The Annals of Statistics
2000, Vol. 28, No. 6, 1670–1695

ASYMPTOTICALLY EFFICIENT STRATEGIES FOR A STOCHASTIC SCHEDULING PROBLEM WITH ORDER CONSTRAINTS

BY CHENG-DER FUH[1] AND INCHI HU[2]

Academia Sinica and Hong Kong University of Science and Technology

Motivated by an application in computerized adaptive tests, we consider the following sequential design problem. There are J jobs to be processed according to a predetermined order. A single machine is available to process these J jobs. Each job under processing evolves stochastically as a Markov chain and earns rewards as it is processed, not otherwise. The Markov chain has transition probabilities parameterized by an unknown parameter θ. The objective is to determine how long each job should be processed so that the total expected rewards over an extended time interval is maximized. We define the regret associated with a strategy as the shortfall from the maximum expected reward under complete information on θ. Therefore the problem is equivalent to minimizing the regret. The asymptotic lower bound for the regret associated with any uniformly good strategy is characterized by a deterministic constraint minimization problem. In ignorance of the parameter value, we construct a class of efficient strategies, which achieve the lower bound, based on the theory of sequential testing.

1. Introduction. Consider the following sequential design problem. There are J jobs to be processed by a single machine. When job j is processed for a unit time, one observes the current state x of the job and receives a reward $g_j(x)$, which depends on the job nature j and the job state x. The probability law that governs the evolution of the job under processing from one period to the next follows a Markov chain.

The main difference of the problem considered here from common stochastic scheduling problems is that these J jobs have to be processed in a predetermined order. When one decides to stop processing job j and move onto job $(j+1)$, it is not allowed to return to job j later. An adaptive strategy is defined to be a sequence of random variables $\phi = \{\phi_t\}$ taking values in the set $\{1, 2, \dots, J\}$, such that the event $\{\phi_t = j\}$ (process job j at step t), $j = 1, \dots, J$ belongs to the σ-field $\mathscr{F}_{t-1}$ generated by $\phi_1, X_1, \dots, \phi_{t-1}, X_{t-1}$, where X_n denotes the state of the job being processed at the nth step. We shall study the problem of designing an adaptive strategy which maximizes

Received October 1999; revised October, 2000.

[1]Supported in part by the National Science Council of ROC.

[2]Visiting National Taiwan University and Academia Sinica. Supported in part by Hong Kong Research Grant Council.

AMS 2000 *subject classifications*. Primary 62L05; secondary 62N99.

Key words and phrases. Computerized adaptive tests, Markov chains, multiarmed bandits, Kullback–Leibler number, likelihood ratio, sequential testing, sequential design, Wald's equation.

the expected value of the sum of one-step rewards,

$$J_N(\theta) = E_\theta \sum_{t=1}^{N} g_{\phi_t}(X_t) \quad \text{as } N \to \infty, \tag{1.1}$$

under the constraint

$$\phi_t \le \phi_{t+1} \quad \text{for } 1 \le t \le N-1, \tag{1.2}$$

where N is the total processing time. Here E_θ denotes the expectation with initial distribution ν_j for each job $\gamma = 1, 2, \ldots, J$, which is assumed to be finite.

If the parameter θ were known, the best strategy would be to process only the job with greatest one-step expected reward. In ignorance of θ, an optimal adaptive strategy needs to trade off a reduced reward for better knowledge of the unknown parameter θ so that the total reward is as large as possible. In order to achieve the optimal trade-off, we first establish an asymptotic lower bound for the shortfall from complete information maximum reward then construct efficient strategies that achieve this lower bound.

Our formulation is motivated by a sequential design problem from computerized adaptive tests. Suppose we divide a computerized adaptive test into J stages. At each stage, a sequential sample of test items was selected from a stratum of test items with the same discrimination parameter value [see (5.1)] and the discrimination parameter value increases from one stage to the next. Within each stage, the selection was based on the up-and-down method on the difficulty parameter values. That is, if the examinee answers the current test item correctly, the difficulty level of the next item would increase by one unit and decrease by one unit otherwise. The number of items from each stage depends on a stopping rule with the purpose of maximizing the total Fisher information from the test for estimating the latent trait of the examinee. The first sequential aspect of the preceding design, the up-and-down method within each stage, is justified by its superior Fisher information per test item over the conventional test. The purpose of the second sequential aspect, stopping rules from one stage to the next, is two-fold. First, it can curb the exposure rate of test items with high discrimination parameter values. Second, when the latent trait value is away from the center, the stopping rule can reduce the information loss by using more test items with lower discrimination parameter values. The details are given in Section 5.

The rest of the paper is organized as follows. In Section 2, we describe the model and the assumptions. The asymptotic lower bound for the regret is derived in Section 3. In Section 4, we use a class of sequential tests to construct efficient strategies. In Section 5, the motivating example in computerized adaptive tests is discussed and analyzed using the method described in Sections 2–4.

2. The model.

2.1. *Relation to bandit problem.* It is possible to formulate the scheduling problem described in Section 1 as a bandit problem with constraint on arm-pulling. To be more precise, let each of the J jobs correspond to an arm in the bandit problem such that arms have to be pulled in a predetermined order: once an arm i is pulled, then any arm $j < i$ is no longer available in a later stage. Pulling an arm once corresponds to processing a job for a unit time. We now proceed to provide more details for the equivalence of the aforementioned control problem to an irreversible bandit problem.

Given J statistical populations $\Pi_1, \dots, \Pi_J$. For each $j = 1, \dots, J$, the observations from Π_j follow a Markov chain on a state space D with σ-algebra $\mathscr{D}$. It is assumed that the transition probability P_j^θ has a density function $p_j(x, y; \theta)$ with respect to some nondegenerate measure Q, where $p_j(x, y; \cdot)$ is known and θ is an unknown parameter belonging to some parameter space Θ. We also assume that the stationary measure of the Markov chain exists with the density function $\pi_j(\cdot; \theta)$ with respect to Q. Let $g_j(x)$ be the reward when population Π_j is sampled and x is observed. At each step we are required to pull one arm obeying the order constraint. The constraint (1.2) indicates that once a sample has been taken from Π_j, no further sampling is allowed from $\Pi_1, \dots, \Pi_{j-1}$. Then the adaptive strategy for the scheduling problem described in Section 1 corresponds to the adaptive irreversible allocation rule for the bandit problem, and the objective function for the bandit problem is the same as (1.1).

Note that the arms in the irreversible bandit problem considered here are "correlated," that is, when an arm is pulled, in addition to providing information on the reward distribution about the arm pulled it also gives distribution information about other arms as well. The difference between the "correlated" and the independent multiarmed bandit lies in the structure of the parameter space Θ. In the "correlated" multiarmed bandit problem, the parameter $\theta \in \Theta$ parameterizes *all* the arms Π_j, $j = 1, \dots, J$, whereas in the independent multiarmed bandit problem, the parameter $(\theta_1, \dots, \theta_J) \in \Theta$, and each θ_j parameterizes the *individual* arm Π_j for $j = 1, \dots, J$.

2.2. *Optimality criterion.* For each $j = 1, \dots, J$ and for all $\theta \in \Theta$. Let the initial state of the job under processing have distribution $\nu_j(\cdot; \theta)$. Let $\pi_j(\cdot; \theta)$ be the stationary distribution of the Markov chain and assume that $\int_{x\in D} \int_{y \in D} \times |g_j(y)| p_j(x, y; \theta) \pi_j(x; \theta)\, dQ(y)\, dQ(x) < \infty$ and let

$$\mu_j(\theta) = \int_{x\in D} \int_{y\in D} g_j(y) p_j(x, y; \theta) \pi_j(x; \theta)\, dQ(y)\, dQ(x) \tag{2.1}$$

be the one-step mean reward of job j under the stationary distribution. Let

$$T_N(j) = \sum_{t=1}^{N} I_{\{\phi_t = j\}} \tag{2.2}$$

be the amount of time job j is processed when the total processing time for all jobs is N. It follows from the Wald's equation for Markov chains (see Theorem 3 in the Appendix) that

$$J_N(\theta) = \sum_{t=1}^{N} \sum_{j=1}^{J} E_\theta\{E_\theta[g_j(X_t) I_{\{\phi_t = j\}} | \mathscr{F}_{t-1}]\} = \sum_{j=1}^{J} \mu_j(\theta) E_\theta T_N(j) + C, \tag{2.3}$$

where C is a constant. Here $E_\theta T_N(j)$ denotes the expectation under the initial distribution $\nu_j(\cdot, \theta)$, which is assumed to be finite.

REMARK. Equation (2.3) plays a pivotal role in our solution to the stochastic scheduling problem. That is, (2.3) allows us to reduce the problem to the study of stopping times $T_N(j)$, $j = 1, \dots, J$. Note that (2.3) as well as (A.5) of Theorem 3 in the Appendix give Wald's equation for Markov chains with arbitrary initial distributions. Here, the importance of Wald's equation for Markov chains with arbitrary initial distribution lies not so much on its generality but on the statistical significance. Because the stationary distribution depends on θ, which is unknown in practice, we have to start the Markov chain with an initial distribution independent of θ. Our proofs work for any initial distribution no matter whether it depends on θ or not. When the initial distribution is the stationary distribution, the constant C is zero by (A.5).

Hence, the objective of maximizing $J_N(\theta)$ is equivalent to that of minimizing the regret

$$\begin{aligned} R_N(\theta) &:= N\mu^*(\theta) - J_N(\theta) - C \\ &= \sum_{j:\, \mu_j(\theta) < \mu^*(\theta)} (\mu^*(\theta) - \mu_j(\theta)) E_\theta T_N(j), \end{aligned} \tag{2.4}$$

where $\mu^*(\theta) = \max_{1 \le j \le J} \mu_j(\theta)$.

