathcal{C}}) = f((X_t,Y_t)_{t\in\boldsymbol{N}}, (\widehat{G}_C)_{C\in\mathcal{C}})$$

almost surely on Ω_x, and $\boldsymbol{E}_\infty[R(\theta;i)\mathbf{1}_{\Omega_x} \mid G_x] = f((X_t,Y_t)_{t\in\boldsymbol{N}}, (\widehat{G}_C)_{C\in\mathcal{C}})\mathbf{1}_{\Omega_x}$ almost surely, where we use the fact that $\mathbf{1}_{\Omega_x} \in \mathcal{F}_\infty$.

Finally, by the tower property and the $\mathcal{F}_\infty$-measurability of $\mathbf{1}_{\Omega_x}$,

$$\begin{aligned}\boldsymbol{E}_\infty[R(\theta;i)\mathbf{1}_{\Omega_x}] &= \boldsymbol{E}_\infty[\boldsymbol{E}_\infty[R(\theta;i)\mathbf{1}_{\Omega_x} \mid G_x]] = \boldsymbol{E}_\infty\left[f((X_t,Y_t)_{t\in\boldsymbol{N}}, (\widehat{G}_C)_{C\in\mathcal{C}})\mathbf{1}_{\Omega_x}\right] \\ &= f((X_t,Y_t)_{t\in\boldsymbol{N}}, (\widehat{G}_C)_{C\in\mathcal{C}})\mathbf{1}_{\Omega_x} = \boldsymbol{E}_\infty[R(\theta;i)\mathbf{1}_{\Omega_x} \mid G_x].\end{aligned}$$

Proof of Lemma 3.4: Fix $x \in \mathcal{X}$, and define the event $\Omega_x = \{N_{\infty,x} = \infty\}$. Now, fix $i \in \mathcal{I}$ and consider the sequence of conditional expectations $(\boldsymbol{E}_t[R(\theta;i) \mid G_x])_{t\in\boldsymbol{N}}$. Since $R(\theta;i)$ is integrable, this is a uniformly integrable martingale with respect to the filtration $(\mathcal{F}_t \vee \sigma\{G_x\})_{t\in\boldsymbol{N}}$, and converges almost surely and in L^1 to the integrable random variable $\boldsymbol{E}_\infty[R(\theta;i) \mid G_x]$ (see, e.g., [24] Theorem 6.21). Similarly, $(\boldsymbol{E}_t[R(\theta;i)])_{t\in\boldsymbol{N}}$ is a uniformly integrable martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\boldsymbol{N}}$, and converges almost surely and in L^1 to the integrable random variable $\boldsymbol{E}_\infty[R(\theta;i)]$.

By Lemma 3.3, $\boldsymbol{E}_\infty[R(\theta;i) \mid G_x]\mathbf{1}_{\Omega_x} = \boldsymbol{E}_\infty[R(\theta;i)]\mathbf{1}_{\Omega_x}$. For each $t \in \boldsymbol{N}\cup\{\infty\}$ define two random variables, $R_t := \min_{i\in\mathcal{I}} \boldsymbol{E}_t[R(\theta;i)]$, and $\widetilde{R}_t := \min_{i\in\mathcal{I}} \boldsymbol{E}_t[R(\theta;i) \mid G_x]$. Since $\mathcal{I}$ is a finite set, $R_t \to R_\infty$ almost surely and in L^1, $\widetilde{R}_t \to \widetilde{R}_\infty$ almost surely and in L^1, and $R_\infty\mathbf{1}_{\Omega_x} = \widetilde{R}_\infty\mathbf{1}_{\Omega_x}$ almost surely.

