# IDENTIFICATION AND ADAPTIVE CONTROL

# OF MARKOV CHAINS

by

Vivek Borkar and Pravin Varaiya

Memorandum No. UCB/ERL M79/69

3 October 1979

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

# IDENTIFICATION AND ADAPTIVE CONTROL OF MARKOV CHAINS

Vivek Borkar and Pravin Varaiya
Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

ABSTRACT

Consider a countable state controlled Markov chain whose transition probability is specified up to an unknown parameter $\alpha$ taking values in a compact metric space A. To each $\alpha$ is associated a prespecified stationary control law $\zeta(\alpha)$. The adaptive control law selects at each time t the control action $\zeta(\alpha_t, x_t)$ where $x_t$ is the state and $\alpha_t$ is the maximum likelihood estimate of $\alpha$. The asymptotic behavior of this control scheme is investigated for the cases when the true parameter value $\alpha_0$ does or does not belong to A, and for the case when $\zeta$ is chosen to minimize an average cost criterion. The analysis uses an appropriate extension of the notions of recurrence to non-stationary Markov chains.

Research sponsored in part by the National Science Foundation Grant ENG79-03879 and the Joint Services Electronics Program Contract F49620-79-C-0178.

## 1. INTRODUCTION

We consider a controlled Markov chain $\{x_n, n \geq 0\}$, characterized by

(i) a countable state space $S = \{1,2,\ldots\}$,

(ii) A control parameter $z(i)$, taking values in a compact separable metric space $Z(i)$, for each $i$ in $S$,

(iii) An unknown parameter $\alpha$ taking values in a compact separable metric space $A$,

(iv) and a function $p(i,j;z,\alpha)$, $i,j \in S$, $z \in Z(i)$, $\alpha \in A$, which is the probability of transition from $i$ to $j$ when control $z$ is used and if $\alpha$ is the true parameter.

The following assumptions are made throughout; additional assumptions will be made later as needed.

<u>A1</u> For each $i,j$ $(p(i,j;z,\alpha)$ is continuous in $z,\alpha$.

<u>A2</u> The actual transition probabilities correspond to the parameter value $\alpha_0$. We do <u>not</u> assume a priori that $\alpha_0$ is in $A$.

<u>A3</u> If $p(i,j;z',\alpha_0) = 0$, then $p(i,j;z,\alpha) = 0$ for $z \in Z(i)$, $\alpha \in A$. If $p(i,j;z',\alpha_0) \neq 0$, then there is $\bar{\varepsilon} > 0$, independent of $i,j,z,\alpha$ such that $\bar{\varepsilon} < p(i,j;z,\alpha)\ [p(i,j;z,\alpha_0)]^{-1} < (\bar{\varepsilon})^{-1}$.

<u>A4</u> For any fixed values of $\alpha \in A$ and $z(i) \in Z(i)$, the Markov chain with stationary probabilities $p(i,j;z(i),\alpha)$ has a single communicating class which is positive recurrent.

A <u>control law</u> is any sequence of random variables $\{z_n, n \geq 0\}$ such that

(i) $z_n \in Z(x_n)$,

(ii) $z_n$ is measurable with respect to $\mathcal{F}_n = \sigma(x_0,\ldots,x_n)$.

(iii) $P\{x_{n+1} = k|\mathcal{F}_n\} = p(x_n,k;z_n,\alpha_0)$.

