ADVANCES IN APPLIED MATHEMATICS **17,** 122–142 (1996)
ARTICLE NO. 0007

Optimal Adaptive Policies for Sequential Allocation Problems

Apostolos N. Burnetas

Department of Operations Research, Case Western Reserve University, Cleveland, Ohio 44106-7235

and

Michael N. Katehakis

GSM and RUTCOR, Rutgers University, Newark, New Jersey 07102-1895

Received March 10, 1995

Consider the problem of sequential sampling from m statistical populations to maximize the expected sum of outcomes in the long run. Under suitable assumptions on the unknown parameters $\underline{\theta} \in \Theta$, it is shown that there exists a class C_R of adaptive policies with the following properties: (i) The expected n horizon reward $V_n^{\pi^0}(\underline{\theta})$ under any policy π^0 in C_R is equal to $n\mu^*(\underline{\theta}) - M(\underline{\theta})\log n + o(\log n)$, as $n \to \infty$, where $\mu^*(\underline{\theta})$ is the largest population mean and $M(\underline{\theta})$ is a constant. (ii) Policies in C_R are asymptotically optimal within a larger class C_{UF} of "uniformly fast convergent" policies in the sense that $\overline{\lim}_{n\to\infty}(n\mu^*(\underline{\theta}) - V_n^{\pi^0}(\underline{\theta}))/(n\mu^*(\underline{\theta}) - V_n^{\pi}(\underline{\theta})) \le 1$, for any $\pi \in C_{\mathrm{UF}}$ and any $\underline{\theta} \in \Theta$ such that $M(\underline{\theta}) > 0$. Policies in C_R are specified via easily computable indices, defined as unique solutions to dual problems that arise naturally from the functional form of $M(\underline{\theta})$. In addition, the assumptions are verified for populations specified by nonparametric discrete univariate distributions with finite support. In the case of normal populations with unknown means and variances, we leave as an open problem the verification of one assumption.

1. INTRODUCTION

Consider for any $a = 1, \ldots, m$ the i.i.d. random variables Y_a, Y_{aj}, $j = 1, 2, \ldots$ with univariate density function $f_a(y, \underline{\theta}_a)$, with respect to some known measure ν_a, where $\underline{\theta}_a$ is a vector of unknown parameters $(\theta_{a1}, \ldots, \theta_{ak_a})$. For each a, k_a is known and the vector $\underline{\theta}_a$ belongs to some

0196-8858/96 $18.00

known set Θ_a that in general depends on a and is a subset of $\mathbb{R}^{k_a}$. The functional form of $f_a(\cdot,\cdot)$ is known and allowed to depend on a. The information specified by $(\Theta_a, f_a(\cdot,\cdot), \nu_a)$ is said in the literature to define population $a = 1,\ldots,m$. The practical interpretation of the model is that Y_{aj} represents a reward received the jth time population a is sampled. The objective is to determine an adaptive rule for sampling from the m populations so as to maximize the sum of realized rewards $S_n = X_0 + X_1 + \cdots + X_{n-1}$ $n \to \infty$, where X_t is Y_{ak} if at time t population a is sampled for the kth time.

In their paper [25] on this problem, Lai and Robbins give a method for constructing adaptive allocation policies that converge *fast* to an optimal one under complete information and possess the remarkable property that their finite horizon expected loss due to ignorance (''regret'') attains, asymptotically, a minimum value. The analysis was based on a theorem (Theorem 1) establishing the existence of an asymptotic lower bound for the regret of any policy in a certain class of candidate policies; see UF policies below. The knowledge of the functional form of this lower bound was used to construct, via suitably defined ''upper confidence bounds'' for the sample means of each population, adaptive allocation policies that attain it.

The assumptions that they made for the partial information model restricted the applicability of the method to the case in which each population is specified by a density that depends on a *single* unknown parameter, as is the case of a single parameter exponential family.

The contributions in this paper are the following. (a) It is shown that Theorem 1 holds, under no parametric assumptions, for a suitable unique extension of the coefficient in the lower bound; see Theorem 1 (1), below. (b) We give the explicit form of a new set of indices that are defined as the unique solutions to dual problems that arise naturally from the definition of the (new) lower bound. (c) We give sufficient conditions under which the adaptive allocation policies that are defined by these indices possess the optimality properties of Theorem 1 (2), below. (d) We show that the sufficient conditions hold for an arbitrary nonparametric, discrete, univariate distribution. (e) We discuss the problem of normal populations with unknown variance, where we leave as an open problem the verification of one sufficient condition.

We first discovered the form of the indices used in this paper when we employed the dynamic programming approach to study a Bayes version of this problem [6, 7]. The ideas involved in the present paper are a natural extension of [25]; they are essentially a simplification of work in [8] on dynamic programming.

Our work is related to that of [33], which obtained adaptive policies with regret of order $O(\log n)$, as in our Theorem 1, for general nonparametric

models, under appropriate assumptions on the rate of convergence of the estimates.

Starting with [31, 3], the literature on versions of this problem is large; see [24, 26, 22, 4, 15, 17–20] for work on the so-called multiarmed bandit problem and [16, 29, 30, 1, 12, 27, 14, 5, 2, 28] for more general dynamic programming extensions. For a survey see also [18].

2. THE PARTIAL INFORMATION MODEL

The statistical framework used below is as follows. For any population a let $(\mathscr{Y}_a^{(n)}, \mathscr{B}_a^{(n)})$ denote the sample space of a sample of size n: $(Y_{a1}, \dots, Y_{an})$, $1 \le n \le \infty$. For each $\underline{\theta}_a \in \Theta_a$, let $\mathbf{P}_{\underline{\theta}_a}$ be the probability measure on $\mathscr{B}_a^{(1)}$ generated by $f_a(y, \underline{\theta}_a)$ and ν_a and $\mathbf{P}_{\underline{\theta}_a}^{(n)}$ the measure on $\mathscr{B}_a^{(n)}$ generated by n independent replications of Y_a. In what follows, $\mathbf{P}_{\underline{\theta}_a}^{(n)}$ will often be abbreviated as $\mathbf{P}_{\underline{\theta}_a}$. The joint (product) sample space for the m populations will be denoted by $(\mathscr{Y}^{(n)}, \mathscr{B}^{(n)})$ and the probability measure on $\mathscr{B}^{(n)}$ will be denoted by $\mathbf{P}_{\underline{\theta}}^{(n)}$ and will be abbreviated as $\mathbf{P}_{\underline{\theta}}$, where $\underline{\theta} = (\underline{\theta}_1, \dots, \underline{\theta}_m) \in \Theta := \Theta_1 \times \cdots \times \Theta_m$.

2.1. Sample Paths—Adaptive Policies and Statistics

Let $A_t, X_t, t = 0, 1, \dots$ denote respectively the action taken (i.e., population sampled) and the outcome observed at period t. A *history* or *sample path* at time n is any feasible sequence of actions and observations during the first n time periods, i.e., $\omega_n = a_0, x_0, \dots a_{n-1}, x_{n-1}$. Let $(\Omega^{(n)}, \mathscr{F}^{(n)})$, $1 \le n \le \infty$, denote the sample space of the histories ω_n, where $\Omega^{(n)}$ is the set of all histories ω_n and $\mathscr{F}^{(n)}$ the σ-field generated by $\Omega^{(n)}$. *Events* will be defined on $\mathscr{F}^{(n)}$ or on $\mathscr{B}_a^{(n)}$ and will be denoted by capital letters. The complement of event B will denoted by $\bar{B}$.

A *policy* π represents a generally randomized rule for selecting actions (populations) based on the observed history, i.e., π is a sequence $\{\pi_0, \pi_1, \dots\}$ of history-dependent probability measures on the set of populations $\{1, \dots, m\}$ so that $\pi_t(a) = \pi_t(a, \omega_t)$ is the probability that policy π selects population a at time t when the observed history is ω_t. Any policy π generates a probability measure on $\mathscr{F}^{(n)}$ that will be denoted by $\mathbf{P}_{\underline{\theta}}^{\pi}$ (cf. [10], p. 47]). Let C denote the set of all policies. Expectation under a policy $\pi \in C$ will be denoted by $\mathbf{E}_{\underline{\theta}}^{\pi}$. For notational convenience we may use π_t to denote also the action selected by a policy π at time t.

Given the history ω_n, let $T_n(a)$ denote the number of times population a has been sampled, $T_n(a) := \sum_{t=0}^{n-1} \mathbf{1}\{\pi_t = a\}$. Finally, assume that there are estimators $\hat{\underline{\theta}}_a^{T_n(a)} = g_a(Y_{a1}, \dots, Y_{aT_n(a)}) \in \Theta_a$ for $\underline{\theta}_a$. The initial estimates $\hat{\underline{\theta}}_a^0$

are arbitrary, unless otherwise specified. Properties of the estimators are given by conditions (A2) and (A3) below.

Remark 1. Note the distinction between the policy-dependent $(\Omega^{(n)}, \mathscr{F}^{(n)}, \mathbf{P}_{\underline{\underline{\theta}}}^{\pi})$ and policy-independent $(\gamma_a^{(n)}, \mathscr{B}_a^{(n)}, \mathbf{P}_{\underline{\theta}_a}^{(n)})$ probability spaces, see also [33]. However, since $\hat{\underline{\theta}}_a^j$ is a function of $Y_{a1}, \ldots, Y_{aj}$ only, it is easy to see by conditioning that the following type of relations hold, for any sequence of subsets F_{naj} of Θ_a, $n, j \geq 1$:

$$\mathbf{P}_{\underline{\underline{\theta}}}^{\pi}\left(\hat{\underline{\theta}}_a^{T_n(a)} \in F_{naT_n(a)}, T_n(a) = j\right) \leq \mathbf{P}_{\underline{\theta}_a}\left(\hat{\underline{\theta}}_a^j \in F_{naj}\right). \tag{2.1}$$

$$\mathbf{P}_{\underline{\underline{\theta}}}^{\pi}\left(\hat{\theta}_a^{T_n(a)} \in F_{naT_n(a)}\right) \leq \mathbf{P}_{\underline{\theta}_a}\left(\hat{\underline{\theta}}_a^j \in F_{naj} \text{ for some } j \leq n\right). \tag{2.2}$$

2.2. *Unobservable Quantities*

We next list notation regarding the unobservable quantities such as the population means μ_a, the Kullback–Leibler information number $\mathbf{I}(\underline{\theta}_a, \underline{\theta}_a')$, the set of "optimal" populations, $\mathbf{O}(\underline{\underline{\theta}})$ for any parameter value $\underline{\underline{\theta}}$, the subset $\mathbf{\Delta}\Theta_a(\underline{\theta}_a)$ of the parameter space Θ_a that consists of all parameter values for which population a is uniquely optimal (henceforth called *critical*), the minimum discrimination information for the hypothesis that population a is critical $\mathbf{K}_a(\underline{\underline{\theta}})$, analogous quantities for $\mu^*(\underline{\underline{\theta}}) - \varepsilon$, $\mathbf{\Delta}\Theta_a(\underline{\theta}_a, \varepsilon)$, and $\mathbf{J}_a(\underline{\theta}_a, \underline{\underline{\theta}}; \varepsilon)$, for any $\varepsilon > 0$, the set of all critical populations $\mathbf{B}(\underline{\underline{\theta}})$, and the parameter space constant $\mathbf{M}(\underline{\underline{\theta}})$ as follows:

(1) (a) $\mu_a(\underline{\theta}_a) := \mathbf{E}_{\underline{\theta}_a} Y_a$,
(b) $\mathbf{I}(\underline{\theta}_a, \underline{\theta}_a') := \mathbf{E}_{\underline{\theta}_a} \log(f_a(Y_a; \underline{\theta}_a)/f_a(Y_a; \underline{\theta}_a'))$,