We now show $\boldsymbol{E}_t\widetilde{R}_t$ converges in L^1 to $\boldsymbol{E}_\infty\widetilde{R}_\infty$. We begin by using the triangle inequality to bound $\lim_t \boldsymbol{E}\left[|\boldsymbol{E}_t\widetilde{R}_t - \boldsymbol{E}_\infty\widetilde{R}_\infty|\right]$ above by $\lim_t \boldsymbol{E}\left[\left|\boldsymbol{E}_t\widetilde{R}_t - \boldsymbol{E}_t\widetilde{R}_\infty\right|\right] + \lim_t \boldsymbol{E}\left[\left|\boldsymbol{E}_t\boldsymbol{E}_\infty\widetilde{R}_\infty - \boldsymbol{E}_\infty\widetilde{R}_\infty\right|\right]$. We show that this upper bound is zero. We rewrite the first term and bound it above via Jensen's inequality to obtain,

$$\lim_t \boldsymbol{E}\left[\left|\boldsymbol{E}_t\widetilde{R}_t - \boldsymbol{E}_t\widetilde{R}_\infty\right|\right] = \lim_t \boldsymbol{E}\left[\left|\boldsymbol{E}_t\left[\widetilde{R}_t - \widetilde{R}_\infty\right]\right|\right] \le \lim_t \boldsymbol{E}\left[\left|\widetilde{R}_t - \widetilde{R}_\infty\right|\right] = 0,$$

which is zero since $\widetilde{R}_t$ converges to $\widetilde{R}_\infty$ in L^1. Examining the second term, $\widetilde{R}_\infty$ is an integrable random variable, so $\left(\boldsymbol{E}_t\widetilde{R}_\infty\right)_{t\in\boldsymbol{N}}$ is a uniformly integrable martingale converging in L^1 to $\boldsymbol{E}_\infty\widetilde{R}_\infty$. This shows the second term is also zero, implying that $\lim_t \boldsymbol{E}\left[|\boldsymbol{E}_t\widetilde{R}_t - \boldsymbol{E}_\infty\widetilde{R}_\infty|\right] = 0$, and $\boldsymbol{E}_t\widetilde{R}_t$ converges in L^1 to $\boldsymbol{E}_\infty\widetilde{R}_\infty$.

Conditioning on $\mathcal{F}_t \vee \sigma\{G_x\}$ and using Jensen's inequality with the tower property shows that $\boldsymbol{E}_t\left[\boldsymbol{E}_{t+1}\widetilde{R}_{t+1}\right] \leq \boldsymbol{E}_t\widetilde{R}_t$, and so $(\boldsymbol{E}_t\widetilde{R}_t)_{t\in\boldsymbol{N}}$ is a supermartingale. Since it is nonnegative, it converges almost surely, and this convergence must be to the same random variable $\boldsymbol{E}_\infty\widetilde{R}_\infty$ to which L^1 convergence is shown above. This implies $(\boldsymbol{E}_t\widetilde{R}_t)\mathbf{1}_{\Omega_x}$ converges almost surely to $\left(\boldsymbol{E}_\infty\widetilde{R}_\infty\right)\mathbf{1}_{\Omega_x} = \boldsymbol{E}_\infty\left[\widetilde{R}_\infty\mathbf{1}_{\Omega_x}\right] = \boldsymbol{E}_\infty\left[R_\infty\mathbf{1}_{\Omega_x}\right] = R_\infty\mathbf{1}_{\Omega_x}$.

Also converging almost surely to $R_\infty\mathbf{1}_{\Omega_x}$ is $(R_t\mathbf{1}_{\Omega_x})_{t\in\boldsymbol{N}}$, and since both $(\boldsymbol{E}_t\widetilde{R}_t)\mathbf{1}_{\Omega_x}$ and $R_t\mathbf{1}_{\Omega_x}$ converge almost surely to the same random variable, we have that $g(K_t;\{x\})\mathbf{1}_{\Omega_x} = \left(-R_t + \boldsymbol{E}_t\widetilde{R}_t\right)\mathbf{1}_{\Omega_x}$ converges to zero almost surely. This shows that $(\Omega\setminus\Omega_x)\cup\{\lim_t g(K_t;\{x\}) = 0\}$ is almost sure, completing the proof.