The framework above is intended to cover two situations. In the first, $\{z_n\}$ is a deterministic sequence representing a known non-stationarity of the Markov process $\{x_n\}$. In the second, $z_n$ is a random variable chosen on the basis of the known history, $\mathcal{F}_n$, of the state process in such a way as to satisfy some performance criterion. It is this second situation which will be addressed in Section 4. Since this is the main motivation for our work we elaborate a little more. More specifically, our interest is in adaptive control laws which are constructed as follows. Suppose that for each $\alpha$ we are given a control function $\zeta(\alpha) = [\zeta(\alpha,1),\zeta(\alpha,2),\ldots]$ such that if $\alpha$ is the true parameter then the sequence $z_n = \zeta(\alpha,x_n)$ results in a performance which is satisfactory (or optimal for some criterion). Next let $\alpha_n$ be the estimate of $\alpha_0$ at time n obtained by some estimating scheme. The adaptive control law given by this estimating scheme and the function $\zeta$ is the random sequence $z_n = \zeta(\alpha_n,x_n)$, $n \geq 0$. Such an adaptive law seem to have been rigorously first explored in the context of linear systems by Åström and Wittenmark [1] where it was called a self-tuning regulator. A similar scheme for finite state Markov chains was studied by Mandl [2] under the assumption $\alpha_0 \in A$. In [2] the control function $\zeta$ was chosen to minimize the time average of the expected cost and a class of estimators for $\alpha_0$ were considered on the basis of "contrast" functions. This class includes the maximum likelihood estimate (MLE). Mandl established the a.s. convergence of $\alpha_n$ to $\alpha_0$ and of the time average of the cost to the minimum, by imposing an additional condition which in the case of MLE is the following: for any $\alpha \neq \beta$ in A, there is $i \in S$ such that

$$[p(i,1;z,\alpha),p(i,2;z,\alpha),\ldots] \neq [p(i,1;z,\beta),p(i,2;z,\beta),\ldots] \text{ for all } z \in Z(i). \tag{1.1}$$

However such an "identifiability" condition may, in applications, be too restrictive as seen from the work on linear systems ([1,3], see also the discussion in [4]) where (1.1) does not hold and it was observed that the estimates need not converge and, even if they do, not necessarily to the true value $\alpha_0$.

The problem we consider is essentially the same as Mandl's in a more general setting: (i) S may be countable, (ii) (1.1) is relaxed, and (iii) the control function $\zeta$ used to specify the adaptive law is any arbitrary map so that no explicit reference to a cost function is needed. (The case where S and A are both finite has been treated in [4].) Our problem is less general than Mandl's in one sense, namely, we restrict ourselves to MLE and do not consider other estimators.

The paper is organized as follows. Section 2 consists of results on recurrence of controlled Markov chains. These results will be used in Sections 3,4,5. Section 3 introduces the likelihood ratio and the MLE for Markov chains and studies their asymptotic properties. These results are applied to adaptive laws in Section 4 under the assumption that $\alpha_0 \in A$. The behavior of the MLE when $\alpha_0 \notin A$ is examined in Section 5. Some concluding remarks are collected in Section 6.

## 2. RECURRENCE IN CONTROLLED MARKOV CHAINS

We begin with some definitions.

Definition 2.1 Let $\{A_n, n \geq 0\}$ be a sequence of random events and $\{I(A_n)\}$ the corresponding indicator functions. The sequence $\{A_n\}$ is rare along a sample path $\omega$ if $\lim \frac{1}{n} \sum_{m=0}^{n} I(A_n)(\omega) = 0$, and frequent otherwise, i.e. if $\overline{\lim}\, \frac{1}{n} \Sigma I(A_n)(\omega) > 0$. $\{A_n\}$ occurs almost always if $\{A_n^c\}$ is rare, where $A_n^c = \Omega - A_n$ is the complement of $A_n$. If $\{A_n\}$ is rare along all sample paths outside a set of zero probability it is <u>rare a.s.</u> <u>Frequent a.s.</u> and <u>almost always a.s.</u> are similarly defined.

<u>Definition 2.2</u> Let $\{y_n\}$ be a sequence of real numbers and y one of its limit points. Then y* is a <u>frequent limit point</u> of $\{y_n\}$ if for every neighborhood 0 of y*, the sequence of events $\{y_n \in 0\}$ is frequent. A limit point which is not frequent is rare.

The following lemma is useful to test whether a sequence is rare.