The objective is to find an adaptive strategy ϕ that is optimal for all $\theta \in \Theta$ and large N. In general, no such strategy exists. Hence, we consider the class of all (asymptotically) *uniformly good* irreversible adaptive strategies, with regret satisfying

$$R_N(\theta) = o(N^\alpha) \quad \text{for all } \alpha > 0,\ \theta \in \Theta. \tag{2.5}$$

Such strategies have regret that does not increase too rapidly for any $\theta \in \Theta$. We would like to find a strategy that minimizes the increasing rate of the regret within the class of uniformly good irreversible adaptive strategies.

2.3. *Bad sets.* Here we shall describe the concept of bad set, which plays a central role in deriving the regret lower bound and constructing efficient adaptive strategies. For $j = 1, \dots, J$, define the Kullback–Leibler information number as

$$I_j(\theta, \theta') := \int_{x \in D} \int_{y \in D} \log \frac{p_j(x, y; \theta)}{p_j(x, y; \theta')} p_j(x, y; \theta) \pi_j(x; \theta)\, dQ(y)\, dQ(x). \tag{2.6}$$

Then, $0 \le I_j(\theta, \theta') \le \infty$, and we shall assume that $I_j(\theta, \theta') < \infty$, for all $\theta, \theta' \in \Theta$.

Let

$$\Theta_j := \{\theta \in \Theta: \mu_j(\theta) > \mu_i(\theta) \text{ for } i < j, \text{ and } \mu_j(\theta) \ge \mu_{i'}(\theta) \text{ for } i' > j\} \tag{2.7}$$

be the subset of Θ for which job j yields higher expected rewards than any job that comes before it and no lower expected rewards than any job comes after it. That is to say, Θ_j makes job j the first optimum job. Then the whole parameter space Θ can be decomposed as a union of Θ_j which means $\Theta := \bigcup_{j=1}^{J} \Theta_j$. For $\theta \in \Theta_j \subset \Theta$, we define the bad set

$$B(\theta) = \{\theta' \in \Theta \setminus \Theta_j: I_j(\theta, \theta') = 0\}. \tag{2.8}$$

Thus, $B(\theta)$ is the set of "bad" parameter values associated with θ, namely, those parameter values θ' for which the probability distribution of Π_j is the same under θ and θ', when the optimal strategy under θ is to process only job j while the optimal strategy under θ' is to process another job. The point is that if the true parameter is θ' when processing job j, we would be led to believe it were θ, and be trapped into believing job j is optimum unless we deliberately experiment with other seemingly inferior jobs. It is clear that, when $\theta \in \Theta_j$, in order for our problem, which is equivalent to a irreversible bandit problem, to have a solution, the bad set $B(\theta)$ cannot have nonempty intersection with Θ_i, $i = 1, \dots, j-1$. We shall discuss this assumption along with others in the following subsection.

2.4. *Assumptions.*

A1. For all $j = 1, \dots, J$ and $\theta \in \Theta$, the observations $\{X_{jt}, t \ge 0\}$ is a Markov chain on a general state space D with σ-algebra $\mathscr{D}$, which is irreducible with respect to a maximal irreducible measure on $(D, \mathscr{D})$ and aperiodic. Furthermore, we assume X_{jt} is *Harris recurrent* in the sense that there exists a set $E_j \in \mathscr{D}$, some $r_j \ge 1$, $\alpha_j > 0$ and a probability measure φ_j on E_j such that $P_j^{\theta}\{X_{jt} \in E_j \text{ i.o. } | X_{j0} = x\} = 1$ for all $x \in D$ and

$$P_j^{\theta}(X_{r_j} \in A | X_{j0} = x) \ge \alpha_j \varphi_j(A), \tag{2.9}$$

holds for all $x \in E_j$ and $A \in \mathscr{D}$.

For all $j = 1, \dots, J$ and $\theta \in \Theta$, let $a(t)$ be a probability distribution on the set of nonnegative integers, and let $K_{ja}^{\theta} = \sum_{t=0}^{\infty} a(t) P_j^{\theta^t}$, where $P_j^{\theta^t}$ denotes the t step transition of P_j^{θ}. A set $E_j \in D$ is called *petite*, if there exists a nontrivial measure φ_j on $\mathscr{D}$ such that

$$K_{ja}^{\theta}(x, A) \ge \varphi_j(A), \tag{2.10}$$

for all $x \in E_j$ and $A \in \mathscr{D}$. Assume that the parameter space Θ is a metric space and let ρ be its metric.

A2. For all $j = 1, \ldots, J$ and $\theta \in \Theta$, given a function $w_j\colon D \to [1, \infty)$, a petite set $E_j \in \mathscr{D}$, a constant $b_j < \infty$ and an extended real valued function $V_j\colon D \to [0, \infty]$ such that for all $x \in D$, we have

$$P_j^\theta V_j(x) \le V_j(x) - w_j(x) + b_j I_{E_j}(x),$$

where $P_j^\theta V_j(x) = \int V_j(y) P_j^\theta(x, dy)$ and I denotes indicator function.

A3. For all $\theta \in \Theta_{l+1}$, $1 \le j \le l$, $I_j(\theta, \theta')$ is continuous in $\theta' \in \bigcup_{j=1}^{l} \Theta_j$.

A4. If $\theta \in \Theta_{l+1}$, then $\inf_{\theta' \in B_j(\theta)} I_j(\theta, \theta') > 0$, for all $j = 1, \cdots, l$.

A5. For every $\theta \in \Theta$ and $j = 1, \ldots, J$, there exists $\delta_\theta > 0$ and $r_\theta > 2$ such that

$$\int_{x \in D} \int_{y \in D} \sup_{\theta' : \rho(\theta, \theta') \le \delta_\theta} \left| \log \frac{p_j(x, y; \theta)}{p_j(x, y; \theta')} \right|^{r_\theta} p_j(x, y; \theta) \pi_j(x; \theta) \times dQ(y)\, dQ(x) < \infty.$$

Assumption A1 is a recurrence condition general enough to allow a substantive theory and to cover interesting examples. Assumption A2 is a drift criterion for Markov chains, that ensures the boundedness of the constants appearing in the Wald's equations for Markov chains. Assumption A3 is a continuity assumption on $I_j(\theta, \theta')$. Assumption A4 assures that when sampling from a population one can consistently determine whether the current population is optimal or the optimal population is still ahead. Assumption A5 is a moment assumption on the log-likelihood ratio statistics.

2.5. *Related literature.* Although a dynamic programming method has been developed to solve the adaptive control problem described above and it can yield useful insight into some adaptive control problems, the difficulty of computation makes it less applicable. One reason for adopting the approach of this paper is to obtain an explicit (or semiexplicit) solution. Even though this paper is motivated by multistage computerized adaptive tests, the problem considered also relates to manufacturing job-shop and multiphase project management. We formulate the problem of irreversible bandits with Markovian rewarding, give an asymptotic lower bound for the regret to be defined below and provide an asymptotically efficient strategy. This approach was first introduced by Lai and Robbins (1985). When the irreversibility constraint is removed and the reward distribution is i.i.d., the preceding control problem is the classical multiarmed bandit problem; see Robbins (1952), Berry and Fristedt (1985) and Gittins (1989). When the observations in each arm are independent and identically distributed (i.i.d.) random variables, Hu and Wei (1989) studied the irreversible multiarm bandit problem, under a monotone structure on one-dimensional parameter space. The concept of bad sets, first introduced in Agrawal, Teneketzis and Anantharam (1989), plays an important role in our solution to the adaptive control problem. Other related works can be found in Anantharam, Varaiya and Walrand (1987), Gittins (1989), Glazebrook (1991, 1996) and Presman and Sonin (1990).

3. A lower bound for the regret. Note that for any fixed $j = 1, \ldots, J$, if

$$0 = I_j(\theta, \theta') = \int_{x \in D} \int_{y \in D} \log \frac{p_j(x, y; \theta)}{p_j(x, y; \theta')} p_j(x, y; \theta) \times \pi_j(x; \theta)\, dQ(y)\, dQ(x), \tag{3.1}$$

then $p_j(x, y; \theta) = p_j(x, y; \theta')$ for Q-almost all x and y, and therefore, $\mu_j(\theta) = \mu_j(\theta')$.

The following theorem gives an asymptotic lower bound for the regret (2.4) of uniformly good adaptive strategies.

THEOREM 1. *Assume that* A1 *holds and let* $\theta \in \Theta_{l+1}$ *for* $0 < l \le J-1$. *Then, for any uniformly good adaptive strategy* ϕ, *we have*

$$\liminf_{N \to \infty} R_N(\theta)/\log N \ge r(\theta, l), \tag{3.2}$$

where $r(\theta, l)$ *is the solution of the following minimization problem.*

PROBLEM A. *Minimize* $\sum_{j=1}^{l} (\mu^*(\theta) - \mu_j(\theta)) z_j(\theta)$, *subject to*

$$\inf_{\theta' \in B_1(\theta)} \{I_1(\theta, \theta') z_1(\theta)\} \ge 1,$$

$$\inf_{\theta' \in B_2(\theta)} \{I_1(\theta, \theta') z_1(\theta) + I_2(\theta, \theta') z_2(\theta)\} \ge 1,$$

$$\vdots$$

$$\inf_{\theta' \in B_l(\theta)} \{I_1(\theta, \theta') z_1(\theta) + \cdots + I_l(\theta, \theta') z_l(\theta)\} \ge 1$$

and

$$z_j(\theta) \ge 0 \quad \text{for } j = 1, 2, \ldots, l,$$

where $B_j(\theta) = B(\theta) \cup \Theta_j$, $j = 1, \ldots, J$.

REMARK. Agrawal, Teneketzis and Anantharam (1989) studied a controlled independent process with finite parameter space, and introduced a finite set $B(\theta)$ of bad parameter values associated with θ. In (2.8), for general parameter space Θ to be finite, we define the bad set $B(\theta)$ via the decomposition $\Theta = \bigcup_{j=1}^{n} \Theta_j$ and apply it to Theorem 1 for the regret lower bound of irreversible bandits with Markovian rewards.