(2) (a) $\mu^* = \mu^*(\underline{\underline{\theta}}) := \max_{a=1,\ldots,m}\{\mu_a(\underline{\theta}_a)\}$,
(b) $\mathbf{O}(\underline{\underline{\theta}}) := \{a: \mu_a = \mu^*(\underline{\underline{\theta}})\}$,

(3) (a) $\mathbf{\Delta}\Theta_a(\underline{\theta}_a, \varepsilon) := \{\underline{\theta}_a' \in \Theta_a: \mu_a(\underline{\theta}_a') > \mu^*(\underline{\underline{\theta}}) - \varepsilon\}$, for $\varepsilon > 0$,
(b) $\mathbf{w}_a(\underline{\theta}_a, z) := \inf\{\mathbf{I}(\underline{\theta}_a, \theta_a'): \mu_a(\underline{\theta}_a') > z\}$, for $-\infty \leq z \leq \infty$,
(c) $\mathbf{J}_a(\underline{\theta}_a, \underline{\underline{\theta}}; \varepsilon) := \mathbf{w}_a(\underline{\theta}_a, \mu^*(\underline{\underline{\theta}}) - \varepsilon) = \inf\{\mathbf{I}(\underline{\theta}_a, \underline{\theta}_a'): \underline{\theta}_a' \in \bar{\mathbf{\Delta}}\Theta_a(\underline{\theta}_a, \varepsilon)\}$, for $\varepsilon \geq 0$,

(4) (a) $\mathbf{\Delta}\Theta_a(\underline{\theta}_a) := \mathbf{\Delta}\Theta_a(\underline{\theta}_a, 0) = \{\underline{\theta}_a' \in \mathbf{\Theta}_a: \mu_a(\underline{\theta}_a') > \mu^*(\underline{\underline{\theta}})\}$,
(b) $\mathbf{B}(\underline{\underline{\theta}}) := \{a: a \notin \mathbf{O}(\underline{\underline{\theta}}) \text{ and } \mathbf{\Delta}\Theta_a(\underline{\theta}_a) \neq \varnothing\}$,

(5) (a) $\mathbf{K}_a(\underline{\underline{\theta}}) := \mathbf{J}_a(\underline{\theta}_a, \underline{\underline{\theta}}; 0) = \inf\{\mathbf{I}(\underline{\theta}_a, \underline{\theta}_a'): \underline{\theta}_a' \in \mathbf{\Delta}\Theta_a(\underline{\theta}_a)\}$, for $a \in \mathbf{B}(\underline{\underline{\theta}})$,
(b) $\mathbf{M}(\underline{\underline{\theta}}) := \sum_{a \in \mathbf{B}(\underline{\underline{\theta}})}(\mu^*(\underline{\underline{\theta}}) - \mu_a(\underline{\theta}_a))/\mathbf{K}_a(\underline{\underline{\theta}})$.

In the definition of $\mathbf{M}(\underline{\underline{\theta}})$ we have used the fact that $\mathbf{K}_a(\underline{\underline{\theta}}) \in (0, \infty)$, $\forall a \in \mathbf{B}(\underline{\underline{\theta}}) \subseteq \bar{\mathbf{O}}(\underline{\underline{\theta}})$, which is a consequence of the fact that $\mathbf{I}(\underline{\theta}_a, \underline{\theta}_a') = 0$ only when $\underline{\theta}_a = \underline{\theta}_a'$.

Under the assumptions made in[25] the constant $\mathbf{K}_a(\underline{\theta})$ reduces to $\mathbf{I}(\theta_a, \theta^*)$, thus giving the form for $\mathbf{M}(\underline{\theta})$ used in that paper.

2.3. Optimality Criteria

Let $V_n^{\pi}(\underline{\theta}) = \mathbf{E}_{\underline{\theta}}^{\pi} \sum_{t=0}^{n-1} X_t$ and $V_n(\underline{\theta}) = n\mu^*(\underline{\theta})$ denote respectively the expected total reward during the first n periods under policy π and the optimal n-horizon reward which would be achieved if the true value $\underline{\theta}$ were known to the experimenter. Let $R_n^{\pi}(\underline{\theta}) = V_n(\underline{\theta}) - V_n^{\pi}(\underline{\theta})$ represent the loss or regret, due to partial information, when policy π is used; maximization of $V_n^{\pi}(\underline{\theta})$ with respect to π is equivalent to minimization of $R_n^{\pi}(\underline{\theta})$.

In general it is not possible to find an adaptive policy that minimizes $R_n^{\pi}(\underline{\theta})$ uniformly in $\underline{\theta}$) uniformly in $\underline{\theta}$. However, if we let $g^{\pi}(\underline{\theta}) = \underline{\lim}_{n\to\infty} V_n^{\pi}(\underline{\theta})/n$, then $\overline{\lim}_{n\to\infty}(V_n(\underline{\theta}) - V_n^{\pi}(\underline{\theta}))/n = \overline{\lim}_{n\to\infty} R_n^{\pi}(\underline{\theta}) = \mu^*(\underline{\theta}) - g^{\pi}(\underline{\theta}) \geq 0$, $\forall \underline{\theta}$.

A policy π will be called *uniformly convergent* (UC) or *uniformly fast convergent* (UF) if $\forall \underline{\theta} \in \Theta$ as $n \to \infty$, $R_n^{\pi}(\underline{\theta}) = o(n)$ (for UC) or $R_n^{\pi}(\underline{\theta}) = o(n^{\alpha})$, $\forall \alpha > 0$ (for UF).

A UF Policy π^0 will be called *uniformly maximal convergence rate* (UM) if $\overline{\lim}_{n\to\infty} R_n^{\pi^0}(\underline{\theta})/R_n^{\pi}(\underline{\theta}) \leq 1$, $\forall \underline{\theta} \in \Theta$ such that $M(\underline{\theta}) > 0$, for all UF policies π. Note that according to this definition a UM policy has maximum rate of convergence only for those values of the parameter space for which $M(\underline{\theta}) > 0$; when $M(\underline{\theta}) = 0$ it is UF. Let $C_{\mathrm{UC}} \supset C_{\mathrm{UF}} \supset C_{\mathrm{UM}}$ denote the classes of UC, UF, UM policies, respectively.

3. THE MAIN THEOREM

We start by giving the explicit form of the indices $\mathbf{U}_a(\omega_k)$ which define a class of adaptive policies that will be shown to be UM under conditions (A1)–(A3) below. For $a = 1, \ldots, m$, $\underline{\theta}_a \in \Theta_a$, and $0 \leq \gamma \leq \infty$, let

$$\mathbf{u}_a(\underline{\theta}_a, \gamma) = \sup_{\underline{\theta}'_a \in \Theta_a} \{\mu_a(\underline{\theta}'_a) : \mathbf{I}(\underline{\theta}_a, \underline{\theta}'_a) < \gamma\}. \tag{3.1}$$

Given the history ω_k and the statistics $T_k(a)$, $\hat{\underline{\theta}}_a = \hat{\underline{\theta}}_a^{T_k(a)}$ for $\underline{\theta}_a$, define the index $\mathbf{U}_a(\omega_k)$ as

$$\mathbf{U}_a(\omega_k) = \mathbf{u}_a\big(\hat{\underline{\theta}}_a, \log k/T_k(a)\big), \tag{3.2}$$

for $k \geq 1$; $\mathbf{U}_a(\omega_0) = \mu_a(\hat{\underline{\theta}}_a^0)$. We assume, throughout, that, when $j = 0$ in a ratio of the form $\log k/j$, the latter is equal to ∞.

Note also that $\mathbf{U}_a(\omega_k)$ is a function of k, $T_k(a)$, and $\hat{\underline{\theta}}_a^{T_k(a)}$ only and that we allow $T_k(a) = 0$ in (3.2), in which case $\mathbf{U}_a(\omega_k) = \sup_{\underline{\theta}'_a \in \Theta_a}\{\mu_a(\underline{\theta}'_a)\}$; in applications this will be equivalent to taking some small number of samples from each population to begin with.

Remark 2. (a) For all a and ω_k, $\mathbf{U}_a(\omega_k) \geq \mu_a(\hat{\underline{\theta}}_a)$, i.e., the indices are inflations of the current estimates for the means. In addition, $\mathbf{U}_a(\omega_k)$ is increasing in k and decreasing in $T_k(a)$, thus giving higher chance for the "under sampled" actions to be selected. In the case of a one-dimensional parameter vector, they yield the same value as those in [25, 23].

(b) The analysis remains the same if in the definition of $\mathbf{U}_a(\omega_k)$ we replace $\log k/j$ by a function of the form $(\log k + h(\log k))/j$, where $h(t)$ is any function of t with $h(t) = o(t)$ as $t \to \infty$. Up to this equivalence, the index $\mathbf{U}_a(\omega_k)$ is *uniquely* defined.

(c) We note that $\mathbf{u}_a(\underline{\theta}_a, \gamma)$ and $\mathbf{w}_a(\underline{\theta}_a, z)$ are connected by the following duality relation. The condition $\mathbf{u}_a(\underline{\theta}_a, \gamma) > z$ implies $\mathbf{w}_a(\underline{\theta}_a, z) \leq \gamma$. In addition, when for $\gamma = \gamma_0$, the supremum in $\mathbf{u}_a(\underline{\theta}_a, \gamma_0)$ is attained at some $\underline{\theta}_a^0 = \underline{\theta}_a^0(\gamma_0) \in \Theta_a$ such that $\mathbf{I}(\underline{\theta}_a, \underline{\theta}_a^0) = \gamma_0$ (as is the case, for example, when $\mu_a(\underline{\theta}'_a)$ is a linear function of $\underline{\theta}'_a$), $\underline{\theta}_a^0$ also attains the infimum in $\mathbf{w}_a(\underline{\theta}_a, z_0)$ for $z_0 = \mathbf{u}_a(\underline{\theta}_a, \gamma_0)$, i.e., $\mathbf{u}_a(\underline{\theta}_a, \gamma_0) = \mu_a(\underline{\theta}_a^0) = z_0$, and $\mathbf{w}_a(\underline{\theta}_a, z_0) = \mathbf{I}(\underline{\theta}_a, \underline{\theta}_a^0) = \gamma_0$. This type of duality is well known in finance [32, p. 113].

(d) For $z \in \mathbb{R}$, let $\mathbf{W}_a(\omega_k, z) = \mathbf{w}_a(\hat{\underline{\theta}}_a^{T_n(a)}, z)$. It follows from (c) above that $\forall \omega_k$, the condition $\mathbf{U}_a(\omega_k) > z$ implies the condition $\mathbf{W}_a(\omega_k, z) \leq \log k/T_k(a)$.

Furthermore, when the supremum in $\mathbf{u}_a(\hat{\underline{\theta}}_a, \log k/T_k(a))$ is attained at some $\underline{\theta}_a^0 = \underline{\theta}_a^0(\omega_k) \in \Theta_a$ such that $\mathbf{I}(\hat{\underline{\theta}}_a, \underline{\theta}_a^0) = \log k/T_k(a)$,

$$\mathbf{U}_a(\omega_k) = \mu_a(\underline{\theta}_a^0), \tag{3.3}$$

$$\mathbf{W}_a(\omega_k, \mu_a(\underline{\theta}_a^0)) = \mathbf{I}(\hat{\underline{\theta}}_a, \underline{\theta}_a^0) = \log k/T_k(a) = \mathbf{J}_a(\hat{\underline{\theta}}_a, \underline{\bar{\theta}}; \mu^*(\underline{\bar{\theta}}) - \mu_a(\underline{\theta}_a^0)). \tag{3.4}$$

The conditions given below are *sufficient conditions* for Theorem 1 (2).

Condition A1. $\forall \underline{\theta} \in \Theta$ and $\forall a \notin \mathbf{O}(\underline{\theta})$ such that $\Delta\Theta_a(\underline{\theta}_a, 0) = \varnothing$ and $\mathbf{\Delta}\Theta_a(\underline{\theta}_a, \varepsilon) \neq \varnothing$, $\forall \varepsilon > 0$, the following relation holds: $\lim_{\varepsilon \to 0} \mathbf{J}_a(\underline{\theta}_a, \underline{\bar{\theta}}; \varepsilon) = \infty$.