<u>Lemma 2.1</u> Let $\{a_n\}$ be a nonnegative sequence such that $n^{-1} \sum^n a_m \to 0$. For $\varepsilon > 0$ let $k_n^\varepsilon$ be the number of terms in $\{a_1,\ldots,a_n\}$ larger than $\varepsilon$. Then $n^{-1}k_n^\varepsilon \to 0$. If $\{a_n\}$ is bounded the converse also holds.

<u>Proof</u> $\frac{1}{n} \sum^{n-1} a_m \geq \frac{1}{n} \sum_{\{a_m<\varepsilon\}} a_m + \frac{1}{n} \sum_{\{a_m \geq \varepsilon\}} \varepsilon \geq \frac{1}{n} \sum_{\{a_m \geq \varepsilon\}} \varepsilon = \frac{\varepsilon}{n} k_n^\varepsilon.$

Hence $n^{-1}k_n^\varepsilon \to 0$. Conversely, suppose $a_n \leq M$ for all n and $n^{-1}k_n^\varepsilon \to 0$. Then

$$\frac{1}{n} \sum^{n-1} a_m = \frac{1}{n} \sum_{\{a_m<\varepsilon\}} a_m + \frac{1}{n} \sum_{\{a_m \geq \varepsilon\}} a_m \leq \varepsilon + \frac{M}{n} k_n^\varepsilon \to \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $n^{-1} \Sigma a_m \to 0$. ¤

Some properties of frequent limit points are collected below.

<u>Lemma 2.2</u> Let $\{a_n\}$ be a sequence in a compact metric space with metric d. Then

i) The set A* of limit points of $\{a_n\}$ is compact and $a_n \to A^*$, i.e. there is a sequence $\{a_n^*\}$ in A* such that $d(a_n, a_n^*) \to 0$.

ii) $\{a_n\}$ has at least one frequent limit point.

iii) The set A** of frequent limit points is compact.

iv) For any open set $0 \supset A^{**}$, the sequence of events $\{a_n \in 0\}$ occurs almost always.

<u>Proof</u> i) This is well known. ii) Let $\varepsilon > 0$ and cover A* by finitely many balls of radius $\varepsilon$, say $B_1,\ldots,B_{m_1}$. Then $\{a_n\}$ is eventually in

$\cup B_i$, and for at least one i, say $i = k$, the sequence $\{a_n \in B_k\}$ is frequent. Cover $\bar{B}_k$, the closure of $B_k$, by finitely many balls of radius $\varepsilon/2$, say $B_{k1},\ldots,B_{km_2}$. Then for some j the sequence $\{a_n \in B_k \cap B_{kj}\}$ is frequent. Cover $\overline{B_k \cap B_{kj}}$ by finitely many balls or radius $\varepsilon/4$. Continuing in this way we form a sequence of balls with radius $\varepsilon/2^n$, $n = 1,2,\ldots$ such that the sequence of events that $a_n$ belongs to any of this spheres is frequent. By compactness the centers of these balls have a limit point a* which it is easy to check must be a frequent limit point of $\{a_n\}$. (iii) Let $\tilde{a} \in \overline{A^{**}}$, and $\tilde{O}$ a neighborhood of $\tilde{a}$. Then there are $a^* \in A^{**}$ and a neighborhood $O^*$ of $a^*$ such that $a^* \in O^* \subset \tilde{O}$. Since $\{a_n \in O^*\}$ is frequent so is the sequence $\{a_n \in \tilde{O}\}$, hence $a \in A^{**}$. (iv) Let $B = O^c \cap A^*$. Then B is compact and $B \cap A^{**} = \phi$. Let $O_1$, $O_2$ be disjoint open sets such that $B \subset O_1$, $A^{**} \subset O_2$. For $\varepsilon > 0$ let $D_1',\ldots,D_m'$ be balls of radius $\varepsilon$ which cover B. Then $D_1 = D_1' \cap O_1,\ldots,D_m = D_m' \cap O_1$ is a finite cover of B that does not intersect $A^{**}$. Now if $\{a_n \in O^c\}$ is frequent then so is $\{a_n \in D_k\}$ for some k. Proceeding exactly as in (ii) we can find a frequent limit point in B. This contradicts $B \cap A^{**} = \phi$, hence $\{a_n \in O^c\}$ is rare. ⌑