Proof of Theorem 4.1: Define four events,

$$\Omega_1 := \left\{K_t \in \widetilde{\mathcal{K}},\ \forall t \in \boldsymbol{N}\cup\{\infty\}\right\}, \quad \Omega_2 := \left\{K_\infty = \lim_t K_t \text{ exists}\right\},$$
$$\Omega_3 := \bigcap_{x\in\mathcal{X}}\left[\{N_{\infty,x} < \infty\}\cup\left\{\lim_t g(K_t;\{x\}) = 0\right\}\right],$$
$$\Omega_4 := \bigcap_{A\subseteq 2^{\mathcal{X}}}\left[\left\{\sum_{t\in\boldsymbol{N}}\Pi(t,K_t,A) < \infty\right\}\cup\left\{\sum_{t\in\boldsymbol{N}}\mathbf{1}_{\{X_{t+1}\in A\}} = \infty\right\}\right].$$

Each of these events is almost sure: Ω_1 is almost sure by the assumption of the theorem; Ω_2 is almost sure by Lemma 3.1; Ω_3 is almost sure by Lemma 3.3; and Ω_4 is almost sure by the extended Borel-Cantelli Lemma (see, e.g., [24] Corollary 6.20), since $\left\{\sum_{t\in\boldsymbol{N}}\mathbf{1}_{\{X_{t+1}\in A\}} = \infty\right\}$ is almost sure on $\left\{\sum_{t\in\boldsymbol{N}}\boldsymbol{P}_t\{X_{t+1}\in A\} = \infty\right\}$ and $\boldsymbol{P}_t\{X_{t+1}\in A\} = \Pi(t,K_t,A)$ almost surely. We then define their intersection $\Omega_0 := \Omega_1\cap\Omega_2\cap\Omega_3\cap\Omega_4$ and note that it too is almost sure.

Choose $\omega\in\Omega_0$ and suppose for contradiction that $K_\infty(\omega)\notin M_*$. Since $\omega\in\Omega_1$ implies $K_\infty(\omega)\in\widetilde{\mathcal{K}}$, (4.1) implies $K_\infty(\omega)$ has an open neighborhood U such that $\limsup_t c_t < 1$, where $c_t := \sup_{k'\in U\cap\widetilde{\mathcal{K}}\cap\mathcal{K}}\Pi(t,k',A_{K_\infty(\omega)})$.

Since $\lim_t K_t(\omega) = K_\infty(\omega)$, there exists a $t'\in\boldsymbol{N}$ such that $K_t(\omega)\in U$ for all $t > t'$. Furthermore, for all $t > t'$ the finiteness of t, together with $\omega\in\Omega_1$, implies $K_t(\omega)\in U\cap\widetilde{\mathcal{K}}\cap\mathcal{K}$. Thus

$$\sum_{t\in\boldsymbol{N}}\Pi(t,K_t(\omega),\mathcal{X}\setminus A_{K_\infty(\omega)}) \geq \sum_{t>t'}\Pi(t,K_t(\omega),\mathcal{X}\setminus A_{K_\infty(\omega)}) \geq \sum_{t>t'}(1-c_t) = \infty,$$

where the final equality with infinity is due to $\limsup_t c_t < 1$.

Since $\omega\in\Omega_4$, this implies $\sum_{t\in\boldsymbol{N}}\mathbf{1}_{\{X_{t+1}(\omega)\notin A_{K_\infty(\omega)}\}} = \infty$. In particular, since $\mathcal{X}$ is finite, there must exist some $x\notin A_{K_\infty(\omega)}$ satisfying $\sum_{t\in\boldsymbol{N}}\mathbf{1}_{\{X_{t+1}(\omega)=x\}} = \infty$. Finally, $\omega\in\Omega_3$ implies $\lim_t g(K_t(\omega);\{x\}) = 0$, implying that $K_\infty(\omega)\in M_x$, and $x\in A_{K_\infty(\omega)}$. This contradicts $x\notin A_{K_\infty(\omega)}$.

This contradiction shows that $K_\infty(\omega)\in M_*$ for all $\omega\in\Omega_0$, implying $\lim_t g(K_t;\{x\}) = 0$ almost surely for all $x\in\mathcal{X}$. This, together with Assumption 1 and Lemma 3.2, implies consistency.

Proof of Lemma 5.1: Since μ_i is $\boldsymbol{P}^{(k)}$ integrable for each $k\in\mathcal{K}$, the inequality $R(\theta;i) = \max_j\mu_j - \mu_i \leq \sum_j|\mu_j|$ shows that $R(\theta;i)$ is also $\boldsymbol{P}^{(k)}$ integrable for each $k\in\mathcal{K}$, and $\mathrm{dom}(g) = \mathcal{K}$.