Condition A2. $\mathbf{P}_{\underline{\theta}_a}(\|\hat{\underline{\theta}}_a^k - \underline{\theta}_a\| > \varepsilon) = o(1/k)$, as $k \to \infty$, $\forall \varepsilon > 0$, and $\forall \underline{\theta}_a \in \Theta_a$, $\forall a$.

Condition A3. $\mathbf{P}_{\underline{\theta}_a}(\mathbf{u}_a(\hat{\underline{\theta}}_a^j, \log k/j) < \mu(\underline{\theta}_a) - \varepsilon$, for some $j \leq k) = o(1/k)$, as $k \to \infty$, $\forall \varepsilon > 0$, $\forall \underline{\theta}_a \in \Theta_a$, $\forall a$.

Remark 3. To see the significance of condition (A1) consider the next examples.

EXAMPLE 1. Take $m = 2$, $\Theta_1 = [0, 1]$, $\Theta_2 = [0, .5]$, $\mu^*(\underline{\theta}) = \mu_1(\theta_1) = 0.5 > \mu_2(\theta_2)$, $f_a(y; \theta_a) = \theta_a^x(1 - \theta_a)^{(1-x)}$, $x = 0, 1$.

EXAMPLE 2. Take $\Theta_2 = [0, 0.51]$ in Example 1.

EXAMPLE 3. Take $\Theta_1 = \Theta_2 = [0, 1]$, $\mu^*(\underline{\theta}) = \mu_1(\theta_1) = 1 > \mu_2(\theta_2)$, in Example 1.

Situations such as in Example 1 are excluded, while Examples 2 and 3 satisfy (A1).

Remark 4. (a) Note that (A2) is a condition on the rate of convergence of $\hat{\underline{\theta}}_a^k$ to $\underline{\theta}_a$ and it holds in the usual case that $\hat{\underline{\theta}}_a^k$ is either equal to or follows the same distribution as the mean of i.i.d. random variables Z_j with finite moment generating function in a neighborhood around zero. In this case (A2) can be verified using large deviation arguments. This implies that $\sum_{k=1}^{n-1} \mathbf{P}_{\underline{\theta}_a}(\|\hat{\underline{\theta}}_a^k - \underline{\theta}_a\| > \varepsilon) = o(\log n)$, as $n \to \infty$.

(b) From the continuity of $\mathbf{I}(\underline{\theta}_a, \underline{\theta}_a')$ and, hence, of $\mathbf{J}_a(\underline{\theta}_a, \underline{\theta}; \varepsilon)$ in $\underline{\theta}_a$, it follows that the event $\{\mathbf{J}_a(\hat{\underline{\theta}}_a^k, \underline{\theta}; \varepsilon) < \mathbf{J}_a(\underline{\theta}_a, \underline{\theta}; \varepsilon) - \delta\}$ is contained in the event $\{\|\hat{\underline{\theta}}_a^k - \underline{\theta}_a\| > \eta\}$, for some $\eta = \eta(\delta) > 0$. Thus, condition (A2) implies $\mathbf{P}_{\underline{\theta}_a}[\mathbf{J}_a(\hat{\underline{\theta}}_a^k, \underline{\theta}; \varepsilon) < \mathbf{J}_a(\underline{\theta}_a, \underline{\theta}; \varepsilon) - \delta] = o(1/k)$, as $k \to \infty$. The last can be written in the form below, as required for the proof of Proposition 2 (a): $\forall \delta > 0$

$$\sum_{k=1}^{n-1} \mathbf{P}_{\underline{\theta}_a}\left[\mathbf{J}_a\left(\hat{\underline{\theta}}_a^k, \underline{\theta}; \varepsilon\right) < \mathbf{J}_a\left(\underline{\theta}_a, \underline{\theta}; \varepsilon\right) - \delta\right] = o(\log n) \qquad \text{as } n \to \infty.$$

Remark 5. Condition (A3) can be written as $\sum_{k=1}^{n-1} \mathbf{P}_{\underline{\theta}_a}[\mathbf{u}_a(\hat{\underline{\theta}}_a^j, \log k/j) < \mu(\underline{\theta}_a) - \varepsilon$, for some $j \le k] = o(\log n)$, as $n \to \infty$. It is used in this form for the proof of Proposition 2 (b).

Let C_R denote the class of policies which in every period select any action with the largest index value $\mathbf{U}_a(\omega_k) = \mathbf{u}_a(\hat{\underline{\theta}}_a, \log k/T_k(a))$. We can now state the following theorem.

THEOREM 1. (1) *For any* $\underline{\theta} \in \Theta$, $a \in \mathbf{B}(\underline{\theta})$, *such that* $\mathbf{K}_a(\underline{\theta}) \neq 0$ *the following is true,* $\forall \pi \in C_{\mathrm{UF}}$,

$$\varliminf_{n \to \infty} \mathbf{E}_{\underline{\theta}}^{\pi} T_n(a)/\log n \ge 1/\mathbf{K}_a\left(\underline{\theta}\right).$$

(2) *If conditions* (A1), (A2), *and* (A3) *hold and* $\pi^0 \in C_R$, *then*, $\forall \underline{\theta} \in \Theta$ *and* $\forall a \notin \varnothing(\underline{\theta})$,

$$\text{(a)} \quad \begin{aligned} &\overline{\lim_{n\to\infty}}\, \mathbf{E}_{\underline{\theta}}^{\pi^0} T_n(a)/\log n \le 1/\mathbf{K}_a\big(\underline{\theta}\big), && \text{if } a \in \mathbf{B}\big(\underline{\theta}\big), \\ &\overline{\lim_{n\to\infty}}\, \mathbf{E}_{\underline{\theta}}^{\pi^0} T_n(a)/\log n = 0, && \text{if } a \notin \mathbf{B}\big(\underline{\theta}\big), \end{aligned}$$

(b) $0 \le R_n^{\pi^0}(\underline{\theta}) = \mathbf{M}(\underline{\theta})\log\, n + o(\log n)$, *as* $n \to \infty$, $\forall \underline{\theta} \in \Theta$,

(c) $C_R \subset C_{\mathrm{UM}}$.

Proof. Parts (1) and (2a) are proved in Propositions (1) and (2), respectively.

For part (2b), note first that

$$\begin{aligned} R_n^{\pi}\big(\underline{\theta}\big) &= n\mu^*\big(\underline{\theta}\big) - \sum_{a=1}^{m} \mathbf{E}_{\underline{\theta}}^{\pi} \sum_{t=1}^{T_n(a)} Y_{at} \\ &= \sum_{a=1}^{m} \Big(\mu^*\big(\underline{\theta}\big) - \mu_a(\underline{\theta}_a)\Big)\mathbf{E}_{\underline{\theta}}^{\pi} T_n(a), \qquad \forall \pi \in C. \end{aligned}$$

Using the definition of $\mathbf{M}(\underline{\theta})$ in subsection 2.3, $\forall \pi^0 \in C_R$, $\forall \underline{\theta} \in \Theta$,

$$\underline{\lim_{n\to\infty}}\, R_n^{\pi^0}\big(\underline{\theta}\big)/\log n \le \mathbf{M}\big(\underline{\theta}\big), \qquad \text{(from part (2a))};$$

hence, $C_R \subseteq C_{\mathrm{UF}}$. Thus, it follows from part (1) that

$$\underline{\lim_{n\to\infty}}\, R_n^{\pi^0}\big(\underline{\theta}\big)/\log n \ge \mathbf{M}\big(\underline{\theta}\big),$$

and the proof is easy to complete, using the above observations.

To show the last chain we need only to divide both $R_n^{\pi^0}(\underline{\theta})$ and $R_n^{\pi}(\underline{\theta})$ by $\mathbf{M}(\underline{\theta})\log n$, when $\mathbf{M}(\underline{\theta}) > 0$. ∎

Remark 6. (a) It is instructive to compare the maximum expected finite horizon reward under complete information about $\underline{\theta}$ (Eq. (3.5)) with the asymptotic expression for expected finite horizon reward for a UM policy π^0, under partial information about $\underline{\theta}$ (Eq. (3.6)), established by Theorem 1:

$$V_n\big(\underline{\theta}\big) = n\mu^*\big(\underline{\theta}\big) \tag{3.5}$$

$$V_n^{\pi^0}\big(\underline{\theta}\big) = n\mu^*\big(\underline{\theta}\big) - M\big(\underline{\theta}\big)\log n + o(\log n) \qquad (\text{as } n \to \infty). \tag{3.6}$$

(b) The results of Theorem 1 can be expressed in terms of the rate of convergence of $V_n^{\pi}(\underline{\theta})/n$ to $\mu^*(\underline{\theta})$, as follows. If $\pi \in C_{\mathrm{UC}}$ then $\lim_{n\to\infty} V_n^{\pi}(\underline{\theta})/n = \mu^*(\underline{\theta})$ for all $\underline{\theta}$. No claim regarding the rate of convergence can be made. If $\pi \in C_{\mathrm{UF}}$ then it is also true that $|V_n^{\pi}(\underline{\theta})/n - \mu^*(\underline{\theta})| = o(n^{\alpha})$ for all $\underline{\theta}$ (and $\forall \alpha > 0$); therefore $V_n^{\pi}(\underline{\theta})/n$ converges to $\mu^*(\underline{\theta})$ at least as fast as $\log n/n$. The UM policy is such that for all $\underline{\theta} \in \Theta$ with $M(\underline{\theta}) > 0$, the rate of convergence of $V_n^{\pi}(\underline{\theta})$ to $\mu^*(\underline{\theta})$ is the maximum among all policies in $C_{\mathrm{UF}} \subseteq C_{\mathrm{UC}} \subseteq C$, and is equal to $M(\underline{\theta})\log n/n$.

(c) For $\underline{\theta} \in \Theta$ such that $M(\underline{\theta}) = 0$, it is true that $V_n^{\pi^0}(\underline{\theta}) = n\mu^*(\underline{\theta}) + o(\log n)$; therefore $V_n^{\pi^0}(\underline{\theta})/n$ converges to $\mu^*(\underline{\theta})$ faster than $\log n/n$. However, this does not necessarily represent the fastest rate of convergence.

In the proof of the Proposition 1 below, we use the notation $\underline{\theta}_a' := (\underline{\theta}_1, \ldots, \underline{\theta}_{a-1}, \underline{\theta}_a', \underline{\theta}_{a+1}, \ldots, \underline{\theta}_m)$, $\forall \underline{\theta}_a' \in \mathbf{\Delta}\Theta_a(\underline{\theta}_a)$, and the following remark.

Remark 7. (a) The definition of $\mathbf{\Delta}\Theta_a(\underline{\theta}_a)$ implies that if $a \in \mathbf{B}(\underline{\theta}) \neq \emptyset$, then $\mathbf{O}(\underline{\theta}_a') = \{a\}$, $\forall \underline{\theta}_a' \in \mathbf{\Delta}\Theta_a(\underline{\theta}_a)$, and thus $\mathbf{E}_{\underline{\theta}_a'}^{\pi}(n - T_n(a)) = o(n^{\alpha})$, $\forall \alpha > 0$ and $\forall \underline{\theta}_a' \in \mathbf{\Delta}\Theta_a(\underline{\theta}_a)$, $\forall \pi \in C_{\mathrm{UF}}$, the latter being a consequence of the definition of C_{UF}.