Lemma 2.3 Let $\{n_k, k \geq 0\}$ be a sequence of positive integers such that $\overline{\lim}\, \frac{1}{n} \sum_{m=0}^{n} I(m = n_k \text{ for some } k) > 0$. Let $\{a_n\}$ be a sequence in a compact metric space. Then the subsequence $\{a_{n_k}, k \geq 0\}$ has a limit point which is a frequent limit of $\{a_n\}$.

Proof Let A* be the limit points of $\{a_{n_k}\}$. For $\varepsilon > 0$, cover A* by finitely many balls of radius $\varepsilon$, say $B_1,\ldots,B_m$. Then $a_{n_k} \in \cup B_i$ eventually and by the assumption regarding $\{n_k\}$ it follows that the sequence $\{a_n \in \cup B_i\}$ is frequent. Proceeding as in the proof of Lemma 2.2 (ii) leads to the result. ⌑

We can now define recurrence concepts for controlled Markov chains.

<u>Definition 2.3</u> A state $i \in S$ is said to be <u>recurrent</u> along a sample path $\omega$ if $x_n(\omega) = i$ infinitely often (i.o). If the sequence $\{x_n=i\}$ is frequent along $\omega$ it is said to be <u>positive recurrent</u> along $\omega$; otherwise it is <u>null recurrent</u> along $\omega$.

<u>Lemma 2.4</u> Recall the assumptions A1-A4. If $i \in S$ is recurrent on a set D of positive probability, then every state is recurrent a.s. on D. Moreover, for any j,k in S such that $p(j,k;z,\alpha_0) > 0$ for some (hence all) $z \in Z(i)$, the events $\{x_n=j, x_{n+1}=k\}$ occur i.o. a.s. on D.

<u>Proof</u> Suppose $p(i,i';z,\alpha_0) > 0$. The series $\sum^n I(x_m=i')$ and $\sum^n E\{I(x_m=i')|\mathcal{F}_{m-1}\} = \sum^n p(x_{m-1},i';z_{m-1},\alpha_0)$ converge or diverge together (see [5,pp. 96-97].) The latter series diverges on D since $\sum^n p(x_{m-1},i';z_{m-1},\alpha_0) \geq \min_{z\in Z(i)} p(i,i';z,\alpha_0) \sum^n I(x_m=i)$ and, by hypothesis, $\sum^n I(x_m=i) \to \infty$ on D. Hence $\sum^n I(x_m=i')$ diverges a.s. on D, i.e., i' is recurrent a.s. on D. Now by A4 there is a single communicating class. Hence for any $\ell \in S$ there is a finite path of strictly positive probability from i to $\ell$ so that a repeated application of the preceding argument establishes the a.s. recurrence of $\ell$ on D. The second part of the lemma is proved in a similar manner. ¤

<u>Lemma 2.5</u> If $i \in S$ is positive recurrent on a set D of positive probability, then every state is positive recurrent a.s. on D. Moreover, for any j,k in S such that $p(j,k;z,\alpha_0) > 0$ for $z \in Z(i)$, the sequence $\{x_n=j, x_{n+1}=k\}$ is frequent a.s. on D.