We now prove the statement about continuity of g. For $k \in \mathrm{dom}(g)$ and $C \subseteq \mathcal{X}$,

$$g(k;C) = \min_i \boldsymbol{E}^{(k)}\left[R(\theta;i)\right] - \boldsymbol{E}^{(k)}\left[\min_i \boldsymbol{E}^{(k)}\left[R(\theta;i) \mid (\mu_j,\lambda_j), j \in \mathcal{I}_C\right]\right].$$

Noting the $\boldsymbol{P}^{(k)}$ integrability of $\max_{j'} \mu_{j'}$, and the two relations $\min_i \boldsymbol{E}^{(k)}\left[R(\theta;i)\right] = \boldsymbol{E}^{(k)}\left[\max_{j'} \mu_{j'}\right] - \max_i \hat{\mu}_i(k)$, and $\boldsymbol{E}^{(k)}\left[\min_i \boldsymbol{E}^{(k)}\left[R(\theta;i) \mid (\mu_j,\lambda_j), j \in \mathcal{I}_C\right]\right] = \boldsymbol{E}^{(k)}\left[\max_{j'} \mu_{j'}\right] - \boldsymbol{E}^{(k)}\left[\max_i \boldsymbol{E}^{(k)}\left[\mu_i \mid (\mu_j,\lambda_j), j \in \mathcal{I}_C\right]\right]$, we have

$$\begin{aligned} g(k;C) &= \boldsymbol{E}^{(k)}\left[\max_i \boldsymbol{E}^{(k)}\left[\mu_i \mid (\mu_j,\lambda_j), j \in \mathcal{I}_C\right]\right] - \max_i \hat{\mu}_i(k) \\ &= \boldsymbol{E}^{(k)}\left[\max\left(\max_{i \in \mathcal{I}_C} \mu_i, \max_{i \notin \mathcal{I}_C} \hat{\mu}_i(k)\right)\right] - \max_{i \in \mathcal{I}} \hat{\mu}_i(k). \end{aligned} \tag{7.1}$$

This expression agrees with (5.2) for $k \in \mathrm{dom}(g)$, and so it only remains to show that (5.2) is continuous on $\mathrm{cl}(\mathrm{dom}(g))$.

Let $f(\theta,k) = \max(\max_{i \in \mathcal{I}_C} \mu_i, \max_{i \notin \mathcal{I}_C} \hat{\mu}_i(k))$. Since $k \mapsto \hat{\mu}_i(k)$ is continuous and finite on $\mathrm{cl}(\mathrm{dom}(g))$, to show continuity of (5.2) it is sufficient to show continuity of $k \mapsto \boldsymbol{E}^{(k)}\left[f(\theta,k)\right]$ on $\mathrm{cl}(\mathcal{K})$. To this end, let $(k_t) \subseteq \mathrm{cl}(\mathrm{dom}(g))$ be a sequence converging to $k_* \in \mathrm{cl}(\mathrm{dom}(g))$. We show $\lim_t \boldsymbol{E}^{(k_t)}\left[f(\theta,k_t)\right] = \boldsymbol{E}^{(k_*)}\left[f(\theta,k_*)\right]$ through Lebesgue's dominated convergence theorem together with the identity

$$\boldsymbol{E}^{(k)}\left[f(\theta,k)\right] = \int_{\boldsymbol{R}_+} \boldsymbol{P}^{(k)}\left\{f(\theta,k) > u\right\} - \boldsymbol{P}^{(k)}\left\{f(\theta,k) < -u\right\}\, du. \tag{7.2}$$