Note that for a uniformly good adaptive strategy, we have for every $\theta \in \Theta$,

$$\lim_{N \to \infty} N^{-1} J_N(\theta) = \mu^*(\theta). \tag{3.3}$$

The strategies that satisfy (3.3) are said to be *consistent*. Under the assumptions of Theorem 1, the strategies that satisfy for $\theta \in \Theta_{l+1}$,

$$\sum_{j=l+2}^{J} E_\theta T_N(j) = O(1) \quad \text{if } l < J-1, \tag{3.4}$$

$$R_N(\theta) \sim r(\theta, l) \log N \quad \text{if } l > 0, \tag{3.5}$$

are said to be *asymptotically efficient*. The result of (3.4) implies that

$$R_N(\theta) = O(1) \quad \text{if } \theta \in \Theta_1. \tag{3.6}$$

This specifies the order of the regret when the best population is the first one, while (3.5) gives the order when it is not. The results (3.4) and (3.5) also imply that for all $\theta \in \Theta_{l+1}$,

$$E_\theta(N - T_N(l+1)) = O(\log N). \tag{3.7}$$

We need the following two lemmas for the proof of Theorem 1, where P_θ and E_θ denote the probability and expectation, respectively, under the initial distribution $\nu_j(\cdot, \theta)$.

LEMMA 1. *Assume that* A1 *holds. Let ϕ_N be a uniformly good irreversible allocation rule. Then, for every $\theta \in \bigcup_{j=l+1}^{J} \Theta_j$ and every $\theta' \in B_l(\theta)$,*

$$\liminf_{N\to\infty} \left[\sum_{j=1}^{l} I_j(\theta, \theta') E_\theta T_N(j) \right] \Big/ \log N \geq 1. \tag{3.8}$$

PROOF. To prove (3.8), it suffices to show that for every $\theta' \in B_l(\theta)$ and for $\alpha > 0$, $\delta > \alpha > 0$,

$$\lim_{N\to\infty} P_\theta \left\{ \sum_{j=1}^{l} I_j(\theta, \theta') T_N(j) \geq (1-\delta) \log N \right\} = 1. \tag{3.9}$$

Since ϕ_N is uniformly good, $E_{\theta'}(N - T_N(l)) = o(N^\alpha)$ for $\alpha > 0$.

Because $\theta' \in B_l(\theta)$ and $\theta \in \Theta_j$ for some $j = l+1, \ldots, J$, we have $\mu^*(\theta') > \mu_j(\theta')$ and $I_j(\theta, \theta') = 0$. Hence, $\mu_j(\theta) = \mu_j(\theta') < \mu^*(\theta')$ implying that $\mu^*(\theta') > \mu^*(\theta)$. For $\theta' \in \Theta_l$, $\mu_l(\theta') = \mu^*(\theta') > \mu^*(\theta) \geq \mu_l(\theta)$; therefore, $I_l(\theta, \theta') > 0$.

It then follows that for all large N,

$$
\begin{aligned}
&P_{\theta'}\left\{\sum_{j=1}^{l} I_j(\theta,\theta')T_N(j) < (1-\delta)\log N\right\}\\
&\quad = P_{\theta'}\left\{NI_l(\theta,\theta') - \sum_{j=1}^{l} I_j(\theta,\theta')T_N(j) > NI_l(\theta,\theta') - (1-\delta)\log N\right\}\\
&\quad \le \frac{E_{\theta'}\big[NI_l(\theta,\theta') - \sum_{j=1}^{l} I_j(\theta,\theta')T_N(j)\big]}{O(N)}\\
&\quad = \frac{1}{O(N)} E_{\theta'}\left[(N - T_N(l))I_l(\theta,\theta') - \sum_{j=1}^{l-1} I_j(\theta,\theta')T_N(j)\right]\\
&\quad = o(N^{\alpha-1}).
\end{aligned}
\tag{3.10}
$$

Let ν_j be the initial distribution of the Markov chain $\{X_{j1}, X_{j2}, \dots, X_{jn_j}\}$, which are the successive realizations from job J. And let

$$
\begin{aligned}
L_{n_1,\dots,n_l} = \sum_{j=1}^{l}\sum_{t=1}^{n_j} \log[&\nu_j(X_{j0};\theta)\\
&\times p_j(X_{jt}, X_{j(t+1)};\theta)/\nu_j(X_{j0};\theta')p_j(X_{jt}, X_{j(t+1)};\theta')]
\end{aligned}
$$

be the log likelihood ratio of θ with respect to θ'. Let

$$
\begin{aligned}
G_N = \Bigg\{ &\sum_{j=1}^{l} I_j(\theta,\theta')T_N(j) < (1-\delta)\log N\\
&\text{and } L_{T_N(1),\dots,T_N(l)} \le (1-\alpha)\log N\Bigg\}.
\end{aligned}
$$

Then, by (3.10), $P_{\theta'}(G_N) = o(N^{\alpha-1})$. By Wald's likelihood ratio identity for Markov chains,

$$
\begin{aligned}
&P_{\theta'}\{T_N(1) = n_1, \dots, T_N(l) = n_l, L_{n_1,\dots,n_l} \le (1-\alpha)\log N\}\\
&\quad = \int_{\{T_N(1)=n_1,\dots,T_N(l)=n_l,\, L_{n_1,\dots,n_l}\le(1-\alpha)\log N\}} \frac{dP_{\theta'}}{dP_\theta}\, dP_\theta\\
&\quad = \int_{\{T_N(1)=n_1,\dots,T_N(l)=n_l,\, L_{n_1,\dots,n_l}\le(1-\alpha)\log N\}} \exp(-L_{n_1,\dots,n_l})\, dP_\theta\\
&\quad \ge N^{\alpha-1} P_\theta\{T_N(1) = n_1, \dots, T_N(l) = n_l, L_{n_1,\dots,n_l} \le (1-\alpha)\log N\}.
\end{aligned}
\tag{3.11}
$$

We can then sum (3.11) over $n_1, \dots, n_l$ to get

$$
P_\theta(G_N) \le N^{1-a} P_{\theta'}(G_N) = N^{1-a} o(N^{a-1}) = o(1). \tag{3.12}
$$

Under Assumption A1 of the Markov chains, we have by the strong law of large numbers [cf. Theorem 17.0.1 of Meyn and Tweedie (1993)], as $\sum_{j=1}^{l} m_j \to \infty$,

$$\left| L_{m_1,\ldots,m_l} - \sum_{j=1}^{l} I_j(\theta, \theta')m_j \right| = o\left(\sum_{j=1}^{l} m_j \right) \quad \text{a.s. } P_\theta,$$

and by Lemma 2 below, we have as $\sum_{j=1}^{l} m_j \to \infty$,

$$\max_{\sum_{j=1}^{l} I_j(\theta,\theta')n_j \le \sum_{j=1}^{l} m_j} \left(\frac{L_{n_1,\cdots,n_l} - \sum_{j=1}^{l} I_j(\theta, \theta')n_j}{\sum_{j=1}^{l} m_j} \right) \longrightarrow 0 \quad \text{a.s. } P_\theta.$$

Since $1-\alpha > 1-\delta$, it then follows that as $N \to \infty$, $P_\theta\{L_{n_1,\ldots,n_l} > (1-\alpha)\log N$ for some $n_1, \ldots, n_l$ such that $\sum_{j=1}^{l} I_j(\theta, \theta')n_j < (1-\delta)\log N\} \to 0$.

Therefore,

$$P_\theta \left\{ \sum_{j=1}^{l} I_j(\theta, \theta')T_N(j) < (1-\delta)\log N \text{ and } L_{T_N(1),\ldots,T_N(l)} > (1-\alpha)\log N \right\} \longrightarrow 0.$$

This combined with (3.12) gives (3.9), from which (3.8) follows by letting $\delta \downarrow 0$. □

LEMMA 2. *Let S_{1n_1}, S_{2n_2} be two independent sequences of random variables with positive means μ_1 and μ_2, respectively. Let $n = n_1 + n_2$ such that $n_i/n \to \lambda_i$ for $i = 1, 2$. If, as $n \to \infty$,*

$$\frac{S_{1n_1} + S_{2n_2}}{n} \longrightarrow \mu = \lambda_1\mu_1 + \lambda_2\mu_2 \quad a.s. \tag{3.13}$$

then, as $n \to \infty$,

$$\max_{i\mu_1 + j\mu_2 \le n} \frac{S_{1i} + S_{2j}}{n} \longrightarrow \mu \quad a.s. \tag{3.14}$$

PROOF. For all $\varepsilon > 0$,

$$\begin{aligned} &P\left\{ \max_{i\mu_1 + j\mu_2 \le n} S_{1i} + S_{2j} \ge (1+\varepsilon)n\mu \right\} \\ &\quad \le P\left\{ \max_{i\mu_1 \le n_1} S_{1i} \ge \left(1 + \frac{\varepsilon}{2}\right)n_1\mu_1 + o(1),\ \max_{j\mu_2 \le n_2} S_{2j} \ge \left(1 + \frac{\varepsilon}{2}\right)n_2\mu_2 + o(1) \right\} \end{aligned}$$

$$= P\left\{\max_{i\mu_1 \le n_1} S_{1i} \ge \left(1+\frac{\varepsilon}{2}\right)n_1\mu_1 + o(1)\right\}$$
$$\times P\left\{\max_{j\mu_2 \le n_2} S_{2j} \ge \left(1+\frac{\varepsilon}{2}\right)n_2\mu_2 + o(1)\right\}$$
$$\longrightarrow 0 (\text{since as } n_k \to \infty, S_{kn_k}/n_k \to \mu_k \text{ a.s. for } k = 1, 2). \qquad \square$$

Applying Theorem 1 successively for $j = 1, \dots, l$, we have the following.