(b) Let Z_t be i.i.d. random variables such that $S_n/n = \sum_{t=1}^{n} Z_t/n$, converges a.s. ($\mathbf{P}$) to a constant μ, let b_n be an increasing sequence of positive constants such that $b_n \to \infty$, and let $\lfloor b_n \rfloor$ denote the integer part of b_n. Then $\max_{k \le \lfloor b_n \rfloor}\{S_k\}/b_n$ converges a.s. ($\mathbf{P}$) to μ and

$$\forall \delta > 0, \qquad \mathbf{P}\Big(\max_{k \le \lfloor b_n \rfloor} \{S_k\}/b_n > \mu + \delta \Big) = o(1) \qquad (\text{as } n \to \infty).$$

PROPOSITION 1. *If $\pi \in C_{\mathrm{UF}}$ then, for any $a \in \mathbf{B}(\underline{\theta}) \neq \emptyset$,*

$$\varliminf_{n\to\infty} \mathbf{E}_{\underline{\theta}}^{\pi} T_n(a)/\log n \ge 1/\mathbf{K}_a\big(\underline{\theta}\big). \tag{3.7}$$

Proof. The proof is an adaptation of the proof of Theorem 1 in [25] for the constant $\mathbf{K}_a(\underline{\theta})$. Form the Markov inequality it follows that

$$\mathbf{P}_{\underline{\theta}}^{\pi}\Big(T_n(a)/\log n \ge 1/\mathbf{K}_a\big(\underline{\theta}\big)\Big) \le \mathbf{E}_{\underline{\theta}}^{\pi} T_n(a)\mathbf{K}_a\big(\underline{\theta}\big)/\log n, \qquad \forall n > 1.$$

Thus, to show (3.7), it suffices to show that

$$\lim_{n\to\infty} \mathbf{P}_{\underline{\theta}}^{\pi}\Big(T_n(a)/\log n \ge 1/\mathbf{I}_a\big(\underline{\theta}\big)\Big) = 1$$

or, equivalently,

$$\lim_{n\to\infty} \mathbf{P}_{\underline{\underline{\theta}}}^{\pi}\Big(T_n(a)/\log n < (1-\varepsilon)/\mathbf{K}_a\big(\underline{\theta}\big)\Big) = 0, \qquad \forall \varepsilon > 0. \quad (3.8)$$

By the definition of $\mathbf{K}_a(\underline{\theta})$ we have, $\forall \delta > 0$, $\exists \underline{\theta}_a' = \underline{\theta}_a'(\delta) \in \mathbf{\Delta}\Theta_a(\underline{\theta}_a)$ such that $\mathbf{K}_a(\underline{\theta}) < \mathbf{I}(\underline{\theta}_a, \underline{\theta}_a') < (1+\delta)\mathbf{K}_a(\underline{\theta})$.

Fix such a $\delta > 0$ and $\underline{\theta}_a'$, let $\mathbf{I}^{\delta} = \mathbf{I}(\underline{\theta}_a, \underline{\theta}_a')$, and define the sets $\mathbf{A}_n^{\delta} := \{T_n(a)/\log n < (1-\delta)/\mathbf{I}^{\delta}\}$ and $\mathbf{C}_n^{\delta} := \{\log L_{T_n(a)} \le (1-\delta/2)\log n\}$, where $\log L_k = \sum_{i=1}^{k} \log(f_a(Y_{ai}; \underline{\theta}_a)/f_a(Y_{ai}; \underline{\theta}_a'))$.

We will show that $\mathbf{P}_{\underline{\underline{\theta}}}^{\pi}(\mathbf{A}_n^{\delta}) = \mathbf{P}_{\underline{\underline{\theta}}}^{\pi}(\mathbf{A}_n^{\delta}\mathbf{C}_n^{\delta}) + \mathbf{P}_{\underline{\underline{\theta}}}^{\pi}(\mathbf{A}_n^{\delta}\overline{\mathbf{C}}_n^{\delta}) = o(1)$, as $n \to \infty$, $\forall \delta > 0$. Indeed, $\mathbf{P}_{\underline{\underline{\theta}}}^{\pi}(\mathbf{A}_n^{\delta}\mathbf{C}_n^{\delta}) \le n^{1-\delta/2}\mathbf{P}_{\underline{\underline{\theta}}_a'}^{\pi}(\mathbf{A}_n^{\delta}\mathbf{C}_n^{\delta}) \le n^{1-\delta/2}\mathbf{P}_{\underline{\underline{\theta}}_a'}^{\pi}(\mathbf{A}_n^{\delta}) \le n^{1-\delta/2}\mathbf{E}_{\underline{\underline{\theta}}_a'}^{\pi}(n - T_n(a))/(n - (1-\delta)\log n/\mathbf{I}^{\delta}) = o(n^a)/(n^{\delta/2}(1\text{-}O(\log n)/n)) = o(1)$, for $a < \delta/2$.

The first inequality follows from the observation that on $\mathbf{C}_n^{\delta} \cap \{T_n(a) = k\}$ we have $f_a(Y_{a1}; \underline{\theta}_a)\cdots f_a(y_{ak}; \underline{\theta}_a) \le n^{1-\delta/2} f_a(Y_{a1}; \underline{\theta}_a')\cdots f_a(Y_{ak}; \underline{\theta}_a')$; note also that $e^{(1-\delta/2)\log n} = n^{1-\delta/2}$. The third relation is the Markov inequality and the fourth is due to Remark 7(a)) above.

To see that $\mathbf{P}_{\underline{\underline{\theta}}}^{\pi}(\mathbf{A}_n^{\delta}\overline{\mathbf{C}}_n^{\delta}) = o(1)$, note that

$$\begin{aligned}
\mathbf{P}_{\underline{\underline{\theta}}}^{\pi}\big(\mathbf{A}_n^{\delta}\overline{\mathbf{C}}_n^{\delta}\big) &\le \mathbf{P}_{\underline{\underline{\theta}}}^{\pi}\Big(\max_{k\le\lfloor b_n\rfloor}\{\log L_k\} > (1-\delta/2)\log n\Big) \\
&= \mathbf{P}_{\underline{\underline{\theta}}}^{\pi}\Big(\max_{k\le\lfloor b_n\rfloor}\{\log L_k\}/b_n > \mathbf{I}^{\delta}\big(1+\delta/(2(1-\delta))\big)\Big) \\
&\le \mathbf{P}_{\underline{\theta}_a}\Big(\max_{k\le\lfloor b_n\rfloor}\{\log L_k\}/b_n > \mathbf{I}^{\delta}\big(1+\delta/(2(1-\delta))\big)\Big),
\end{aligned}$$

where $b_n := (1-\delta)\log n/\mathbf{I}^{\delta}$ and the last inequality follows using an argument like that in Remark 1. Thus the result follows from Remark 7(b), since $\log L_k/k \to \mathbf{I}^{\delta}$ a.s. $(\mathbf{P}_{\underline{\theta}_a})$.

To complete the proof of (3.8), it suffices to notice that the choices of δ and $\underline{\theta}_a'(\delta)$ imply $(1-\delta)/\mathbf{I}^{\delta} > (1-\delta)/((1+\delta)\mathbf{K}_a(\underline{\theta})) > (1-\varepsilon)/\mathbf{K}_a(\underline{\theta})$, and $\mathbf{P}_{\underline{\underline{\theta}}}^{\pi}(T_n(a)/\log n < (1-\varepsilon)/\mathbf{K}_a(\underline{\theta})) \le \mathbf{P}_{\underline{\underline{\theta}}}^{\pi}(\mathbf{A}_n^{\delta}) = o(1)$, when $\delta < \varepsilon/(2-\varepsilon)$. ∎

To facilitate the proof of the Proposition 2 below we introduce some notation and state a remark.

{For any $\epsilon > 0$, let

$$T_n^{(1)}(a, \varepsilon) = \sum_{k=1}^{n-1} \mathbf{1}\Big(\pi_k = a, \mathbf{U}_a(\omega_k) > \mu^*\big(\underline{\underline{\theta}}\big) - \varepsilon\Big),$$

and

$$T_n^{(2)}(a,\varepsilon) = \sum_{k=1}^{n-1} \mathbf{1}\big(\pi_k = a, \mathbf{U}_a(\omega_k) \le \mu_a(\underline{\theta}_a) - \varepsilon\big).$$

Remark 8. Let Z_t be any sequence sequence of constants (or random variables) and let $t_k := \sum_{t=0}^{k-1} \mathbf{1}\{Z_t = a\}$. This definition of t_k implies that (pointwise if we have random variables)

$$\sum_{k=1}^{n-1} \mathbf{1}\{Z_k = a,\ t_k \le c\} \le c + 1.$$

Indeed, note that $\sum_{k=1}^{n-1} \mathbf{1}\{Z_k = a,\ t_k = i\} \le 1,\ \forall i = 0,\dots,\lfloor c \rfloor$. Therefore, $\sum_{k=1}^{n-1} \mathbf{1}\{Z_k = a,\ t_k \le c\} = \sum_{k=1}^{n-1} \sum_{i=0}^{\lfloor c \rfloor} \mathbf{1}\{Z_k = a,\ t_k = i\} = \sum_{i=0}^{\lfloor c \rfloor} \sum_{k=1}^{n-1} \mathbf{1}\{Z_k = a,\ t_k = i\} \le \lfloor c \rfloor + 1 \le c + 1$.

PROPOSITION 2. *For any $\underline{\theta} \in \Theta$ the following are true*:

(a) *Under* (A1) *and* (A2), *if $\pi^0 \in C_R$ and $a \notin \mathbf{O}(\underline{\theta})$, then*

$$\lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}}\ \mathbf{E}_{\underline{\theta}}^{\pi^0} T_n^{(1)}(a,\varepsilon)/\log n \le 1/\mathbf{K}_a\big(\underline{\theta}\big), \qquad \text{if } a \in \mathbf{B}\big(\underline{\theta}\big), \quad (3.9)$$

$$\lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}}\ \mathbf{E}_{\underline{\theta}}^{\pi^0} T_n^{(1)}(a,\varepsilon)/\log n = 0, \qquad \text{if } a \notin \mathbf{B}\big(\underline{\theta}\big). \quad (3.10)$$

(b) *Under* (A3), *if $\pi^0 \in C_R$, then $\overline{\lim}_{n\to\infty} \mathbf{E}_{\underline{\theta}}^{\pi^0} T_n^{(2)}(a,\varepsilon)/\log n = 0$, $\forall a$ and $\forall \varepsilon > 0$.*

(c) *Under* (A1), (A2), *and* (A3), *if $\pi^0 \in C_R$, then $\overline{\lim}_{n\to\infty} \mathbf{E}_{\underline{\theta}}^{\pi^0} T_n(a)/\log n$ is less than or equal to $1/\mathbf{K}_a(\underline{\theta})$, if $a \in \mathbf{B}(\underline{\theta})$ and it is equal to 0, if $a \notin \mathbf{B}(\underline{\theta})$.*

Proof. (a) fix $\pi^0 \in C_R$, $\underline{\theta} \in \Theta$, $a \notin \mathbf{O}(\underline{\theta})$, i.e., $\mu^* > \mu_a(\underline{\theta}_a)$. Let $\varepsilon \in (0, \mu^* - \mu_a(\underline{\theta}_a))$, and consider two cases.

Case 1. There exists $\varepsilon_0 > 0$ such that $\Delta\Theta_a(\underline{\theta}_a, \varepsilon_0) = \varnothing$. For any $\varepsilon < \varepsilon_0$ and any $\underline{\theta}_a \in \Theta_a$ it is true that $\mu_a(\underline{\theta}_a) \le \mu^*(\underline{\theta}) - \varepsilon_0 < \mu^*(\underline{\theta}) - \varepsilon$; therefore, $T_n^{(1)}(a,\varepsilon) = 0$, for all $\varepsilon < \varepsilon_0$ and (3.10) holds.