<u>Proof</u> Suppose $p(i,i';z,\alpha_0) > 0$. By the Martingale Stability Theorem ([6, p. 387]),

$$\frac{1}{n}\sum^{n}[I(x_m = i') - E\{I(x_m = i'|\mathcal{F}_{m-1}\}] = \frac{1}{n}\sum^{n}[I(x_m = i') - p(x_{m-1},i';z_{m-1},\alpha_0)]$$

$$\to 0 \text{ a.s.,}$$

and so

$$\overline{\lim}\,\frac{1}{n}\sum^{n} I(x_m = i') \geq [\min_{Z(i)} p(i,i';z,\alpha_0)]\ \overline{\lim}\,\frac{1}{n}\sum^{n} I(x_m = i) \text{ a.s.}$$

By hypothesis, the expression on the right is strictly positive on D. Hence i' is also positive recurrent a.s. on D. By arguing exactly as in the proof of Lemma 2.4 the remaining results may be established. ¤

The use of the word "recurrent" in Definition 2.3 is clearly appropriate. Use of the phrases "positive recurrent" and "null recurrent" is justified by the next result.

Lemma 2.6. For Markov chains with stationary transition probabilities and a single communicating class the preceding definitions of positive and null recurrence coincide with the usual ones.

Proof For fixed $i \in S$ let $y_n = \frac{1}{n}\sum^{n-1} I(x_m{=}i)$ so $y_n^{-1} = n[\sum^{n-1} I(x_m{=}i)]^{-1}$. Let $\tau_0$ be the first time i is reached and $\tau_k$, $k = 1,2,\ldots$, the kth return time for i. If $m_n$ is the number of visits to i up to time n, then

$$\frac{1}{m_n}\sum_{k=0}^{m_n}\tau_k \leq \frac{1}{y_n} \leq \frac{1}{m_n}\sum_{k=0}^{m_n+1}\tau_k.$$

Now $\tau_0$ is finite a.s. and the $\tau_k$, $k \geq 1$, are independent and identically distributed. By the strong law of large numbers, therefore,

$$\lim_n \frac{1}{n}\sum^{m_n}\tau_k = \lim_n \frac{1}{m_n}\sum^{m_n+1}\tau_k = E\tau_1 \text{ a.s.,}$$

where the possibility $E\tau_1 = \infty$ is included. Hence $\lim_n y_n = (E\tau_1)^{-1}$ a.s. Therefore $\lim y_n = 0$ iff $E\tau_1 = \infty$ iff i is null recurrent in either sense, and $\lim y_n > 0$ iff $E\tau_1 < \infty$ iff i is positive recurrent in either sense. ¤

We now introduce a condition which bears a resemblance to the notion of "tightness" of a family of distributions.

Condition T There exists a null set N and for each $\varepsilon > 0$ there exists $J_\varepsilon < \infty$ such that $\overline{\lim}\ \frac{1}{n} \sum_{m=0}^{n} I(x_m = i, i > J_\varepsilon) < \varepsilon$ for every sample path $\omega \notin N$.

The next result is immediate.

Lemma 2.7 (i) For a finite state Markov chain condition T is always satisfied. (ii) For a Markov chain with stationary transition probabilities positive recurrence implies condition T.

Lemma 2.8 Under condition T all states are positive recurrent a.s.

Proof Suppose $i \in S$ is null recurrent on a set D of positive probability. By Lemma 2.5 all states are null recurrent a.s. on D i.e., $n^{-1} \sum^{n} I(x_m = j) \to 0$ a.s. on D for all $j \in S$. Hence for any

$$\frac{1}{n} \sum_{m=0}^{n-1} I(x_m = j, j \leq J) = \frac{1}{n} \sum_{j=1}^{J} \sum_{m=0}^{n-1} I(x_m = j) \to 0 \text{ a.s. on D,}$$

so that condition T cannot hold. ¤

To obtain a condition in terms of the transition probabilities $p(i,j;z,\alpha_0)$ which implies condition T we need the following. Let $\zeta = [z(1), z(2), \ldots]$ with $z(i) \in Z(i)$ be a fixed control function and consider the control law $\{z_n\}$ with $z_n = z(x_n)$. Such a law $\{z_n\}$, or equivalently $\zeta$, will be called a stationary law. Consider the following assumption.