COROLLARY 1. *Assume that* A1 *holds. Let* ϕ_N *be a uniformly good irreversible allocation rule. Then, for every* $\theta \in \Theta_{l+1}$ *and* $\theta_j \in B_j(\theta)$ *for* $1 \le j \le l$, *we have*

$$\liminf_{N\to\infty} I_1(\theta, \theta_1) E_\theta T_N(1)/\log N \ge 1,$$
$$\vdots$$
$$\liminf_{N\to\infty} \sum_{j=1}^{l} I_j(\theta, \theta_l) E_\theta T_N(j)/\log N \ge 1.$$

Because our goal is to minimize $\sum_{j=1}^{l}(\mu^*(\theta) - \mu_j(\theta)) E_\theta T_N(j)$, Corollary 1 leads us to consider Problem A. Assume that $\Lambda_l = \Theta_1 \times \cdots \times \Theta_l$ is nonempty. For each $\lambda = (\theta_1, \dots, \theta_l) \in \Theta_1 \times \cdots \times \Theta_l$ and $\theta \in \Theta_{l+1}$. Problem A has a solution [cf. Duffin, Peterson and Zener (1967)] and we denote the minimum by $r(\theta, l, \lambda)$.

PROOF OF THEOREM 1. Assume Λ_l is nonempty; we first show that for every $\theta \in \Theta_{l+1}$,

$$\liminf_{N\to\infty} R_N(\theta)/\log N \ge \sup_{\lambda \in \Lambda_l} r(\theta, l, \lambda). \tag{3.15}$$

If $\liminf_{N\to\infty} R_N(\theta)/\log N = \infty$, then (3.15) clearly holds. On the other hand assume $\liminf_{N\to\infty} R_N(\theta)/\log N = c < \infty$, then $\liminf_{N\to\infty} E_\theta T_N(j)/\log N < \infty$ for $1 \le j \le l$ because $\mu^*(\theta) - \mu_j(\theta) \ge 0$ for $1 \le j \le l$ and $R_N(\theta) \ge \sum_{j=1}^{l}(\mu^*(\theta) - \mu_j(\theta)) E_\theta T_N(j)$. Therefore, we can choose a subsequence N_n such that $\lim_{n\to\infty} R_{N_n}(\theta)/\log N_n = c$ and $\lim_{n\to\infty} E_\theta T_{N_n}(j)/\log N_n = z_j(\theta)$, for $1 \le j \le l$. It is clear that

$$z_j(\theta) \ge 0 \quad \text{for } 1 \le j \le l \tag{3.16}$$

and

$$c \ge \sum_{j=1}^{l}(\mu^*(\theta) - \mu_j(\theta)) z_j(\theta). \tag{3.17}$$

For each $\lambda \in \Lambda_l$, by Corollary 1, we have

$$I_1(\theta, \theta_1) z_1(\theta) \geq 1,$$
$$\vdots$$
$$I_1(\theta, \theta_l) z_1(\theta) + \cdots + I_l(\theta, \theta_l) z_l(\theta) \geq 1.$$

This along with (3.16) leads to $\sum_{j=1}^{l} (\mu^*(\theta) - \mu_j(\theta)) z_j(\theta) \geq r(\theta, l, \lambda)$. Hence,

$$\sum_{j=1}^{l} (\mu^*(\theta) - \mu_j(\theta)) z_j(\theta) \geq \sup_{\lambda \in \Lambda_l} r(\theta, l, \lambda). \tag{3.18}$$

Combining (3.17) and (3.18), we obtain (3.15).

Next, we want to show that

$$r(\theta, l) = \sup_{\lambda \in \Lambda_l} r(\theta, l, \lambda). \tag{3.19}$$

For each $\lambda \in \Lambda_l$, $\theta_j \in B_j(\theta)$, $j = 1, \ldots, l$ and $\theta \in \Theta_{l+1}$, we have

$$\inf_{\theta' \in B_j(\theta)} I_i(\theta, \theta') \leq I_i(\theta, \theta_j), \qquad 1 \leq i \leq l,\ 1 \leq j \leq l. \tag{3.20}$$

Now let $z_j(\theta)$, $j = 1, \ldots, l$ be a solution of Problem A. In view of (3.20), $z_j(\theta)$, $j = 1, \ldots, l$ also satisfies

$$I_1(\theta, \theta_1) z_1(\theta) \geq 1,$$
$$\vdots$$
$$I_1(\theta, \theta_l) z_1(\theta) + \cdots + I_l(\theta, \theta_l) z_l(\theta) \geq 1.$$

Consequently,

$$r(\theta, l) = \sum_{j=1}^{l} (\mu^*(\theta) - \mu_j(\theta)) z_j(\theta) \geq r(\theta, l, \lambda).$$

Hence,

$$r(\theta, l) \geq \sup_{\lambda \in \Lambda_l} r(\theta, l, \lambda). \tag{3.21}$$

Next, we choose $\lambda_n = (\lambda_1(n), \ldots, \lambda_l(n)) \in \Lambda_l$ such that

$$\lim_{n \to \infty} \lambda_n = (\theta_1, \ldots, \theta_l). \tag{3.22}$$

Fix $\theta \in \Theta_{l+1}$. Let $z_n = (z_1(n), \ldots, z_l(n))$ be a solution of Problem A. Set

$$c_j(n) = \max\{I_j(\theta, \lambda_j(n))/I_j(\theta, \theta_j), \ldots, I_j(\theta, \lambda_l(n))/I_j(\theta, \theta_l)\}. \tag{3.23}$$

In view of (3.22) and (3.23),

$$\lim_{n \to \infty} c_j(n) = 1 \quad \text{for } 1 \leq j \leq l. \tag{3.24}$$

By (3.23), $(c_1(n)z_1(n), \dots, c_l(n)z_l(n))$ solve Problem A. Hence,

$$\left[\max_{1\le j\le l} c_j(n)\right] r(\theta, l, \lambda_n) \ge \sum_{j=1}^{l} (\mu^*(\theta) - \mu_j(\theta)) c_j(n) z_j(n) \ge r(\theta, l).$$

Applying (3.24), we obtain

$$\sup_{\lambda\in\Lambda_l} r(\theta, l, \lambda) \ge r(\theta, l).$$

Now (3.19) follows from this and (3.21). □

4. Construction of asymptotically efficient strategies. To understand the main idea behind our construction of efficient strategies, first note that the goal of any reasonable strategy is to determine whether the job currently under processing is optimum or not based on sequentially observed job states. The job under processing, say job j, is optimum if $\theta \in \Theta_j$. Thus, the problem of constructing an efficient strategy reduces to that of finding an optimum test of the hypothesis $\theta \in \Theta_j$ based on a sequential sample.

The efficient adaptive strategy to be described below is based on a sequential test. The lower bound discussed in Section 3 gives us valuable information about the sample size of the sequential test. In particular, it suggests that for $\theta \in \Theta_{l+1}$, the amount of processing time for job j ($j = 1, \dots, l$) should be at least of the order $(z_j(\theta) + o(1)) \log N$, where the $z_j(\theta)$ solve the following minimization problem:

$$\text{Minimize } \tilde{r}(\theta, l) = \sum_{j=1}^{l} (\mu^*(\theta) - \mu_j(\theta)) z_j(\theta), \tag{4.1}$$

subject to

$$\inf_{\theta'\in B_1(\theta)} \{I_1(\theta,\theta') z_1(\theta)\} = 1,$$

$$\inf_{\theta'\in B_2(\theta)} \{I_1(\theta,\theta') z_1(\theta) + I_2(\theta,\theta') z_2(\theta)\} = 1,$$

$$\vdots$$

$$\inf_{\theta'\in B_l(\theta)} \{I_1(\theta,\theta') z_1(\theta) + \cdots + I_l(\theta,\theta') z_l(\theta)\} = 1.$$

One can use sequential likelihood ratio tests of composite hypotheses in Markov chains to test the null hypothesis that $\theta \in \Theta_j$, with prescribed error probability of wrongly rejecting the null hypothesis and with asymptotically minimal expected sample size to correctly reject the null hypothesis when it is false.

To construct an irreversible asymptotically efficient adaptive strategy, we need to decide when to stop processing the current job and move to the next one. Let $T_N(j)$ be the amount of processing time for job j. To be more concrete, for each $1 \le j \le J$, let $\{X_{jt}\}$ be a random sample from Π_j. For any adaptive strategy ϕ, the associated $T_N(j)$ is an $\mathscr{F}_N(j)$-stopping time. Then $\phi = \{\phi_t\}$ is said to be an irreversible adaptive strategy if

$$\phi_t = l \quad \text{if } \sum_{j=0}^{l-1} T_N(j) < t \le \sum_{j=0}^{l} T_N(j) \quad \text{and } T_N(0) = 0. \tag{4.2}$$

Our goal is, therefore, to construct $\mathscr{F}_N(j)$-stopping time $T_N(j)$ so that the regret $R_N(\theta) = O(1)$ for $\theta \in \Theta_1$ and $R(\theta) \sim \sum_{j=1}^{l}(\mu^*(\theta) - \mu_j(\theta))z_j(\theta)\log N$ for $\theta \in \Theta_{l+1}$, under some regularity conditions. To this end, for each l, let F_l be a probability distribution with support $\bigcup_{j=l+1}^{J}\Theta_j$. For nonnegative integers $n_1, \dots, n_l$, let

$$U_l(n_1,\dots,n_l) = \frac{\int_{\bigcup_{j=l+1}^{J}\Theta_j} \prod_{j=1}^{l} \nu_j(X_{j0};\theta)\prod_{t=0}^{n_j} p_j(X_{jt}, X_{j(t+1)};\theta)\,dF_l(\theta)}{\sup_{\theta' \in B_l(\theta)} \prod_{j=1}^{l} \nu_j(X_{j0};\theta')\prod_{t=0}^{n_j} p_j(X_{jt}, X_{j(t+1)};\theta')}. \tag{4.3}$$

Define $T_N(j)$, $1 \le j \le J$, inductively by

$$\begin{aligned}
T_N(0) &= 0,\\
&\vdots\\
\tau_N(l) &= \inf\{n\colon U_l(T_N(1), \dots, T_N(l-1), n) > N\},\\
T_N(l) &= \min\left\{\tau_N(l), N - \sum_{j=1}^{l-1} T_N(j), \quad \text{for } 1 \le l < J\right\}, \qquad (4.4)\\
&\vdots\\
T_N(J) &= N - \sum_{j=1}^{J-1} T_N(j).
\end{aligned}$$

THEOREM 2. *Assume that* A1–A5 *hold and let* $\theta \in \Theta_{l+1}$. *Then, the strategy described in* (4.3) *and* (4.4) *satisfies*

$$\text{(i)} \qquad \sum_{j=l+2}^{J} E_\theta T_N(j) = O(1) \text{ if } l+1 < J, \tag{4.5}$$

$$\text{(ii)} \qquad \limsup_{N\to\infty} \inf_{\theta' \in B_l(\theta)} \left[\sum_{j=1}^{l} I_j(\theta, \theta')E_\theta T_N(j)\right] \Big/ \log N \le 1, \tag{4.6}$$

where $E_\theta T_N(j)$ *denote the expectation under the initial distribution* $\nu_j(\cdot, \theta)$.