Case 2. $\Delta\Theta_a(\underline{\theta}_a, \varepsilon) \ne \varnothing$, $\forall \varepsilon > 0$. Note that $\mathbf{J}_a(\underline{\theta}_a, \underline{\theta}; \varepsilon) > 0$, $\forall \varepsilon \in (0, \mu^* - \mu_a(\underline{\theta}_a))$. Let $\mathbf{J}_\varepsilon = \mathbf{J}_a(\underline{\theta}_a, \underline{\theta}; \varepsilon)$ and $\hat{\mathbf{J}}_\varepsilon = \mathbf{J}_a(\hat{\underline{\theta}}_a^{T_k(a)}, \underline{\theta}; \varepsilon)$; then, $\forall \delta > 0$, we have sample path-wise:

$$\begin{aligned} T_n^{(1)}(a,\varepsilon) &\le \sum_{k=1}^{n-1} \mathbf{1}\big(\pi_k^0 = a, \hat{\mathbf{J}}_\varepsilon \le \log k/T_k(a)\big) \\ &= \sum_{k=1}^{n-1} \mathbf{1}\big(\pi_k^0 = a, \hat{\mathbf{J}}_\varepsilon \le \log k/T_k(a), \hat{\mathbf{J}}_\varepsilon \ge \mathbf{J}_\varepsilon - \delta\big) \\ &\quad + \sum_{k=1}^{n-1} \mathbf{1}\big(\pi_k^0 = a, \hat{\mathbf{J}}_\varepsilon \le \log k/T_k(a), \hat{\mathbf{J}}_\varepsilon < \mathbf{J}_\varepsilon - \delta\big) \end{aligned}$$

$$\leq \sum_{k=1}^{n-1} \mathbf{1}\left(\pi_k^0 = a, T_k(a) \leq \log n/(\mathbf{J}_\varepsilon - \delta)\right)$$
$$+ \sum_{k=1}^{n-1} \mathbf{1}\left(\pi_k^0 = a, \hat{\mathbf{J}}_\varepsilon < \mathbf{J}_\varepsilon - \delta\right)$$
$$\leq \log n/(\mathbf{J}_\varepsilon - \delta) + 1 + \sum_{k=1}^{n-1} \mathbf{1}\left(\pi_k^0 = a, \hat{\mathbf{J}}_\varepsilon < \mathbf{J}_\varepsilon - \delta\right).$$

For the first inequality, we have used an immediate consequence of the "duality" relations of Remark 2(c) with $z_0 = \mu^*(\underline{\underline{\theta}}) - \varepsilon$, which imply that the event $\{\mathbf{U}_a(\omega_k) > \mu^*(\underline{\underline{\theta}}) - \varepsilon\}$ is contained in the event $\{\mathbf{J}_a(\hat{\theta}_a^{T_k(a)}, \underline{\underline{\theta}}; \varepsilon) \leq \log k/T_k(a)\}$. For the last inequality, we have used Remark 8.

Taking expectations of the first and last terms and using remark 4(b), it follows that, since δ is arbitrarily small, $\mathbf{E}_{\underline{\underline{\theta}}}^{\pi^0} T_n^{(1)}(a, \varepsilon)/\log n \leq 1 + \log n/\mathbf{J}_a(\underline{\theta}_a, \underline{\underline{\theta}}; \varepsilon) + \mathbf{E}_{\underline{\underline{\theta}}}^{\pi^0}[\sum_{k=1}^{n-1} \mathbf{1}(\pi_k^0 = a, \hat{\mathbf{J}}_\varepsilon < \mathbf{J}_\varepsilon - \delta)]$. In addition,

$$\mathbf{E}_{\underline{\underline{\theta}}}^{\pi^0}\left[\sum_{k=1}^{n-1} \mathbf{1}\left(\pi_k^0 = a, \hat{\mathbf{J}}_\varepsilon < \mathbf{J}_\varepsilon - \delta\right)\right]$$
$$= \mathbf{E}_{\underline{\underline{\theta}}}^{\pi^0}\left[\sum_{j=0}^{T_n(a)} \mathbf{1}\left(\mathbf{J}_a\left(\hat{\underline{\theta}}_a^j, \underline{\underline{\theta}}; \varepsilon\right) < \mathbf{J}_\varepsilon - \delta\right)\right]$$
$$\leq \mathbf{E}_{\underline{\underline{\theta}}}^{\pi^0}\left[\sum_{j=0}^{n-1} \mathbf{1}\left(\mathbf{J}_a\left(\hat{\underline{\theta}}_a^j, \underline{\underline{\theta}}; \varepsilon\right) < \mathbf{J}_\varepsilon - \delta\right)\right]$$
$$\leq \mathbf{E}_{\underline{\theta}_a}\left[\sum_{j=0}^{n-1} \mathbf{1}\left(\mathbf{J}_a\left(\hat{\underline{\theta}}_a^j, \underline{\underline{\theta}}; \varepsilon\right) < \mathbf{J}_\varepsilon - \delta\right)\right] = o(\log n),$$

where the second inequality follows from Remark 1 and the last equality follows from Remark 4(b). Therefore, $\mathbf{E}_{\underline{\underline{\theta}}}^{\pi^0} T_n^{(1)}(a, \varepsilon)/\log n \leq \log n/\mathbf{J}_a(\underline{\theta}_a, \underline{\underline{\theta}}; \varepsilon) + o(\log n)$.

Thus the proof of part (a) is complete since $\lim_{\varepsilon \to 0} \mathbf{J}_a(\underline{\theta}_a, \underline{\underline{\theta}}; \varepsilon) = \mathbf{K}_a(\underline{\underline{\theta}})$, if $\mathbf{\Delta}\Theta_a(\underline{\theta}_a, 0) \neq \varnothing$, from the definition of $\mathbf{K}_a(\underline{\underline{\theta}})$, and $\lim_{\varepsilon \to 0} \mathbf{J}_a(\underline{\theta}_a, \underline{\underline{\theta}}; \varepsilon) = \infty$, if $\mathbf{\Delta}\Theta_a(\underline{\theta}_a, 0) = \varnothing$, from (A1).

(b) Note first that for $\pi^0 \in C_R$, the following inequality holds *pointwise* on $\Omega^{(n)}$:

$$T_n^{(2)}(a, \varepsilon) \leq \sum_{k=1}^{n-1} \mathbf{1}\left(\mathbf{u}_{a^*}\left(\hat{\underline{\theta}}_{a^*}^j, \log k/j\right) \leq \mu^*\left(\underline{\underline{\theta}}\right) - \varepsilon, \text{ for some } j \leq k\right),$$

$$\forall a^* \in \mathbf{O}\left(\underline{\underline{\theta}}\right), \forall \underline{\underline{\theta}} \in \Theta.$$

Indeed, $T_n^{(2)}(a, \varepsilon) = \sum_{k=1}^{n-1} \mathbf{1}(\pi_k^0 = a, \mathbf{U}_a(\omega_k) \le \mu^*(\underline{\theta}) - \epsilon)$, and, since $\pi \in C_R$, the condition $\pi_k^0 = a$ implies that $\mathbf{U}_a(\omega_k) = \max_{a'} \mathbf{U}_{a'}(\omega_k) \ge \mathbf{U}_{a^*}(\omega_k)$; thus, the event $\{\pi_k^0 = a, \mathbf{U}_a(\omega_k) \le \mu^*(\underline{\theta}) - \varepsilon\}$ is contained in the event $\{\mathbf{U}_{a^*}(\omega_k) \le \mu^*(\underline{\theta}) - \varepsilon\}$, for any $a^* \in O(\underline{\theta})$. The latter event is contained in the event $\{\mathbf{u}_{a^*}(\hat{\underline{\theta}}_{a^*}^j, k, j) \le \mu^*(\underline{\theta}) - \varepsilon$, for some $j \le k)\}$. Therefore, using also (2.2),

$$\mathbf{E}_{\underline{\theta}}^{\pi^0} T_n^{(2)}(a, \varepsilon) \le \sum_{k=1}^{n-1} \mathbf{P}_{\underline{\theta}_{a^*}}\Big(\mathbf{u}_{a^*}\big(\hat{\underline{\theta}}_{a^*}^j, k, j\big) \le \mu^*\big(\underline{\theta}\big) - \varepsilon, \text{ for some } j \le k\Big) = o(\log n),$$

by Condition (A3).

The proof of (c) follows from (a) and (b) when we let $\varepsilon \to 0$, since $T_n(a) \le 1 + T_n^{(1)}(a, \varepsilon) + T_n^{(2)}(a, \varepsilon)$, $\forall n \ge 1$, $\forall \varepsilon > 0$. ∎

4. APPLICATIONS OF THEOREM 1

4.1. *Discrete Distributions with Finite Support*

Assume that the observations Y_{aj} from population a follow a univariate discrete distribution, i.e., $f_a(y, \underline{p}_a) = p_{ay}\mathbf{1}\{Y_a = y\}$, $y \in S_a = \{r_{a1}, \ldots, r_{ad_a}\}$, where the unknown parameters p_{ay} are in $\Theta_a = \{\underline{p}_a \in \mathbb{R}^{d_a}: p_{ay} > 0, \forall y = 1, \ldots, d_a, \sum_y p_{ay} = 1\}$, and r_{ay} are known. Here we use the notation $\underline{\theta}_a = \underline{p}_a$, $\underline{\theta} = \underline{p} = (\underline{p}_1, \ldots, \underline{p}_m)$ and ν_a is the counting measure on $\{r_{a1}, \ldots, r_{ad_a}\}$.

Thus we can write $\mathbf{I}(\underline{p}_a, \underline{q}_a) = \sum_{y=1}^{d_a} p_{ay} \log(p_{ay}/q_{ay})$, $\mu_a(\underline{p}_a) = \underline{r}_a' \underline{p}_a = \sum_y r_{ay} p_{ay}$, $\mu^* = \mu^*(\underline{p}) = \max_a \{\underline{r}_a' \underline{p}_a\}$, $\mathbf{\Delta}\Theta_a(\underline{p}, \varepsilon) = \{\underline{q}_a: \mu_a(\underline{q}_a) > \mu^*(\underline{p}) - \varepsilon\}$, where $\underline{r}_a'$ denotes the transpose of the vector $\underline{r}_a$. Note that computation of the *constant* $\mathbf{K}_a(\underline{p})$ as a function of $\underline{p}$ involves the minimization of $\underline{p}$ involves the minimizatin of a convex function subject to two linear constraints; hence,

$$\mathbf{K}_a\big(\underline{p}\big) = \mathbf{w}_a\Big(\underline{p}_a, \mu^*\big(\underline{p}\big)\Big) = \min_{\underline{q}_a \ge \underline{0}} \left\{ \mathbf{I}(\underline{p}_a, \underline{q}_a): \underline{r}_a' \underline{q}_a \ge \mu^*\big(\underline{p}\big), \sum_{y=1}^{d_a} q_{ay} = 1 \right\}. \tag{4.1}$$

For any estimators $\hat{\underline{p}}_a^t$ of $\underline{p}_a$, the computation of the *index* $\mathbf{U}_a(\omega_k)$ involves the solution of the dual problem of (4.1) (with $\underline{p}_a$ replaced by $\hat{\underline{p}}_a^t$) in (4.1)), which, in this case, is a problem of maximization of a linear function

subject to a constraint with convex level sets and a linear constraint; hence,

$$\mathbf{U}_a(\omega_k) = \mathbf{u}_a\left(\hat{\mathrm{p}}_a^t, \log k/t\right) = \max_{\underline{\mathrm{q}}_a \geq \underline{0}} \left\{ \underline{\mathrm{r}}_a' \underline{\mathrm{q}}_a : \mathbf{I}\left(\hat{\mathrm{p}}_a^t, \underline{\mathrm{q}}_a\right) \leq \log k/t, \sum_{y=1}^{d_a} \mathrm{q}_{ay} = 1 \right\}. \quad (4.2)$$

In Proposition 3 below, it is shown that Conditions (A1), (A2), and (A3) are satisfied for estimators defined from the observations from population a. Given $\omega\kappa$ with $T_k(a) = t$ define:

(1) For $t \geq 1$, $n_t(y; a) = \sum_{j=1}^t \mathbf{1}(Y_{aj} = r_{a,y})$, $\mathrm{f}_t(y; a) = n_t(y; a)/t$ and $\underline{\mathrm{f}}_t(a) = [\mathrm{f}_t(y; a)]_{y \in S_a}$.