The integrand in (7.2) is bounded above by $\boldsymbol{P}^{(k)}\left\{|f(\theta,k)| > u\right\}$, which is in turn bounded above by

$$\begin{aligned} \boldsymbol{P}^{(k)}\left\{\max\left(\max_{i \in \mathcal{I}_C} |\mu_i|, \max_{i \notin \mathcal{I}_C} |\hat{\mu}_i(k)|\right) > u\right\} &= 1 - \left[\prod_{i \in \mathcal{I}_C} \boldsymbol{P}^{(k)}\left\{|\mu_i| \le u\right\}\right]\left[\prod_{i \notin \mathcal{I}_C} \mathbf{1}_{\{|\hat{\mu}_i(k)| \le u\}}\right] \\ &\le \begin{cases} 1, & \text{if } \max_i |\hat{\mu}_i(k)| \ge u, \\ 1 - \prod_{i \in \mathcal{I}_C} \boldsymbol{P}^{(k)}\left\{|\mu_i| \le u\right\}, & \text{otherwise.} \end{cases} \end{aligned}$$

Construct $\tilde{k}$ so that $\hat{\mu}_i(\tilde{k}) = \inf_t |\hat{\mu}_i(k_t)|$, $\sigma_i^2(\tilde{k}) = \sup_t \sigma_i^2(k_t)$ and note that $\tilde{k} \in \widetilde{\mathcal{K}}$.

For $u > |\hat{\mu}_i(k)|$, $\boldsymbol{P}^{(k)}\left\{|\mu_i| \le u\right\}$ is decreasing in $|\hat{\mu}_i(k)|$ and increasing in $\sigma_i^2(k)$, so for all $u \ge \bar{u} := \max_{i \in \mathcal{I}_C} \sup_t |\hat{\mu}_i(k_t)|$ we have

$$1 - \prod_{i \in \mathcal{I}_C} \boldsymbol{P}^{(k_t)}\left\{|\mu_i| \le u\right\} \le 1 - \prod_{i \in \mathcal{I}_C} \boldsymbol{P}^{(\tilde{k})}\left\{|\mu_i| \le u\right\} = \boldsymbol{P}^{(\tilde{k})}\left\{\max_{i \in \mathcal{I}_C} |\mu_i| > u\right\}.$$

Thus, the integrand in (7.2) is dominated by $\mathbf{1}_{\{u<\bar{u}\}} + \mathbf{1}_{\{u \ge \bar{u}\}} \boldsymbol{P}^{(\tilde{k})}\left\{\max_{i \in \mathcal{I}_C} |\mu_i| > u\right\}$, a function whose integral over $\boldsymbol{R}_+$ is given by

$$\bar{u} + \int_{[\bar{u},\infty)} \boldsymbol{P}^{(\tilde{k})}\left\{\max_{i \in \mathcal{I}_C} |\mu_i| > u\right\}\, du \le \bar{u} + \boldsymbol{E}^{(\tilde{k})}\left[\max_{i \in \mathcal{I}_C} |\mu_i|\right] \le \bar{u} + \sum_{i \in \mathcal{I}_C} \boldsymbol{E}^{(\tilde{k})}\left[|\mu_i|\right],$$

which is finite because $\bar{k} \in \widetilde{\mathcal{K}}$. Thus, by the dominated convergence theorem,

$$\lim_t \boldsymbol{E}^{(k_t)}\left[f(\theta,k_t)\right] = \int_{\boldsymbol{R}_+} \lim_t \boldsymbol{P}^{(k_t)}\left\{f(\theta,k_t) > u\right\} - \boldsymbol{P}^{(k_t)}\left\{f(\theta,k_t) < -u\right\}\, du, \tag{7.3}$$

provided that the limit of the integrand exists for all but at most countably many u.

Note that there at most countably many $u \in \boldsymbol{R}$ with $\boldsymbol{P}^{(k_*)}\{|f(\theta,k_*)| = u\} > 0$, and choose any other u. We show $\lim_t \boldsymbol{P}^{(k_t)}\{f(\theta,k_t) > u\} = \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) > u\}$, and a similar argument may be used to show $\lim_t \boldsymbol{P}^{(k_t)}\{f(\theta,k_t) < -u\} = \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) < -u\}$. By the triangle inequality, the quantity $\lim_t \left|\boldsymbol{P}^{(k_t)}\{f(\theta,k_t) > u\} - \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) > u\}\right|$ is bounded above by the sum of $\lim_t \left|\boldsymbol{P}^{(k_t)}\{f(\theta,k_*) > u\} - \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) > u\}\right|$ and $\lim_t \left|\boldsymbol{P}^{(k_t)}\{f(\theta,k_t) > u\} - \boldsymbol{P}^{(k_t)}\{f(\theta,k_*) > u\}\right|$.