Therefore, its regret satisfies

$$\liminf_{N\to\infty} R_N(\theta)/\log N = \tilde{r}(\theta, l), \tag{4.7}$$

where $\tilde{r}(\theta, l)$ *is the solution of the minimization problem* (4.1).

REMARKS. Consider the following additional assumption to Theorem 2. Suppose that for any $\theta \in \Theta_{l+1}$ and any $1 \le i \le l$,

$$\inf_{\theta' \in B_i} \{ I_j(\theta, \theta')/[\mu^*(\theta) - \mu_j(\theta)] \} \tag{4.8}$$

is monotone increasing as j increases for $1 \le j \le l$. Following an argument similar to that of Theorem 3 of Hu and Wei (1989), we can establish that $\tilde{r}(\theta, l) = r(\theta, l)$.

The proof of Theorem 2 involves a detailed analysis of the log-likelihood ratio statistics and suitable applications of Wald's equations for Markov chains. We leave the proof of Wald's equation for Harris recurrent Markov chains to the Appendix.

PROOF OF THEOREM 2. (i) Under Assumption A5 we can construct the twisting transformation (A.3) given in the Appendix. A simple change of measure argument, as that of Theorem 6.1 in Woodroofe (1982), leads to that for all $x \in D$,

$$P_\theta\{\tau_N(l) < \infty | X_0 = x\} \le \frac{1}{N}. \tag{4.9}$$

This implies that $\sup_x P_\theta\{\sum_{j=1}^{l} T_N(j) = N | X_0 = x\} \ge \sup_x P_\theta\{\tau_N(l) = \infty | X_0 = x\} \ge 1 - 1/N$. Consequently,

$$\sum_{j=l+1}^{J} E_\theta T_N(j) \le (J-l) N \sup_x P_\theta \left\{ \sum_{j=l+1}^{J} E_\theta T_N(j) > 0 | X_0 = x \right\} \le J - l. \tag{4.10}$$

(ii) For simplicity, in the remaining part of this proof, we shall use T_j in instead of $T_N(j)$. Here the initial distribution ν_j is a point mass at x_{j0} in (4.3). Let $S_{T_j} = \sum_{t=1}^{T_j} X_{jt}$, $T = \min\{\tau_N(l) - 1, T_l\}$, $S_T = \sum_{t=1}^{T} X_{lt}$ and

$$\begin{aligned} \Delta(\lambda, \gamma) = {} & \left(\sum_{j=1}^{l-1} (\lambda - \gamma) S_{T_j} - T_j(\Lambda_j(\lambda) - \Lambda_j(\gamma)) \right) \\ & + \Big((\lambda - \gamma) S_T - T(\Lambda_l(\lambda) - \Lambda_l(\gamma)) \Big), \end{aligned} \tag{4.11}$$

where $\Lambda_j := \log \lambda_j$ and λ_j is the maximal simple eigenvalue of the operator defined in (A.2) for Π_j. By the definition of T and the twisting transformation (A.3),

$$\begin{aligned}\log N &\geq \log U_l(T_1, \ldots, T_{l-1}, T)\\ &= \inf_{\theta' \in B_l(\theta)} \Bigg\{ \log \int_{\bigcup_{j=l+1}^{J} \Theta_j} \left(\prod_{j=1}^{l} \frac{r_j(X_{jT_j}; \lambda)}{r_j(x_{j0}; \lambda)} \frac{r_j(x_{j0}; \theta')}{r_j(X_{jT_j}; \theta')} \right)\\ &\qquad\qquad \times \exp(\Delta(\lambda, \theta'))\, dF_l(\lambda) \Bigg\} \qquad (4.12)\\ &= \inf_{\theta' \in B_l(\theta)} \Bigg\{ \Delta(\theta, \theta') + \log \left(\sum_{j=1}^{l} \frac{r_j(x_{j0}; \theta')}{r_j(X_{jT_j}; \theta')} \right)\\ &\qquad + \log \int_{\bigcup_{j=l+1}^{J} \Theta_j} \left(\prod_{j=1}^{l} \frac{r_j(X_{jT_j}; \lambda)}{r_j(x_{j0}; \lambda)} \frac{r_j(x_{j0}; \theta)}{r_j(X_{jT_j}; \theta)} \right) \exp(\Delta(\lambda, \theta))\, dF_l(\lambda) \Bigg\}.\end{aligned}$$

Since $\alpha_u := \max_{i=1,\ldots,l} \sup_{x \in D} r_i(x; \theta) < \infty$ and $\alpha_l := \min_{i=1,\ldots,l} \inf_{x \in D} r_i(x; \theta) > 0$, we obtain that

$$\prod_{j=1}^{l} \frac{r_j(X_{jn_j}; \lambda)}{r_j(x; \lambda)} \leq (\alpha_u / \alpha_l)^l. \qquad (4.13)$$

Combining this and (4.12), there exists a constant $C_1 > 0$ such that, for any δ such that θ is in the δ-neighborhood of $\bigcup_{j=1}^{l} \Theta_j$, we have

$$\begin{aligned}\log N \geq \inf_{\theta' \in B_l(\theta)} \Bigg\{ &\Delta(\theta, \theta') + \log \left(\prod_{j=1}^{l} \frac{r_j(X_{jT_j}; \theta)}{r_j(x_{j0}; \theta)} \frac{r_j(x_{j0}; \theta')}{r_j(X_{jT_j}; \theta')} \right)\\ &+ \log C_1 \int_{\lambda \in N_\delta(\theta)} \exp(\Delta(\lambda, \theta))\, dF_l(\lambda) \Bigg\},\end{aligned} \qquad (4.14)$$

where $N_\delta(\theta) := \{\lambda \colon \rho(\lambda, \theta) < \delta\}$ is a δ-neighborhood of θ. By Jensen's inequality,

$$\begin{aligned}&\int_{\lambda \in N_\delta(\theta)} \exp(\Delta(\lambda, \theta))\, dF_l(\lambda) / F_l(N_\delta(\theta))\\ &\qquad \geq \exp \int_{\lambda \in N_\delta(\theta)} \Delta(\lambda, \theta) dF_l(\lambda) / F_l(N_\delta(\theta)).\end{aligned} \qquad (4.15)$$

In view of (4.14) and (4.15),

$$\begin{aligned}\log N \geq &\inf_{\theta' \in B_l(\theta)} \left\{ \Delta(\theta, \theta') + \log \left(\prod_{j=1}^{l} \frac{r_j(X_{jT_j}; \theta)}{r_j(x_{j0}; \theta)} \frac{r_j(x_{j0}; \theta')}{r_j(X_{iT_i}; \theta')} \right) \right\}\\ &+ \log C_1 + \log F_l(N_\delta(\theta))\\ &+ \int_{\lambda \in N_\delta(\theta)} \Delta(\lambda, \theta)\, dF_l(\lambda) / F_l(N_\delta(\theta)).\end{aligned} \qquad (4.16)$$

Substituting $p_j(x, y; \theta)$ in the twisting formula (A.3) in (2.6) and a simple calculation yields

$$I_j(\lambda, \gamma) = (\lambda - \gamma)\mu_j(\lambda) - \big(\Lambda_j(\lambda) - \Lambda_j(\gamma)\big) + \int_{x\in D}\int_{y\in D} \log\left(\frac{r_j(y;\lambda)}{r_j(x;\lambda)}\frac{r_j(x;\gamma)}{r_j(y;\gamma)}\right) \times p_j(x, y; \lambda)\pi_j(x;\lambda)\,dQ(y)\,dQ(x). \tag{4.17}$$

Combining this with (4.13), there exists a constant C_2 such that

$$\Delta(\lambda, \gamma) = \left(\sum_{j=1}^{l-1}(\lambda - \gamma)(S_{T_j} - \mu_j(\lambda)T_j)\right) + (\lambda - \gamma)(S_T - \mu_l(\lambda)T) + \sum_{j=1}^{l-1} I_j(\lambda, \gamma)T_j + I_l(\lambda, \gamma)T + C_2. \tag{4.18}$$

Because $T_j \le N$ for all j, Wald's equation for Markov chains in Theorem 3(i) implies that there exists a constant C_3 such that

$$E_\theta\Delta(\theta, \theta') = \sum_{j=1}^{l-1} I_j(\theta, \theta')E_\theta T_j + I_l(\theta, \theta')E_\theta(T + 1) - (\theta - \theta')E_\theta(X_{l,T+1} - \mu_l(\theta)) + C_3. \tag{4.19}$$

By the Markov–Wald equation for squared sums in Theorem 3(ii) and Hölder's inequality, there exists a finite constant $C_4 > 0$, such that

$$E_\theta|X_{l,T+1} - \mu_l(\theta)| \le E_\theta\left[\sum_{t=1}^{T+1}(X_{lt} - \mu_l(\theta))^2\right]^{1/2} \le \sigma_\theta[E_\theta(T + 1)]^{1/2} + C_4. \tag{4.20}$$

In view of this, we have

$$E_\theta\int_{\lambda\in N_\delta(\theta)} \Delta(\lambda, \theta)\,dF_l(\lambda) \ge \left\{-\varepsilon\left[\sum_{j=1}^{l-1} E_\theta T_j + E_\theta(T + 1)\right] - C_5(E_\theta(T+1))^{1/2}\right\}F_l(N_\delta(\theta)) - C_4, \tag{4.21}$$

for a suitable constant C_5. Applying (4.19) and (4.21) to (4.16), we obtain that

$$\begin{aligned}\log N \geq \inf_{\theta' \in B_l(\theta)} & \left\{ \sum_{j=1}^{l-1} (I_j(\theta, \theta') - \varepsilon) E_\theta T_j + (I_l(\theta, \theta') - \varepsilon) E_\theta (T+1) \right\} \\ & - 2C_5 (E_\theta (T+1))^{1/2} + \log F_l(N_\delta(\theta)) \\ & - C_4 - (\theta - \theta') E_\theta (X_{l,T+1} - \mu_l(\theta)) + C_3. \end{aligned} \tag{4.22}$$

By the definition of T and T_l, $T \geq T_l - 1$. Note that by Assumption A4, $\inf_{\theta' \in B_j(\theta)} I_j(\theta, \theta') > 0$ for all $j = 1, \dots, l$. Since ε can be arbitrary small, (4.6) follows from (4.22).