(2) For $t \geq 0$ let $\hat{\mathrm{p}}_{a,y}^t = (1 - w_t)/d_a + w_t \mathrm{f}_t(y; a)$, where $w_t = t/(d_a + t)$, and let $\hat{\mathrm{p}}_n^t = [\hat{\mathrm{p}}_{a,y}^t]_{y \in S_a}$.

In the proof of Proposition 3 we make use of the following quantities and properties:

(1) For $a = 1, \ldots, m$ and $\underline{\mathrm{p}}_a, \underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2 \in \Theta_a$ let $\lambda(\underline{\mathrm{p}}_a; \underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) = \mathbf{I}(\underline{\mathrm{p}}_a, \underline{\mathrm{q}}_2) - \mathbf{I}(\underline{\mathrm{p}}_a, \underline{\mathrm{q}}_1) := \sum_{y \in S_a} \mathrm{p}_{ay} \log[\mathrm{q}_{1y}/\mathrm{q}_{2y}]$.

(2) Let $\Lambda_t(\underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) = \prod_{j=1}^t \mathrm{q}_{1,Y_{aj}}/\mathrm{q}_{2,Y_{aj}}$.

(3) For $\underline{\mathrm{p}}_a \in \Theta_a$, let $F_{kt}(\underline{\mathrm{p}}_a) = \{\underline{\mathrm{q}} \in \Theta_a : \mathbf{I}(\underline{\mathrm{p}}_a, \underline{\mathrm{q}}) \leq \log k/t\}$.

Note that $\mathbf{U}_a(\omega_k) = \sup\{\underline{\mathrm{r}}_a' \underline{\mathrm{q}} : \underline{\mathrm{q}} \in F_{k,T_k(a)}(\hat{\mathrm{p}}_a^{T_k(a)})\}$.

Remark 9. (a) $\log \Lambda_t(\underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) = \lambda(\underline{\mathrm{f}}_t(a); \underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2)$.

(b) $\sup_{\underline{\mathrm{q}}_1} \lambda(\underline{\mathrm{p}}_a; \underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) = \lambda(\underline{\mathrm{p}}_a; \underline{\mathrm{p}}_a, \underline{\mathrm{q}}_2) = \mathbf{I}(\underline{\mathrm{p}}_a, \underline{\mathrm{q}}_2)$.

(c) $e^{t\mathbf{I}(\underline{\mathrm{f}}_t(a), \underline{\mathrm{q}}_2)} = e^{t\lambda(\underline{\mathrm{f}}_t(a); \underline{\mathrm{f}}_t(a), \underline{\mathrm{q}}_2)} = \sup_{\underline{\mathrm{q}}_1} \Lambda_t(\underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2)$.

(d) $t \cdot \lambda(\hat{\mathrm{p}}_a^t; \underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) = w_t[b(\underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) + \lambda(\underline{\mathrm{f}}_t(a); \underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2)]$, where $b(\underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) = \sum_y \log[\mathrm{q}_{1y}/\mathrm{q}_{2y}]$.

(e) $t\mathbf{I}(\hat{\mathrm{p}}_a^t, \underline{\mathrm{q}}_2) \leq w_t[b_0(\underline{\mathrm{q}}_2) = t\mathbf{I}(\underline{\mathrm{f}}_t(a), \underline{\mathrm{q}}_2)]$, where $b_0(\underline{\mathrm{q}}_2 0 = -\sum_y \log \mathrm{q}_{2y}$.

Indeed, (a) follows from the observation that

$$\log \Lambda_t(\underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2) = \sum_{j=1}^t \log\left[\mathrm{q}_{1,Y_{aj}}/\mathrm{q}_{2,Y_{aj}}\right] = \sum_{j=1}^t \sum_{y=1}^{d_a} 1(Y_{aj} = y) \log\left[\mathrm{q}_{1y}/\mathrm{q}_{2y}\right] = t\lambda\left(\underline{\mathrm{f}}_t(a); \underline{\mathrm{q}}_1, \underline{\mathrm{q}}_2\right).$$

(b) is a restatement of the information inequality $-\mathbf{I}(\underline{p}_a, \underline{q}_1) \le 0$. (c) follows from (b). To see (d) and (e), recall that $\hat{\underline{p}}_a^t = (1 - w_t)/d_a + w_t \cdot \underline{f}_t(a)$, where $w_t = t/(t + d_a)$; note that $t(1 - w_t)/d_a = w_t$, and use (a) (for (d)) and (b) (for (e)).

PROPOSITION 3. *The discrete distribution model satisfies Conditions* (A1), (A2), *and* (A3) *of Theorem* 1.

Proof. (1) It is easy to see that Condition (A1) holds. Indeed, note that that $\forall \varepsilon \ge 0$, $\mathbf{\Delta}\Theta_a(\underline{\underline{p}}; \varepsilon) \ne \varnothing$ if and only if $\max_y r_{ay} > \mu^*(\underline{\underline{p}}) - \varepsilon$. Thus if $\mathbf{\Delta}\Theta_a(\underline{\underline{p}}, \varepsilon) \ne \varnothing$, $\forall \varepsilon > 0$ and $\mathbf{\Delta}\Theta_a(\underline{\underline{p}}) = \varnothing$, then $\max_y r_{ay} = \mu^*(\underline{\underline{p}})$ and $\lim_{\varepsilon \to 0} \mathbf{J}(\underline{p}_a, \underline{\underline{p}}; \varepsilon) = \mathbf{I}(\underline{p}_a, \underline{q}_e) = \infty$, where $\underline{q}_e$ is the unit vector of $\mathbb{R}^{d_a}$, with nonzero component corresponding to $\max_y r_{ay}$.

(2) We next show that Condition (A2) holds. Since $1 - w_t \to 0$, it follows from the definition of $\hat{p}_{a,y}^t$ that for any $\varepsilon > 0$ there exists $t_0(\varepsilon) \ge 1$ such that

$$\mathbf{P}_{\underline{p}_a}\left[|\hat{p}_{a,y}^t - p_{ay}| > \varepsilon\right] \le \mathbf{P}_{\underline{p}_a}\left[|f_t(y; a) - p_{ay}| > \frac{\varepsilon}{2}\right], \qquad \forall t \ge t_0.$$

Because $f_t(y; a)$ is the average of i.i.d. Bernoulli random variables with mean p_{ay}, it follows from standard results of large deviations theory (cf. [11; 9, p. 27]) that $\mathbf{P}_{\underline{p}_a}[|f_t(y; a) - p_{ay}| > \varepsilon/2] \le Ce^{-\gamma t}$ for some $C, \gamma > 0$; therefore, for $t \ge t_0(\varepsilon)$, $\mathbf{P}_{\underline{p}_a}[|\hat{p}_{a,y}^t - p_{ay}| > \varepsilon] \le Ce^{-\gamma t} = o(1/t)$.

(3) To show that the model satisfies Condition (A3), we must prove that $\mathbf{P}_{\underline{p}_a}[\cup_{t=0}^k B_{kt}] = o(1/k)$, as $k \to \infty$, $\forall \varepsilon > 0$, $\forall a$, where $B_{kt} = B_{kt}(a, \varepsilon) = \{\mathbf{u}_a(\hat{\underline{p}}_a^t, \log k/t) \le \mu_a(\underline{p}_a) - \varepsilon\}$.

On the event B_{kt} it is true that $\mu_a(\underline{q}) < \mu_a(\underline{p}_a) - \varepsilon$, $\forall \underline{q} \in F_{kt}(\hat{\underline{p}}_a^t)$. Therefore, $B_{kt} \subseteq B'_{kt}$, where $B'_{kt} = \{\mu_a(\underline{q}) < \mu_a(\underline{p}_a) - \varepsilon, \ \forall q \in F_{kt}(\hat{\underline{p}}_a^t)\}$; thus it suffices to prove that $\sum_{t=0}^{k-1} \mathbf{P}_{\underline{p}_a}[B'_{kt}] = o(1/k)$.

From Lemma 1 it follows that for $\varepsilon > 0$ sufficiently small, there exists a probability vector $\underline{q}^0 = \underline{q}^0(\varepsilon) \in \Theta_a$ such that $\mu(\underline{q}^0) = \mu_a(\underline{p}_a) - \varepsilon$ and $\{\underline{q}: \mu(\underline{q}) < \mu_a(\underline{p}_a) - \varepsilon\} = \{\underline{q}: \lambda(\underline{q}; \underline{p}_a, \underline{q}^0) < -c_\varepsilon\}$, where $c_\varepsilon = \mathbf{I}(\underline{q}^0, \underline{p}_a)$.

Hence, $B'_{kt} \subseteq B''_{kt}$, where $B''_{kt} = \{\lambda(\underline{q}; \underline{p}_a, \underline{q}^0) < -c_\varepsilon, \forall \underline{q} \in F_{kt}(\hat{\underline{p}}_a^t)\}$ and it is sufficient to prove that $\sum_{t=0}^{k-1} \mathbf{P}_{\underline{p}_a}[B''_{kt}] = o(1/k)$.

On the event B''_{kt} the following are true. First, since $\hat{\underline{p}}_a^t \in F_{kt}(\hat{\underline{p}}_a^t)$, it follows that $\lambda(\hat{\underline{p}}_a^t; \underline{p}_a, \underline{q}^0) < -c_\varepsilon$. Second, since $\lambda(\underline{q}^0; \underline{p}_a, \underline{q}^0) = -c_\varepsilon$, it follows that $\underline{q}^0 \notin F_{kt}(\hat{\underline{p}}_a^t)$, i.e,. $\mathbf{I}(\hat{\underline{p}}_a^t, \underline{q}^0) > \log k/t$. Therefore, $B''_{kt} \subseteq B'''_{kt}$, where $B'''_{kt} = \{\lambda(\hat{\underline{p}}_a^t; \underline{p}_a, \underline{q}^0) < -c_\varepsilon, \ \mathbf{I}(\hat{\underline{p}}_a^t, \underline{q}^0) > \log k/t\}$. Let $\bar{I}(\underline{q}^0) = \max_{q \in \Theta_a} \mathbf{I}(\underline{q}, \underline{q}^0) = \max_{y \in S_a} \log|q_y^0| < \infty$.

For $t < \log k/\bar{I}(\underline{q}^0)$, it is true that $\mathbf{I}(\hat{\underline{p}}_a^t, \underline{q}^0) \le \bar{I}(\underline{q}^0) < \log k/t$; thus $B'''_{kt} = \varnothing$. Therefore, in order to prove the proposition, it suffices to show that $\sum_{t=\lfloor \log k/\bar{I} \rfloor}^{k-1} \mathbf{P}_{\underline{p}_a}[B'''_{kt}] = o(1/k)$. This follows from Lemma 2 (2); thus the proof is complete. ∎

LEMMA 1. *For any* $\underline{p}_a \in \Theta_a$ *and* $\varepsilon > 0$ *sufficiently small, there exists a vector* $\underline{q}^0(\varepsilon) \in \Theta_a$ *such that* $\mu_a(\underline{q}^0(\varepsilon)) = \mu_a(\underline{p}_a) - \varepsilon$ *and* $\{\underline{q} \in \Theta_a\colon \mu_a(\underline{q}) < \mu_a(\underline{p}_a) - \varepsilon\} = \{\underline{q} \in \Theta_a\colon \lambda(\underline{q}; \underline{p}, \underline{q}^0(\varepsilon)) < -\mathbf{I}(\underline{q}^0(\varepsilon), \underline{p})\}$.