Since $\boldsymbol{P}^{(k_*)}\{f(\theta,k_*) = u\} = 0$, $\boldsymbol{P}^{(k_t)}$ converges weakly to $\boldsymbol{P}^{(k_*)}$, and f is continuous, $\{\theta : f(\theta,k_*) > 0\}$ is an open set whose boundary has measure 0 under $\boldsymbol{P}^{(k_*)}$. This implies $\lim_t |\boldsymbol{P}^{(k_t)}\{f(\theta,k_*) > u\} - \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) > u\}| = 0$ via the Portmanteau theorem ([5]).

Now consider $\lim_t \left|\boldsymbol{P}^{(k_t)}\{f(\theta,k_t) > u\} - \boldsymbol{P}^{(k_t)}\{f(\theta,k_*) > u\}\right|$. Choose any $\epsilon > 0$ and let $\delta > 0$ be small enough that $\boldsymbol{P}^{(k_*)}\{|f(\theta,k_*) - u| < \delta\} < \epsilon$ and $\boldsymbol{P}^{(k_*)}\{|f(\theta,k_*) - u| = \delta\} = 0$.

Since f is uniformly continuous, there exists a $\delta' > 0$ such that, for all $k \in \widetilde{\mathcal{K}}$ satisfying $||k - k_*|| < \delta'$, we have $|f(\theta,k) - f(\theta,k_*)| < \delta$. Here, $||\cdot||$ is the Euclidean norm. Then, since $\lim_t k_t = k_*$, there exists a T such that $||k_t - k_*|| < \delta'$ for all $t > T$. Then, for all $t > T$, when θ satisfies $f(\theta,k_*) > u + \delta$, it also satisfies $f(\theta,k_t) > u$, and when θ satisfies $f(\theta,k_*) < u - \delta$ it also satisfies $f(\theta,k_t) < u$. Thus,

$$\lim_t |\boldsymbol{P}^{(k_t)}\{f(\theta,k_t) > u\} - \boldsymbol{P}^{(k_t)}\{f(\theta,k_*) > u\}| \le \lim_t \boldsymbol{P}^{(k_t)}\{|f(\theta,k_*) - u| < \delta\}.$$

Since $\boldsymbol{P}^{(k_*)}\{|f(\theta,k_*) - u| = \delta\} = 0$, the limit $\lim_t \boldsymbol{P}^{(k_t)}\{|f(\theta,k_*) - u| < \delta\}$ is equal to $\boldsymbol{P}^{(k_*)}\{|f(\theta,k_*) - u| < \delta\}$, again by the Portmanteau theorem, and this is strictly less than ϵ by assumption. Thus we have shown for every $\epsilon > 0$ that $\lim_t |\boldsymbol{P}^{(k_t)}\{f(\theta,k_t) > u\} - \boldsymbol{P}^{(k_t)}\{f(\theta,k_*) > u\}| < \epsilon$, and so this limit must equal 0.

We have shown that $\lim_t \boldsymbol{P}^{(k_t)}\{f(\theta,k_t) > u\} = \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) > u\}$ for all but countably many u. It can be shown similarly, for all but countably many u, that $\lim_t \boldsymbol{P}^{(k_t)}\{f(\theta,k_t) < -u\} = \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) < -u\}$. Thus, by (7.3),

$$\lim_t \boldsymbol{E}^{(k_t)}[f(\theta,k_t)] = \int_{\boldsymbol{R}_+} \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) > u\} - \boldsymbol{P}^{(k_*)}\{f(\theta,k_*) < -u\}\, du = \boldsymbol{E}^{(k_*)}[f(\theta,k_*)].$$

This shows the continuity of (5.2).

Proof of Lemma 5.2: First suppose $\mathcal{I}_C \subseteq \mathcal{I}(k)$. Then, $\max_{i\in\mathcal{I}_C} \mu_i = \max_{i\in\mathcal{I}_C} \hat{\mu}_i(k)$ almost surely under $\boldsymbol{P}^{(k)}$, together with (5.2) from Lemma 5.1, imply that $g(k;C) = 0$.