In order to show that the regret satisfies (4.7), we first claim that

$$\liminf_{N \to \infty} \inf_{\theta' \in B_l(\theta)} \left[\sum_{j=1}^{l} I_j(\theta, \theta') E_\theta T_N(j) \right] \bigg/ \log N \geq 1. \tag{4.23}$$

To show (4.23), we shall apply Lemma 1. In order to apply Lemma 1, we note that by part (ii), of the proof, the adaptive strategy (4.3), (4.4) is uniformly good, and under Assumptions A3 and A4, we obtain (4.23).

Now, by (4.6) of Theorem 2 and (4.23), we have that

$$\liminf_{N \to \infty} \inf_{\theta' \in B_l(\theta)} \left[\sum_{j=1}^{l} I_j(\theta, \theta') E_\theta T_N(j) \right] \bigg/ \log N = 1. \tag{4.24}$$

This implies that any limit point of $(T_N(1)/\log N, \dots, T_N(l)/\log N)$ satisfies (4.1), therefore, (4.7) holds. □

5. An example. To illustrate how our method can be applied, we discuss here one concrete example: computerized adaptive testing. One drawback for the conventional test is that it cannot adaptively adjust the difficulty level of test items according to the ability of the examinee. In sharp contrast with the conventional tests, computerized adaptive tests allow the difficulty level of the next test item to depend on the results of previous responses.

A commonly used statistical criterion for item selection in computerized adaptive testing (CAT) is to maximize item information, estimated from currently available responses from the examinee. The standard three-parameter logistic (3-PL) item response model specifies that, for an examinee with latent trait, θ, the probability that he/she answers an item correctly ($Y = 1$) is

$$P(Y = 1|\theta) = c + (1 - c) \frac{1}{1 + \exp[-a(\theta - b)]}, \tag{5.1}$$

where a, b and c are item parameters and are, respectively, called the discrimination, difficult and guessing parameters [cf. Lord (1980) and Chang and Ying (1999b)].

A key aspect of CAT is how to adaptively select test items to maximize the Fisher information $I(\theta|a, b)$ at the currently estimated ability trait. In the

3-PL model, it is

$$\frac{(1-c)a^2\ \exp(a(\hat{\theta}-b))}{[1+\exp\{a(\hat{\theta}-b)\}]^2[1-c+c\{1+\exp(-a(\hat{\theta}-b))\}]} \tag{5.2}$$

at $\theta = \hat{\theta}$, the estimated ability trait.

It has been noted that the information-based selection rule could lead to extremely skewed item exposure rates in the sense that some items may be overexposed while others may never be used. Remedies to curb high exposure rates have been proposed and studied by Sympson and Hetter (1985), Stocking and Lewis (1995) and Chang and Ying (1999a).

In the case of the 2-PL model ($c = 0$), based on an idea of stratification, Chang and Ying (1999a) proposed a selection procedure for CAT. A modified version is described as follows.

1. Partition the item pool into J strata according to the value a of the discrimination parameter. The first stratum contains items with the smallest a's, the next stratum contains items with the second smallest a's and so on such that the last stratum contains those with the largest a's.
2. Accordingly, partition the entire test of length N into J stages.
3. In the jth stage, let $T_N(j)$ be the number of items selected from the jth stratum. The selection criterion within each stage is based on Dixon and Mood's (1948) up-and-down method on the item difficulty parameter. Starting from an initial difficulty level, a $0--1$ test is performed. When the response is 0 (wrong), we will decrease the difficulty level of the next item by one unit, while if the response is 1 (correct), we will increase the difficulty level by one unit for the next test item. The process continues until the stopping time $T_N(j)$, which is defined in (4.4). After the stopping rule $T_N(j)$ is satisfied, we move to stage $j+1$ and repeat the up-and-down selection method. Note that $T_1+\cdots+T_J = N$. This implies that the whole procedure is terminated when all N items have been selected.

Within each stage $j = 1,\dots,J$, suppose that the difficulty parameter b can take countably many infinite values, $b_1, b_2,\dots$ from the interval $(-3, 3)$ (commonly used in practice). Let the sequence of random variables $\{X_{jt},\ t = 1, 2,\dots\}$, each taking values in the set $\{b_1, b_2,\dots\}$, denote the sequence of random difficulty levels generated by the up-and-down method. Then $\{X_{jt},\ t \geq 0\}$ forms a Markov chain with infinitely many states. Due to the nature of the up-and-down method, the Markov chain X_{jt} visits a small interval of the true ability trait θ infinitely often, and the expected waiting time is finite. This implies that X_{jt} is a positive Harris recurrent Markov chain. Since $E_\pi Y_1 < \infty$, by Theorem 14.1.2 of Meyn and Tweedie (1993), the drift criterion A2 holds. In practice, we shall use a truncated version with a finite transition probability matrix.

For ease of exposition, assume that there are k difficulty levels $b_1, b_2,\dots,b_k$ with $b_i < b_j$ for $i < j$. Recall that Y_{jt} is the response variable for the tth item in stage j. Let $\{(x_{j0}, y_{j0}),\dots,(x_{jT_N(j)}, y_{jT_N(j)})\}$ be the corresponding observations in stage j. It is easy to see that $\{X_{jt}, t = 1, \ldots, T_N(j)\}$ form a Markov chain on the state space $\{b_1, \ldots, b_k\}$ with transition probability matrix

$$P = \begin{pmatrix} 1 - p_{j1} & p_{j1} & 0 & \cdots & 0 \\ 1 - p_{j2} & 0 & p_{j2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & p_{j(k-1)} \\ 0 & \cdots & 0 & 1 - p_{jk} & p_{jk} \end{pmatrix}, \tag{5.3}$$

where $p_{ji} = (1 + \exp\{-a_j(\theta - b_i)\})^{-1}$. For each $j = 1, \ldots, J$, solving the system of equations $p_{ji}\pi_j(i) = [1 - p_{j(i+1)}]\pi_j(i+1)$ for $i = 1, \ldots, k-1$ and $\pi_j(1) + \cdots + \pi_j(k) = 1$, we obtain the stationary distribution for the transition probability matrix (5.3) as $\pi_j(i) = Kd_{ji}$, where $d_{j1} = 1$, $d_{ji} = \prod_{t=2}^{i} p_{j(t-1)}/(1 - p_{jt})$ and $K = 1/(\sum_{i=1}^{k} d_{ji})$, for $i = 1, \ldots, k$.

In the three-parameter logistic model, the parameter space Θ consists of all possible values of the latent trait θ, which is one-dimensional. The probability of getting correct answers $P(Y = 1|\theta)$ is an increasing function of a if $\theta > b$ and a decreasing function of a if $\theta < b$. In the 2-PL model ($c = 0$), the selection criterion for CAT at each stage j is to maximize the expected Fisher information

$$\mathbf{I}_j(\theta) = \sum_{i=1}^{k} \pi_j(i) I(\theta|a_j, b_i) = a_j^2 \sum_{i=1}^{k} \pi_j(i) p_{ji}(1 - p_{ji}). \tag{5.4}$$

Typically, one uses $\theta \in (-4, 4)$, $a \in (0.5, 2)$ and $b \in (-3, 3)$.

There are two sequential aspects for the proposed mutistage CAT. First, within each stage, the up-and-down method sequentially chooses the difficulty level of the next test item. A detailed analysis of the expected Fisher information $\mathbf{I}_j(\theta)$ reveals the superiority of the up-and-down method over the conventional tests. That is, $\mathbf{I}_j(\theta)$ compares favorably to the total Fisher information obtained from the conventional tests using the same discrimination parameter value a_j for all test items. For example $\mathbf{I}_j(\theta)$ dominates the Fisher information obtained from the conventional test with evenly distributed difficulty levels. See Figure 1 for a demonstration of the superiority of the up-and-down method over the conventional test.