Proof. For $\nu \geq 0$ define $\tilde{\underline{q}}(\nu) \in S_a$ as follows: $\tilde{q}_y(\nu) = p_{ay} e^{-\nu r_{ay}}/b(\nu)$, where $b(\nu) = \sum_{y \in S_a} p_{ay} e^{-\nu r_{ay}}$. We prove that for all $\varepsilon > 0$ there exists $\nu = \nu(\varepsilon) > 0$ such that $H_\mu(\varepsilon) = H_\lambda(\nu(\varepsilon))$, where $H_\mu(\varepsilon) = \{\underline{q}\colon \mu_a(\underline{q}) = \mu_a(\underline{p}_a) - \varepsilon\}$, $H_\lambda(\nu) = \{\underline{q}\colon \lambda(\underline{q}; \underline{p}_a, \tilde{\underline{q}}(\nu)) = -c_\nu\}$ and $c_\nu = -\lambda(\tilde{\underline{q}}(\nu); \underline{p}_a, \tilde{\underline{q}}(\nu)) = \mathbf{I}(\tilde{\underline{q}}(\nu), \underline{p}_a)$. Indeed, for all $\varepsilon, \nu > 0$, $H_\mu(\eta)$ and $H_\lambda(\nu)$ are parallel hyperplanes, since by construction of $\tilde{\underline{q}}(\nu)$,

$$\lambda\big(\underline{q}; \underline{p}_a, \tilde{\underline{q}}(\nu)\big) = \sum_{y \in S_a} q_y \log \frac{p_{ay}}{\tilde{q}_y(\nu)}$$

$$= \nu \sum_{y \in S_a} q_y r_{ay} + \log b(\nu) = \nu \mu_a(\underline{q}) + \log b(\nu).$$

In addition, $\tilde{\underline{q}}(\nu) \in H_\lambda(\nu)$, since $\lambda(\tilde{\underline{q}}(\nu); \underline{p}_a, \tilde{\underline{q}}(\nu)) = -\mathbf{I}(\tilde{\underline{q}}(\nu), \underline{p}_a) = -c_\nu$.

Hence, for $H_\mu(\varepsilon) = H_\lambda(\nu(\varepsilon))$, it suffices to choose $\nu(\varepsilon)$ such that $\tilde{\underline{q}}(\nu(\varepsilon)) \in H_\mu(\varepsilon)$, i.e., $\mu_a(\tilde{\underline{q}}(\nu(\varepsilon))) = \mu_a(\underline{p}_a) - \varepsilon$.

For $\nu = 0$ it is true that $b(0) = 1$ and $\tilde{\underline{q}}(0) = \underline{p}_a$; thus, $\mu(\tilde{\underline{q}}(0)) = \mu_a(\underline{p}_a)$. As $\nu \to \infty$,

$$\tilde{\underline{q}}(\nu) \to \tilde{\underline{q}}(\infty) := \begin{cases} p_{ay}/\sum_{z:\, r_{az} = r_a^0} p_{az}, & \text{if } r_{ay} = r_a^0 \\ 0, & \text{otherwise} \end{cases},$$

where $r_a^0 = \min_z r_{az}$. Thus, as $\nu \to \infty$, $\mu_a(\tilde{\underline{q}}(\nu)) \to \mu_\infty := r_a^0 < \mu_a(\underline{p}_a)$.

Therefore, for any $\varepsilon < (\mu_a(\underline{p}_a) - \mu_\infty)$, there exists $\nu(\varepsilon) > 0$ such that $\mu_a(\tilde{\underline{q}}(\nu(\varepsilon))) = \mu_a(\underline{p}_a) - \varepsilon$, because $\mu_a(\underline{q}_\nu^0)$ is continuous in ν.

Let $H_\mu^-(\varepsilon), H_\mu^+(\varepsilon), H_\lambda^-(\nu), H_\lambda^+(\nu)$ denote the corresponding half-spaces of H_μ, H_λ, i.e,. $H_\mu^-(\varepsilon) = \{\underline{q}\colon \mu_a(\underline{q}) < \mu_a(\underline{p}_a) - \varepsilon\}$, etc. To prove that $H_\mu^-(\varepsilon) = H_\lambda^-(\nu(\varepsilon))$, it suffices to show that $\underline{p}_a \in H_\mu^+(\varepsilon)$ and $\underline{p}_a \in H_\lambda^+(\nu(\varepsilon))$. The first is immediate, while for the second we note that $\lambda(\underline{p}_a; \underline{p}_a, \tilde{\underline{q}}(\nu(\varepsilon))) = \mathbf{I}(\underline{p}_a, \tilde{\underline{q}}(\nu(\varepsilon))) > 0 > -\mathbf{I}(\tilde{\underline{q}}(\nu(\varepsilon)), \underline{p}_a)$. Thus the lemma follows with $\underline{q}^0(\varepsilon) = \tilde{\underline{q}}(\nu(\varepsilon))$. ∎

LEMMA 2. (1) $\forall \underline{p}_a, \underline{q} \in \Theta_a$, $\forall c, d > 0$, *and* $\forall b_1, b_2 \in \mathbb{R}$

$$\sum_{t = \lfloor d \log k \rfloor}^{k-1} \mathbf{P}_{\underline{p}_a}\Big[\lambda\big(\underline{f}_t(a); \underline{p}_a, \underline{q}\big) < -c + b_1/t, \mathbf{I}\big(\underline{f}_t(a), \underline{q}\big) > \log k/t + b_2/t\Big]$$

$$= o(1/k), \qquad \text{as } k \to \infty.$$

(2) $\forall c, d > 0$ *and* $\underline{p}_a, \underline{q} \in \Theta_a$

$$\sum_{t = \lfloor d \log k \rfloor}^{k-1} \mathbf{P}_{\underline{p}_a}\Big[\lambda\big(\underline{\hat{p}}_a^t; \underline{p}_a, \underline{q}\big) < -c, \mathbf{I}\big(\underline{\hat{p}}_a^t, \underline{q}\big) > \log k/t\Big] = o(1/k),$$

as $k \to \infty$.

Proof. (1) The proof is an adaptation of Lemma 2 in [25]. From Remark 9(b) it follows that, for all k, t,

$$\mathbf{P}_{\underline{p}_a}\Big[\lambda\big(\underline{f}_t(a); \underline{p}_a, \underline{q}\big) < -c + b_1/t, \mathbf{I}\big(\underline{f}_t(a), \underline{q}\big) > \log k/t + b_2/t\Big]$$
$$= \mathbf{P}_{\underline{p}_a}\left[\Lambda_t(\underline{p}_a, \underline{q}) \le e^{b_1}e^{-ct}, \sup_{\underline{p}} \Lambda_t(\underline{p}, \underline{q}) > e^{b_2}k\right].$$

After a change-or-measure transformation between $\underline{p}_a$ and $\underline{q}$ we obtain

$$\mathbf{P}_{\underline{p}_a}\left[\Lambda_t(\underline{p}_a, \underline{q}) \le e^{b_1}e^{-ct}, \sup_{\underline{p}} \Lambda_t(\underline{p}, q) > e^{b_2}k\right]$$
$$\le e^{b_1}e^{-ct}\mathbf{P}_{\underline{q}}\left[\sup_{\underline{p}} \Lambda_t(\underline{p}, \underline{q}) > e^{b_2}k\right]. \tag{4.3}$$

Since Θ_a is subset of a compact set, for any $\delta > 0$ there exists $M < \infty$ and a finite collection of vectors $\underline{q}^{(i)} \in \Theta_a$, and neighborhoods $\mathcal{N}_i(\delta)$, $i = 1, \ldots, M$, such that $\bigcup_i \mathcal{N}_i(\delta) \supseteq \overline{\Theta}_a$, and $\mathcal{N}_i(\delta) = \{\underline{p} \in \Theta_a : \|\underline{p} - \underline{q}^{(i)}\| < \delta\}$.

For all $i = 1, \ldots, M$ and $y \in S_a$, it is true that $\sup_{\underline{p} \in \mathcal{N}_i(\delta)} p_y/q_y \le (q_y^{(i)} + \delta)/q_y$; thus,

$$\mathbf{E}_{\underline{q}}\left[\sup_{\underline{p} \in \mathcal{N}_i(\delta)} \frac{p_{Y_a}}{q_{Y_a}}\right] \le \mathbf{E}_{\underline{q}}\left[\frac{q_{Y_a}^{(i)} + \delta}{q_{Y_a}}\right] = 1 + |S_a|\delta.$$

Therefore for any $\varepsilon > 0$, selecting $\delta < \varepsilon/|S_a|$, we obtain $\mathbf{E}_{\underline{q}}[\sup_{\underline{p} \in \mathcal{N}_i(\delta)} p_{Y_a}/q_{Y_a}] \le 1 + \varepsilon$, $i = 1, \ldots, M$, and thus

$$\mathbf{P}_{\underline{q}}\left[\sup_{\underline{p} \in \mathcal{N}_i(\delta)} \Lambda_t(\underline{p}, \underline{q}) > e^{b_2}k\right]$$
$$\le e^{-b_2}k^{-1}\mathbf{E}_{\underline{q}}\left[\sup_{\underline{p} \in \mathcal{N}_i(\delta)} \Lambda_t(\underline{p}, \underline{q})\right]$$
$$\le e^{-b_2}k^{-1}\left(\mathbf{E}_{\underline{q}}\left[\sup_{\underline{p} \in \mathcal{N}_i(\delta)} \frac{p_{Y_a}}{q_{Y_a}}\right]\right)^t \le e^{-b_2}k^{-1}(1 + \varepsilon)^t, \tag{4.4}$$

where the first inequality follows from the Markov inequality, the second from the observation that $\sup_{\underline{p} \in \mathcal{N}_i(\delta)} \Lambda_t(\underline{p}, \underline{q}) \le \Pi_{j+1}^t \sup_{\underline{p} \in \mathcal{N}_i(\delta)} p_{Y_{aj}}/q_{Y_{aj}}$, and the third from the fact that Y_{aj}, $j = 1, \ldots, t$, are i.i.d.

Combining (4.3) and (4.4)

$$\mathbf{P}_{\underline{p}_a}\left[\Lambda_t(\underline{p}_a, \underline{q}) \le e^{b_1}e^{-8cu^t}, \sup_{\underline{p}} \Lambda_t(\underline{p}, \underline{q}) > e^{b_2}k\right]$$

$$\le e^{b_1}e^{-ct}\sum_{i=1}^{M}\mathbf{P}_{\underline{q}}\left[\sup_{\underline{p}\in\mathcal{N}_i(\delta)} \Lambda_t(\underline{p}, \underline{q}) > e^{b_2}k\right]$$

$$\le Me^{b_1-b_2}k^{-1}e^{-ct}(1+\varepsilon)^t.$$

Selecting ε so that $e^{-c}(1+\varepsilon) < 1$, we obtain

$$\sum_{t=\lfloor d\log k\rfloor}^{k}\mathbf{P}_{\underline{p}_a}\left[\Lambda_t(\underline{p}_a, \underline{q}) \le b_1e^{-ct}, \sup_{\underline{p}\in\Theta_a} \Lambda_t(\underline{p}, \underline{q}) > b_2k\right]$$

$$\le M'k^{-1}\sum_{t=\lfloor d\log k\rfloor}^{\infty}\left(e^{-c}(1+\varepsilon)\right)^t \le M'k^{-1-d(c-\log(1+\varepsilon))},$$

where $M' = Me^{b_1-b_2}/(1-e^{-c}(1+\varepsilon))$.

Since $d > 0$ and $c - \log(1+\varepsilon) > 0$ by selection of ε, it follows that $-1 - d(c - \log(1+\varepsilon)) < -1$, and the proof is complete.

(2) From Remark 9(c), in the event $\{\lambda(\hat{\underline{p}}_a^t; \underline{p}_a, \underline{q}) < -c\}$ it is true that $w_t(b(\underline{p}_a, \underline{q}) + t\lambda(\underline{f}_t(a); \underline{p}_a, \underline{q})) < -ct$; therefore, since $0 < w_t < 1$, $b(\underline{p}_a, \underline{q}) + t\lambda(\underline{f}_t(a); \underline{p}_a, \underline{q}) < -ct/w_t < -ct$.