Now suppose $\mathcal{I}_C \setminus \mathcal{I}(k)$ is nonempty, and let i' be one of its elements. Then

$$g(k;C) \ge \boldsymbol{E}^{(k)}\left[\max\left(\mu_{i'}, \max_{i\neq i'} \hat{\mu}_i(k)\right)\right] - \max_{i\in\mathcal{I}} \hat{\mu}_i(k), \tag{7.4}$$

where we use the tower property on (5.2) to introduce an inner expectation conditioned on $\mu_{i'}$, then exchange it with the maximums using Jensen's inequality, and remove it using the tower property. We now consider two cases. If $\hat{\mu}_{i'}(k) \le \max_{i\neq i'} \hat{\mu}_i(k)$, then $\max_{i\neq i'} \hat{\mu}_i(k) = \max_{i\in\mathcal{I}} \hat{\mu}_i(k)$ and we have through (7.4) that

$$g(k;C) \ge \boldsymbol{P}^{(k)}\left\{\mu_{i'} > \max_{i\neq i'} \hat{\mu}_i(k) + 1\right\} + \max_{i\neq i'} \hat{\mu}_i(k) - \max_{i\in\mathcal{I}} \hat{\mu}_i(k) > 0.$$

If $\hat{\mu}_{i'}(k) > \max_{i \neq i'} \hat{\mu}_i(k)$, then $\hat{\mu}_{i'}(k) = \max_{i \in \mathcal{I}} \hat{\mu}_i(k)$ and we add and subtract $\mu_{i'}$ to the term in the expectation in (7.4) to obtain

$$\begin{aligned} g(k;C) &\geq \boldsymbol{E}^{(k)} \left[\mu_{i'} + \max \left(0, -\mu_{i'} + \max_{i \neq i'} \hat{\mu}_i(k) \right) \right] - \max_{i \in \mathcal{I}} \hat{\mu}_i(k), \\ &= \boldsymbol{E}^{(k)} \left[\max \left(0, -\mu_{i'} + \max_{i \neq i'} \hat{\mu}_i(k) \right) \right] \geq \boldsymbol{P}^{(k)} \left\{ -\mu_{i'} + \max_{i \neq i'} \hat{\mu}_i(k) > 1 \right\} > 0. \end{aligned}$$

Acknowledgments. This research was supported in part by grant AFOSR contract FA9550-08-1-0195 and the National Science Foundation grant CMMI-0856153.

REFERENCES

[1] D.A. Berry and B. Fristedt, *Bandit Problems: Sequential Allocation of Experiments*, Chapman & Hall, London, 1985.
[2] D.P. Bertsekas and J.N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.
[3] J.E. Bickel and J.E. Smith, *Optimal sequential exploration: A binary learning model*, Decision Analysis, 14 (2006), p. 16.
[4] Peter J. Bickel and Kjell A. Doksum, *Mathematical Statistics - Basic Ideas and Selected Topics Volume 1*, Prentice Hall, Upper Saddle River, NJ, 2001.
[5] Patrick Billingsley, *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, second ed., 1999. A Wiley-Interscience Publication.
[6] J. Calvin and A. Žilinskas, *On the convergence of the P-algorithm for one-dimensional global optimization of smooth functions*, Journal of Optimization Theory and Applications, 102 (1999), pp. 479–495.
[7] C.H. Chen, *An effective approach to smartly allocate computing budget for discrete event simulation*, IEEE Conference on Decision and Control, 34 th, New Orleans, LA, (1995), pp. 2598–2603.
[8] ———, *A lower bound for the correct subset-selection probability and its application to discrete-event system simulations*, IEEE Transactions on Automatic Control, 41 (1996), pp. 1227–1231.
[9] C.H. Chen, L. Dai, and H.C. Chen, *A gradient approach for smartly allocating computing budget for discrete event simulation*, in Proceedings of the 1996 Winter Simulation Conference, J.M. Charnes, D.J. Morrice, D.T. Brunner, and J.J. Swain, eds., Piscataway, NJ, 1996, IEEE, pp. 398–405.
[10] S.E. Chick and K. Inoue, *New two-stage and sequential procedures for selecting the best simulated system*, Operations Research, 49 (2001), pp. 732–743.
[11] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker, *Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments*, Journal of the American Statistical Association, 86 (1991), pp. 953–963.
[12] J. L. Doob, *Application of the theory of martingales*, in Le Calcul des Probabilités et ses Applications., Colloques Internationaux du Centre National de la Recherche Scientifique, no. 13, Centre National de la Recherche Scientifique, Paris, 1949, pp. 23–27.
[13] B. H. Eichhorn and S. Zacks, *Sequential search of an optimal dosage. I*, J. Amer. Statist. Assoc., 68 (1973), pp. 594–598.
[14] P.I. Frazier, *Wiley Encyclopedia of Operations Research and Management Science*, Wiley, 2010, ch. Learning with Dynamic Programming.
[15] P. Frazier and W.B. Powell, *The knowledge-gradient stopping rule for ranking and selection*, Winter Simul. Conf. Proc., 2008, (2008).
[16] P. Frazier, W. B. Powell, and S. Dayanik, *A knowledge gradient policy for sequential information collection*, SIAM Journal on Control and Optimization, 47 (2008).
[17] ———, *The knowledge gradient policy for correlated normal beliefs*, INFORMS Journal on Computing, (2009).
[18] JK Ghosh and RV Ramamoorthi, *Bayesian Nonparametrics*, Springer, 2003.
[19] J.C. Gittins, *Multi-Armed Bandit Allocation Indices*, John Wiley and Sons, New York, 1989.
[20] P. Glynn and S. Juneja, *A large deviations perspective on ordinal optimization*, in Proceedings of the 36th conference on Winter simulation, Winter Simulation Conference, 2004, pp. 577–585.