Second, the number of test items used in each stage is sequentially determined according to the stopping rule specified by (4.3) and (4.4). Further analysis shows that the up-and-down method gives higher Fisher information for higher values of a when θ values are close to the center of the interval (-4,4). When the values θ are away from the center, lower values of a may yield higher Fisher information. Actually, as shown by Figure 2, the set of θ values that makes the Fisher information $\mathbf{I}_j(\theta)$, corresponding to a particular discrimination parameter value a_j, highest moves away from the center as a_j deceases. For examinees with the latent trait values approaching extreme, the stopping rule (4.4) would prefer more test items for earlier stages, which correspond to strata with lower discrimination parameter values. Thus, using the

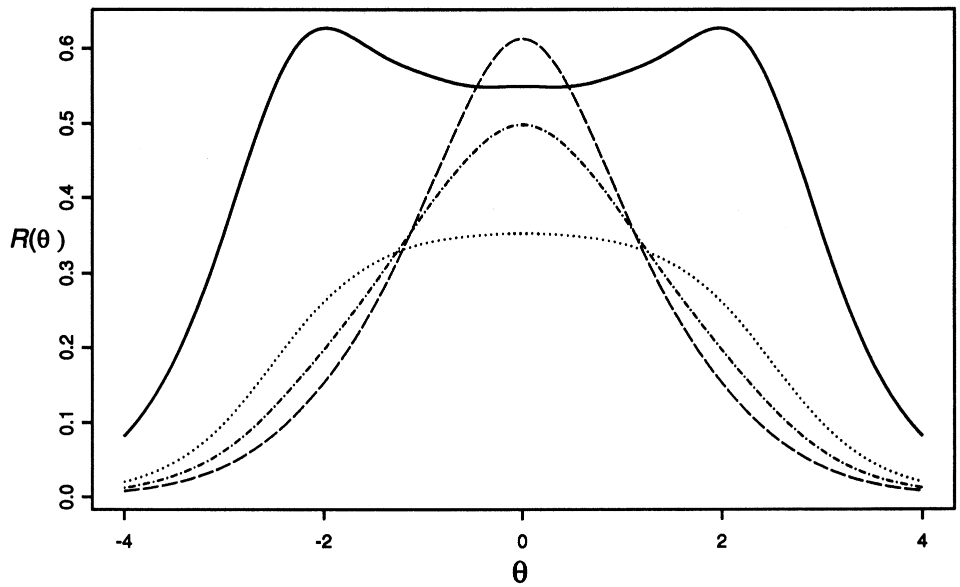

FIG. 1. *Fisher information for fixed and sequential designs*: *Plotted curves have the same discrimination parameter* $a = 1.8$. —— *up-and-down method on difficulty levels* $b \in \{-2, -1, 0, 1, 2\}$; - - - - - *fixed design with difficulty level distribution* 5%($b = -2$), 15%($b = -1$), 60%($b = 0$), 15%($b = 1$), 5%($b = 2$); $\dot{-} \cdot -\cdot$ *fixed design with difficulty level distribution* 10%($b = -2$), 20%($b = -1$), 40%($b = 0$), 20%($b = 1$), 10%($b = 2$); $\cdots\cdots$ *fixed design with uniform distribution over difficulty levels* $b \in \{-2, -1, 0, 1, 2\}$. *Obviously, the up-and-down method compared favorably to the fixed designs.*

stopping rule to determine the number of test items at each stage can reduce information loss when the ability level θ of the examinee is relatively high or low. Interestingly, this reduction of information loss is achieved by using more test items with lower values of a. Hence, as we have previously mentioned, the multi-stage CAT can help to curb the exposure rate of test items with high discrimination values.

It is not difficult to verify that the bad sets in this example are all empty sets. Hence $B_j(\theta) = \Theta_j$. We also found that Θ_J is an interval occupying the center part of $\Theta = (-4, 4)$, and Θ_{J-1} consists of two intervals lying to the immediate left and right of Θ_J. In the same way Θ_{J-2} consists of two intervals lying to the immediate left and right of Θ_{J-1}. The rest of Θ_j, $j = 1, 2, \ldots, J-3$, follows the same pattern. One can check that all assumptions in Section 2.4 hold for this example. Therefore, the efficient strategy described in Section 4 can be employed so that the information obtained from CAT is nearly optimum.

APPENDIX

Wald's equations for Markov chains. Let $\{X_n,\ n \geq 0\}$ be a Markov chain on a state space D with σ-algebra $\mathscr{D}$, which is irreducible with respect to

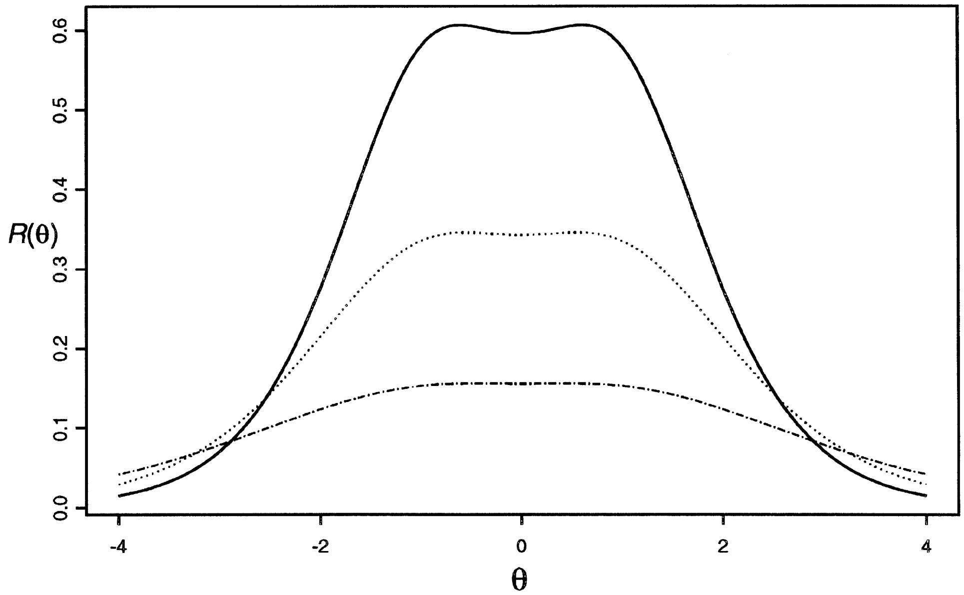

FIG. 2. *Fisher information for up-and-down method*: *Plotted curves correspond to three different discrimination parameters values*, ——- $a = 1.6$; $\cdots\cdots$ $a = 1.2$; $\cdot - \cdot - \cdot$ $a = 0.8$. *The difficulty levels* $b \in \{-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8\}$. *When the latent trait* θ *is relatively large* (>2.5) *or small* (< -2.5), *the Fisher information is larger for smaller discrimination parameter values*.

a maximal irreducibility measure φ on $(D, \mathscr{D})$ and is aperiodic. The transition probability kernel is denoted by $P(\cdot, \cdot)$ and assume X_n has stationary measure $\pi(\cdot)$. Let $S_n = \sum_{t=1}^{n} \xi_t$ be an additive component of X_n, taking values in R^k such that $\{(X_n, S_n),\ n \geq 0\}$ is a Markov chain on $D \times R^k$, with transition probability

$$P\{(X_{n+1}, S_{n+1}) \in A \times (B + s) | (X_n, S_n) = (x, s)\} = P(x, A \times B), \tag{A.1}$$

for all $x \in D$, $A \in \mathscr{D}$ and all $B \in \mathscr{B}(R^k)$, the Borel σ-algebra on R^k. The chain is called a *Markov random walk*.

The purpose of the present section is to establish the Wald's equation for Harris recurrent Markov chains (Assumption A1). It turns out that an even stronger result is possible. We shall prove the Wald's equation under the following condition which is weaker than Assumption A1.

(M) *Minorization condition.* There exists a family of measures $\{h(x, B); B \in \mathscr{B}(R^k)\}$ on R^k, for each $x \in D$, and a probability measure $\{\varphi(A \times B); A \in \mathscr{D}, B \in \mathscr{B}(R^k)\}$ on $D \times R^k$, such that for all $x \in D,\ A \in \mathscr{D}, B \in \mathscr{B}(R^k)$,

$$h(x, \cdot) * \varphi(A, \cdot)(B) \leq P(x, A \times B),$$

where $*$ denotes the convolution of two measures.

Let $x \in D$, $\theta \in R^k$ and let g be a given bounded measurable function on D. Define the linear operators $\mathbf{P}_\theta$, $\mathbf{P}$ by

$$(\mathbf{P}_\theta g)(x) = E_x\{e^{\theta\cdot\xi_1} g(X_1)\}, \qquad (\mathbf{P}g)(x) = E_x\{g(X_1)\}, \tag{A.2}$$

where "·" denotes the inner product. Without loss of generality, we may assume that $E_\pi e^{\theta\cdot\xi_1} < \infty$ for all $\theta \in R^k$. Under the minorization condition (M), Theorem 4.1 of Ney and Nummelin (1987) shows that $\mathbf{P}_\theta$ has a maximal simple eigenvalue $\lambda(\theta)$ with associated right eigenfunction $r(\cdot;\theta)$ such that $\Lambda(\theta) = \log\lambda(\theta)$ is analytic and strictly convex on R^k, and $r(\cdot;\theta)$ is uniformly positive, bounded and analytic on R^k for each $x \in D$. Note that when S_n defined in (A.1) is a deterministic function of X_n, (M) implies that the Harris recurrent assumption A1 of $\{X_n, n \geq 0\}$ holds. Now, for $\theta \in R^k$, define the "twisting" transformation for the transition probability of $\{X_n, n \geq 0\}$,

$$P_\theta(x, dy) = \frac{r(y;\theta)}{r(x;\theta)} e^{-\Lambda(\theta)+\theta\cdot\xi_1} P(x, dy). \tag{A.3}$$

If the function $\Lambda(\theta)$ is normalized so that $\Lambda(0) = \Lambda^{'}(0) = 0$, then $P_0 = P$ is the transition probability of the Markov chain $\{X_n, n \geq 0\}$ with invariant probability $\pi_0 = \pi$.

Let $\mathscr{F}_\infty = \sigma(X_0, (X_1, S_1), (X_2, S_2), \ldots)$. The following proposition is taken from Theorem 1 of Sadowsky (1989).

PROPOSITION 1. *Under* (M), *assume* $E_\pi e^{\theta\cdot\xi_1} < \infty$ *for all* $\theta \in R^k$. *Let* N *be any stopping time. Then, for any* $x \in D$, $B \in \mathscr{F}_N$ *and* $\theta \in R^k$,

$$\begin{aligned} &P_\theta\{B \cap (N < \infty) | X_0 = x\} \\ &\quad = \int_{B\cap(N<\infty)} \frac{r(X_N;\theta)}{r(x;\theta)} \exp(\theta S_N - N\Lambda(\theta))\, dP_x. \end{aligned} \tag{A.4}$$

A version of Wald's equations for uniformly ergodic Markov random walks can be found in Fuh and Lai (1998), where they applied the spectral theory of positive operators related to Markov semigroups. Fuh and Zhang (2000) first derived Poisson equations for Markov random walks and then applied them to establish Wald's equations. Here in (A.2) and (A.3), we applied results in Ney and Nummelin (1987) on Harris recurrent Markov random walks to obtain Wald's equations and characterize the boundedness of the constants appearing in (A.5) and (A.7) via Poisson equations.