Also, in the event $\{\mathbf{I}(\hat{\underline{p}}_a^t, \underline{q}) > \log k/t\}$ it is true that $w_t(b_0(\underline{q}) + t\mathbf{I}(\underline{f}_t(a), \underline{q})) > \log k$; therefore, since $0 < w_t < 1$, $b_0(\underline{q}) + t\mathbf{I}(\underline{f}_t(a), \underline{q}) > \log k/w_t > \log k$.

Hence,

$$\sum_{t=\lfloor d\log k\rfloor}^{k}\mathbf{P}_{\underline{p}_a}\left[\lambda\left(\hat{\underline{p}}_a^t; \underline{p}_a, \underline{q}\right) < -c, \mathbf{I}\left(\hat{\underline{p}}_a^t, \underline{q}\right) > \log k/t\right]$$

$$\le \sum_{t=\lfloor d\log k\rfloor}^{k}\mathbf{P}_{\underline{p}_a}\left[\lambda\left(\underline{f}_t(a); \underline{p}_a, \underline{q}\right) < -c - b(\underline{p}_a, \underline{q})/t, \mathbf{I}\left(\underline{f}_t(a), \underline{q}\right)\right.$$

$$\left. > \log k/t - b_0(\underline{q})/t\right],$$

and the result follows from part (1), with $b_1 = -b(\underline{p}_a, \underline{q}^0)$ and $b_2 = -b_0(\underline{q}^0)$. ∎

4.2. *Normal Distributions*

Assume that the observations Y_{aj} from population a are normally distributed with unknown mean μ_a and known variance σ_a^2, i.e., $\underline{\theta}_a = \mu_a$,

and that $\Theta_a = (-\infty, \infty)$. Given history ω_k, define $\hat{\mu}_a^{T_k(a)} = (\sum_{j=1}^{T_k(a)} Y_{aj})/T_k(a)$.

From the definition of Θ_a it follows that $\mathbf{\Delta}\Theta_a(\underline{\theta}) = (\mu^*(\underline{\theta}), \infty)$; therefore, $\mathbf{B}(\underline{\theta}) = \{1, \ldots, m\}$, $\forall \underline{\theta} \in \Theta$. Also it can be seen after some algebra that $\mathbf{K}_q(\underline{\theta}) = (1/2)\log(1 + (\mu^*(\underline{\theta} - \mu_a)^2/\sigma_a^2)$ and $\mathbf{U}_a(\omega_k) = \hat{\mu}_a^{T_k(a)} + \hat{\sigma}_a^{T_k(a)}(e^{2 \log k / T_k(a)} - 1)^{1/2} = \hat{\mu}_a^{T_k(a)} + \hat{\sigma}_a^{T_k(a)}(k^2/T_k(a) - 1)^{1/2}$.

Therefore, Condition (A1) of Theorem 1 holds. Condition (A2) follows from standard large deviation arguments. Condition (A3) follows from an inequality for the tails of the normal distribution (cf. [13, p. 166]). Therefore, any index policy in C_R is UMCR.

For details and variations of this model see [25, 23, 21].

4.3. *Normal and Discrete Distributions*

Assume that $\exists m_1 < m$ such that:

(1) For $a = 1, \ldots, m_1$, Y_{aj} are normally distributed with unknown mean μ_a and known variance σ_a^2, i.e., $\underline{\theta}_a = \mu_a$, and $\Theta_a = (-\infty, \infty)$.

(2) For $a = m_1 + 1, \ldots, m$, Y_{aj} follow discrete distribution with known support $S_a = \{r_{a1}, \ldots, r_{ad_a}\}$ and unknown parameters $\underline{p}_a \in \Theta_a = \{\underline{p}_a \in \mathbb{R}^{d_a}: p_{ay} > 0, \forall y = 1, \ldots, d_a, \sum_y p_{ay} = 1\}$.

In this case $B(\underline{\theta}) = \{1, \ldots, m_1\} \cup \{a > m_1: \max_y r_{ay} > \mu^*(\underline{\theta})\}$.

Conditions (A1), (A2), and (A3) are satisfied. Indeed, they have been verified separately for $a > m_1$ in subsection 4.1 and for $a \leq m_1$ in subsection 4.2. Thus any index policy in C_R is UM.

4.4. *Normal Distributions with Unknown Variance*

Assume that Y_{aj} are normally distributed with unknown mean μ_a and variance σ_a^2, i.e., $\underline{\theta}_a = (\mu_a, \sigma_a^2)$, and that $\Theta_a = \{\underline{\theta}_a: \mu_a \in \mathbb{R}, \sigma_a^2 \geq 0\}$.

Given history ω_k, define $\hat{\underline{\theta}}_a^{T_k(a)} = (\hat{\mu}_a^{T_k(a)}, \hat{\sigma}_a^{2\,T_k(a)})$, where $\hat{\mu}_a^{T_k(a)} = 1/T_k(a)\sum_{j=1}^{T_k(a)} Y_{aj}$, $\hat{\sigma}_a^{2\,T_k(a)} = s^2(T_k(a)) = 1/T_k(a)\sum_{j=1}^{T_k(a)}(Y_{aj} - \hat{\mu}_a^{T_k(a)})^2$.

From the definition of Θ_a it follows that $\mathbf{\Delta}\Theta_a(\underline{\theta}) \neq \emptyset$, $\forall \underline{\theta} \in \Theta$; therefore, $\mathbf{B}(\underline{\theta}) = \{1, \ldots, m\}$, $\forall \underline{\theta} \in \Theta$. Also it can be seen after some algebra that $\mathbf{K}_a(\underline{\theta}) = (1/2)\log(1 + (\mu^*(\underline{\theta}) - \mu_a)^2/\sigma_a^2)$ and $\mathbf{U}_a(\omega_k) = \hat{\mu}_a^{T_k(a)} + \hat{\sigma}_a^{T_k(a)}(e^{2 \log k / T_k(a)} - 1)^{1/2} = \hat{\mu}_a^{T_k(a)} + \hat{\sigma}_a^{T_k(a)}(k^{2/T_k(a)} - 1)^{1/2}$.

Therefore, Condition (A1) of Theorem 1 holds. It is easy to to see that (A2) holds, using large deviations arguments. However, we have not been able to prove that (A3) is satisfied, so this remains an open problem.

ACKNOWLEDGMENTS

We are grateful to H. Robbins for helpful comments on this paper.

REFERENCES

1. R. Acost-Abreu and O. Hernandez-Lerma, Iterative adaptive control of denumerable state, average-cost Markov systems, *Control Cybernet*. **14** (1985), 313–322.
2. R. Agrawal, D. Teneketzis, and V. Anantharam, Asymtotically efficient adaptive allocation schemes for controlled Markov chains: Finite parameter space, *Trans*. *Automat*. *Control IEEE* **34** (1989), 1249–1259.
3. R. Bellman, ''Dynamic Programming,'' Princeton Univ. Press, Princeton, NJ, 1957.
4. D. A. Berry and B. Fristedt, ''Bandit Problems: Sequential Allocation of Experiments,'' Chapman & Hall, London (1985).
5. F. J. Beutler and D. Teneketzis, Routing in queueing networks under imperfect information: Stochastic dominance and thresholds, *Stochastics Stochastics Rep*. **26** (1989), 81–100.
6. A. N. Burnetas and M. N. Katehakis, ''On Optimal Sequential Allocation Policies for the Finite Horizon One-Armed Bandit Problem,'' Technical Report, Rutgers University, 1989.
7. A. N. Burnetas and M. N. Katehakis, On sequencing two types of tasks on a single processor under incomplete information, *Probab. Engrg*. *Inform*. *Sci*. **7** (1993) 85–119.
8. A. N. Burnetas and M. N. Katehakis, ''Optimal Adaptive Policies for Dynamic Programming,'' Technical Report, Rutgers University, 1994.
9. A. Dembo and O. Zeitouni, ''Large Deviations Techniques and Applications,'' Jones & Bartlett, Boston, 1993.
10. E. B. Dynkin and A. A. Yushkevich, ''Controlled Markov Processes,'' Springer-Verlag, Berlin/New York, 1979.
11. R. S. Ellis, ''Entropy, Large Deviations and Statistical Mechanics,'' Springer-Verlag, Berlin/New York, 1985.
12. A. Federgruen and P. Schweitzer, Nonstatoinary Markov decision problems with converging parameters, *J*. *Optim*. *Theory Appl*. **34** (1981), 207–241.
13. W. Feller, ''An Introduction to Probability Theory and Its Applications,'' Vol. 1, 3rd ed., Wiley, New York, 1967.
14. B. L. Fox and J. F. Rolph, Adaptive policies for Markov renewal programs, *Ann*. *Statist*. **1** (1973), 334–341.
15. J. C. Gittins, Bandit processes and dynamic allocation indices (with discussion), *J*. *Roy*. *Statist*. *Soc*. *Ser*. *B* **41** (1979), 335–340.
16. J. C. Gittins and K. D. Glazebrook, On Bayesian models in stochastic scheduling, *J*. *Appl*. *Probab*. **14** (1977), 556–565.
17. J. C. Gittins and D. M. Jones, A dynamic allocation index for the discounted multarmed bandit problem, *Biometrika* **66** (1979), 561–565.
18. Z. Govindarajulu, ''The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory,'' 2nd ed., Am. Sci. Press, Syracuse, NY, 1987.
19. M. N. Katehakis and C. Derman, Computing optimal sequential allocation rules in clinical trials, *in* ''Adaptive Statistical Procedures and Related Topics,'' (J. van Ryzin, Ed. Vol. 8, pp. 29–39, IMS Lecture Notes—Monograph Series, Inst. Math .Statist., Hayward, CA, 1985.
20. M. N. Katehakis and A. F. Vienott, Jr., The multi-armed bandit problem: Decomposition and computation, *Math*. *Oper*. *Res*. **12** (1987), 262–268.
21. M.N. Katehakis and H. Robbins, ''Sequential Allocation Involving Normal Populations,'' Technical Report, Rutgers University, 1994.
22. P. R. Kumar, A survey of some results in stochastic adaptive control, *SIAM J*. *Control Optim*. **23** (1985) 329–380.
23. T. L. Lai, Adaptive treatment allocation and the multi-armed bandit problem, *Ann*. *Statist*. **15** (1987), 1091–1114.

24. T. L. Lai and H. Robbins, Asymptotically optimal allocation of treatments in sequential experiments, *in* "Design of Experients: Ranking and Selection: Essays in Honor of Robert E. Bechhofer" (T. J. Santner and A. C. Tamhane, Eds.), Vol. 56, pp. 127–142, Dekker, New York, 1984.
25. T. L. Lai and H. Robbins, Asymptotically efficient adaptive allocation rules, *Adv. in Appl. Math.* **6** (1985), 4–22.
26. Z. Li and C. Zhang, Asymptotically efficient allocation rules for two Bernoulli populations, *J. Roy. Statist. Soc. Ser. B* **54** (1992), 609–616.
27. P. Mandl, Estimation and control in Markov chains, *Adv. in Appl. Probab.* **6** (1974), 40–60.
28. K. M. Van Hee, Markov decision processes with unknown transition law: The average return case, *in* "Recent Developments in Markov Decision Processes" (D. J. White R. Hartley, and L. C. Thomas, Eds.), pp. 227–244. Academic Press, New York, 1980.
29. R. A. Milito and J. B. Cruz, Jr., A weak contrast function approach to adaptive semi-Markov decision models, *in* "Stochastic Large Scale Engineering Systems" (S. Tzafestas and C. Watanabe, Eds.), pp. 253–278, Dekker, New York, 1992.
30. U. Rieder and J. Weishaupt, Customer scheduling with incomplete information, *Probab. Engrg. Inform. Sci.* to appear.
31. H. Robbins, Some aspects of the sequential design of experiments, *Bull. Amer. Math. Monthly* **58** (1952), 527–536.
32. H. R. Varian, "Microeconomic Analysis," 2nd ed., Norton, New York, 1984.
33. S. Yakowitz and W. Lowe, Nonparametric bandit methods, *Ann. Oper. Res.* **28** (1991), 297–312.