[21] S.S. Gupta and K.J. Miescke, *Bayesian look ahead one-stage sampling allocations for selection of the best population*, Journal of statistical planning and inference, 54 (1996), pp. 229–244.
[22] DH He, SE Chick, and CH Chen, *Opportunity cost and ocba selection procedures in ordinal optimization for a fixed number of alternative systems*, IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews, 37 (2007), pp. 951–961.
[23] RA Howard, *Information Value Theory*, Systems Science and Cybernetics, IEEE Transactions on, 2 (1966), pp. 22–26.
[24] Olav Kallenberg, *Foundations of Modern Probability*, Springer, New York, 1997.
[25] A. Kapoor and R. Greiner, *Learning and Classifying Under Hard Budgets*, in 16th European Conference on Machine Learning, Porto, Portugal, October 3-7, 2005, Springer-Verlag New York Inc, 2005, pp. 170–181.
[26] H. Kushner, *A versatile stochastic model of a function of unknown and time varying form*, Journal of Mathematical Analysis and Applications, 5 (1962), pp. 150–167.
[27] TL Lai and H. Robbins, *Asymptotically efficient adaptive allocation rules*, Advances in applied mathematics, 6 (1985), pp. 4–22.
[28] M. Locatelli, *Bayesian Algorithms for One-Dimensional Global Optimization*, Journal of Global Optimization, 10 (1997), pp. 57–76.
[29] J. Mockus, V. Tiesis, and A. Zilinskas, *The application of Bayesian methods for seeking the extremum*, in Towards Global Optimisation, L.C.W. Dixon and G.P. Szego, eds., vol. 2, Elsevier Science Ltd., North Holland, Amsterdam, 1978, pp. 117–129.
[30] R.D. Rimey and C.M. Brown, *Control of selective perception using bayes nets and decision theory*, International Journal of Computer Vision, 12 (1994), pp. 173–207.
[31] W. Ruml, *Adaptive tree search*, PhD thesis, Harvard, 2002.
[32] A. Törn and A. Žilinskas, *Global Optimization (or Lecture Notes in Computer Science; Vol. 350)*, Springer-Verlag, Berlin, 1989.
[33] E. Vazquez and J. Bect, *On the convergence of the expected improvement algorithm*. Arxiv preprint arXiv:0712.3744, 2008.
[34] A.G. Zilinskas, *Single-step Bayesian search method for an extremum of functions of a single variable*, Cybernetics and Systems Analysis, 11 (1975), pp. 160–166.