THEOREM 3. *Assume that* (M) *and* A2 *of Section* 2.4 *hold for the Markov chain concerned with corresponding* V, w *and* b, *such that* $\int_D V(x)\, d\pi(x) < \infty$. *Assume* $E_\pi e^{\theta\cdot\xi_1} < \infty$ *for all* $\theta \in R^k$. *Let* N *be a stopping time such that* $E_\nu N < \infty$.

(i) *Suppose* $\sup_x E_x|\xi_1| < \infty$ *and let* $\mu = E_\pi\xi_1$. *Then,*

$$E_\nu S_N = \mu E_\nu N - E_\nu\{r'(X_N;0) - r'(X_0;0)\}, \tag{A.5}$$

where $r'(\cdot;\theta)$ denotes the first derivative of $r(\cdot;\theta)$ with respect to θ. Furthermore, $r'(x;0)$ is bounded on R^k, and is the solution g of the following Poisson equation:

$$(I-\mathbf{P})g=\mathbf{P}(E_x\xi_1-\mu), \tag{A.6}$$

where I is the identity operator and $\mathbf{P}$ is the operator defined in (A.1).

(ii) *Suppose* $\sup_x E_x\xi_1^2<\infty$ *and let* $\sigma^2=E_\pi(\xi_1-\mu)+2\sum_{t=1}^{\infty}\operatorname{Cov}_\pi(\xi_1-\mu,\xi_{1+t}-\mu)$. *Then*,

$$\begin{aligned}E_\nu(S_N-\mu N)^2&=\sigma^2E_\nu N+2E_\nu\{(S_N-\mu N)r'(X_N;0)\}\\&\quad+E_\nu\{r''(X_N;0)-r''(X_0;0)\},\end{aligned} \tag{A.7}$$

where $r''(\cdot,\theta)$ denotes the second derivative of $r(\cdot,\theta)$ with respect to θ. Furthermore, $r''(x;0)+r'^2(x;0)$ is bounded on R^k and is the solution of the following Poisson equation:

$$(I-\mathbf{P})g=\mathbf{P}(E_x\xi_1-\mu-E_\pi r'(X_1;0)+r'(x;0))^2. \tag{A.8}$$

PROOF. (i) We first establish (A.6). Since $r(\cdot;\theta)$ is an eigenfunction of $\lambda(\theta)$ with respect to the operator $\mathbf{P}_\theta$, we have $\mathbf{P}_\theta r(x;\theta)=\lambda(\theta)r(x;\theta)$, which implies that $E_x\{e^{\theta\cdot S_1}r(x;\theta)\}=\lambda(\theta)r(x;\theta)$. A one-term Taylor expansion for $\lambda(\theta)$ and $r(x;\theta)$ with respect to θ around 0 entails $\lambda(\theta)\cong 1+\mu\cdot\theta+o(|\theta|)$ and $r(x;\theta)\cong 1+r'(x;0)+o(|\theta|)$. Therefore,

$$\begin{aligned}&E_x(1+\theta\cdot\xi_1+o(|\theta|))(1+\theta r'(x;0)+o(|\theta|))\\&\quad=(1+\mu\cdot\theta+o(|\theta|))(1+\theta r'(x;0)+o(|\theta|))\\&\quad\Rightarrow\ E_xS_1+E_xr'(x;0)=\mu+r'(x;0)\\&\quad\Rightarrow\ (I-\mathbf{P})r'(x;0)=\mathbf{P}(E_x\xi_1-\mu).\end{aligned}$$

By Assumptions A1, A2 and $E_\pi|E_x\xi_1-\mu|<\infty$, the existence and boundedness of the solution $r'(x;0)$ for the Poisson equation (A.6) follows from (17.38) and Theorem 17.4.2 of Meyn and Tweedie (1993). Furthermore, $|r'(x;0)|\le|E_x\xi_1-\mu|$.

To verify (A.5), let $T(n)=\min(N,n)$. By Proposition 1 and Doob's optional stopping theorem, for $\theta\in R^k$, $E_x\{e^{\theta\cdot S_{T(n)}-\Lambda(\theta)T(n)}r(X_{T(n)};\theta)\}=r(x;\theta)$. Taking derivatives with respect to θ on both sides yields

$$\begin{aligned}E_x\big\{&(S_{T(n)}-\Lambda'(\theta)\cdot T(n))e^{\theta S_{T(n)}-\Lambda(\theta)T(n)}r(X_{T(n)};\theta)\\&+e^{\theta\cdot S_{T(n)}-\Lambda(\theta)T(n)}r'(X_{T(n)};\theta)\big\}=r'(x;\theta).\end{aligned} \tag{A.9}$$

We can interchange expectation and differentiation by the dominated convergence theorem, because $T(n)\le n$ and $\sup_{\theta\in R^k,\,x\in D}\{|\Lambda(\theta)|+|\Lambda'(\theta)|+|r(x;\theta)|+|r'(x;\theta)|+E_x\xi_1\}<\infty$. Setting $\theta=0$ in (A.9) and noting that $\mu=\Lambda'(0)$, we obtain by integrating with respect to ν that

$$E_\nu\{(S_{T(n)}-\Lambda'(0)T(n))+r'(X_{T(n)};0)\}=E_\nu r'(x;0).$$

By the dominated convergence, we have $E_\nu T(n) \to E_\nu N$ and that $E_\nu r'(X_{T(n)};0) \to E_\nu r'(X_N;0)$ as $n \to \infty$. By the monotone convergence theorem, we have $E_\nu(\sum_{j=1}^{T(n)} S_j^+) \to E_\nu(\sum_{j=1}^{N} S_j^+)$ and $E_\nu(\sum_{j=1}^{T(n)} S_j^-) \to E_\nu(\sum_{j=1}^{N} S_j^-)$ as $n \to \infty$. Hence applying the preceding argument separately to $\sum_{j=1}^{T(n)} S_j^+$ and $\sum_{j=1}^{T(n)} S_j^-$ gives the desired conclusion.

To prove (ii), note that $\sigma^2 = \Lambda''(0)$. Following an argument similar to that of Fuh and Lai (1998), taking derivatives with respect to θ in (A.9) and setting $\theta = 0$ yield the desired result. The same argument as (i) leads to (A.8). Because there exists a positive constant c such that $E_\pi(E_x\xi_1 - \mu - E_\pi r'(X_1;0) + r'(x;0))^2 \le c \sup_x E_x\xi_1^2 < \infty$, the existence and boundedness of the solution $r''(x;0)$ of (A.8) follows from (17.38) and Theorem 17.4.2 of Meyn and Tweedie (1993). Furthermore, we have $|r''(x;0)| \le (E_x\xi_1 - \mu - E_\pi r'(X_1;0) + r'(x;0))^2$. $\square$

REFERENCES

AGRAWAL, R., TENEKETZIS, D. and ANANTHARAM, V. (1989). Asymptotically efficient adaptive allocation schemes for controlled i.i.d. processes: finite parameter space. *IEEE Trans. Auto. Control* **35** 258–267.

ANANTHARAM, V., VARAIYA, P. and WALRAND, J. (1987). Asymptotically efficient allocation rules for the multi-armed bandit problem with multiple plays I. IID rewards; II. Markov rewards. *IEEE Trans. Auto. Control* **33** 968–982.

BERRY, D. A. and FRISTEDT, B. (1985). *Bandit Problems.* Chapman and Hall, London.

CHANG, H. H. and YING, Z. L. (1999a). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement* **26** 211–222.

CHANG, H. H. and YING, Z. L. (1999b). Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *Ann. Statist.*

DIXON, W. J. and MOOD, A. M. (1948). A method for obtaining and analyzing sensitivity data. *J. Amer. Statist. Assoc.* **43** 109–126.

DUFFIN, R. J. PETERSEN, E. L. and ZENER, C. (1967). *Geometric Programming.* Wiley, New York.

FUH, C. D. and LAI, T. L. (1998). Wald's equations, first passage times and moments of ladder variables in Markov random walks. *J. Appl. Probab.* **35** 566–580.

FUH, C. D. and ZHANG, C. H. (2000). Poisson equation, moment inequalities and r-quick convergence for Markov random walks. *Stochastic Process. Appl.* **87** 53–67.

GITTINS, (1989). *Multi-armed Bandit Allocation Indices*. Wiley, New York.

GLAZEBROOK, K. D. (1991). Strategy evaluation for stochastic scheduling problems with order constraints. *Adv. Appl. Probab.* **23** 86–104.

GLAZEBROOK, K. D. (1996). On the undiscounted tax problem with precedence constraints. *Adv. Appl. Probab.* **28** 1123–1144.

HU, I. and WEI, C. Z. (1989). Irreversible adaptive allocation rules. *Ann. Statist.* **17** 801–823.

LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22.

LORD, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Erlbaum, Hillsdale, NJ.

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer, New York.

NEY, P. and NUMMELIN, E. (1987). Markov additive processes I. Eigenvalue properties and limit theorems. *Ann. Probab.* **15** 561–592.

PRESMAN, E. L. and SONIN, I. N. (1990). *Sequential Control with Incomplete Information.* Academic Press, San Diego.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 1397–1409.

SADOWSKY, J. S. (1989). A dependent data extension of Wald's identity and its application to sequential test performance computation. *IEEE Trans. Inform. Theory* **35** 834–842.

STOCKING, M. L. and LEWIS, C. (1995). A new method of controlling item exposure in computerized adaptive testing. Research report 95-25, Educational Testing Service, Princeton, NJ.

SYMPSON, J. B. and HETTER, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. In *Proceeding of the 27th Annual Meeting of the Military Testing Association* 973–977. Navy Personal Research and Development Center, San Diego, CA.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI 115
TAIWAN
REPUBLIC OF CHINA

HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
CLEAR WATER BAY
KOWLOON
HONG KONG
E-MAIL: imichu@ust.hu