A5 There is a finite number M such that for any stationary law $\zeta$, there is a state $s_\zeta$ such that the expected time to hit $s_\zeta$ from any i in S is bounded by M.

This is the familiar condition which guarantees the existence of an optimal stationary law in Markov decision processes with the time average cost criterion. (See, e.g. [7, pp. 147-148].) In [8], Federgruen, Hordijk and Tijms have given many equivalent formulations of

A5. In particular, a simple modification of their proof shows that under A1-A5, A5 is equivalent to A5'.

A5' Let $\{\pi(i,\zeta), i \in S\}$ be the invariant probabilities under the stationary law $\zeta$ and let $p_{ij}^n(\zeta) = P\{x_n = j | x_0 = i, \zeta \text{ is used}\}$. Then $\lim_n \frac{1}{n} \sum_{m=1}^{n} p_{ij}^n(\zeta) = \pi(j,\zeta)$ uniformly in $i,\zeta$ for each $j$ in S.

We can now obtain the following useful results.

Lemma 2.9 Under A1-A5 the set of probability measures $\{\pi(i,\zeta), i \in S\}$ on S for $\zeta \in Z = \prod_{i \in S} Z(i)$ is tight i.e. for $\varepsilon > 0$ there exists $J_\varepsilon < \infty$ such that $\sum_{i<J_\varepsilon} \pi(i,\zeta) < \varepsilon$ for all $\zeta \in Z$.

Proof It is easily seen that for each $n,i,j$ $p_{i,j}^n(\zeta)$ is continuous on Z. Since, by A5', $\frac{1}{n} \sum^{n} p_{i,j}^m \to \pi(j,\zeta)$ for each j uniformly in $i,\zeta$, therefore $\pi(j,\zeta)$ is continuous in $\zeta$ for each j. Hence for $J < \infty$, $\sum_{j \le J} \pi(j,\zeta)$ and $\sum_{j>J} \pi(j,\zeta) = 1 - \sum_{j \le J} \pi(j,\zeta)$ are both continuous in $\zeta$. Now as $J \to \infty$, $\sum_{j>J} \pi(j,\zeta)$ decreases monotonically to zero. Since Z is compact under the product topology, it follows by Dini's theorem that $\sum_{j>J} \pi(j,\zeta) \to 0$ as $J \to \infty$, uniformly in $\zeta$. Hence $\lim_J \sup_\zeta \sum_{j>J} \pi(j,\zeta) = 0$ and the result follows. ¤

Lemma 2.10 Let $c(i,j,z)$ be a continuous, nonnegative, bounded function defined for $(i,j,z) \in S \times S \times Z(i)$. Then, under A1-A5, for any control law $\{z_n\}$,

$$\lim_n \frac{1}{n} \sum^{n-1} c(x_m, x_{m+1}, z_m) \le \max_{\zeta \in Z} E_\zeta \, c(x_m, x_{m+1}, \zeta(x_m)) \text{ a.s.}$$

where $E_\zeta$ denotes expectation with respect to the stationary probabilities $\{\pi(i,\zeta)\}$.

Proof This follows from well-known results treating $c(\cdot,\cdot,\cdot)$ as the reward function of a Markov decision process. ¤

<u>Lemma 2.11</u> Under A1-A5 condition T holds.

<u>Proof</u> By the Martingale Stability Theorem, for any j,

$$\lim \frac{1}{n} \sum^{n-1} [I(x_m = j, j > J) - E\{I(x_m = j, j > J) | \mathcal{F}_{m-1}\}] = 0 \text{ a.s.}$$

so

$$\overline{\lim}\ \frac{1}{n} \Sigma\ I(x_m = j, j > J) = \overline{\lim}\ \frac{1}{n} \Sigma\ E\ I(x_m = j, j > J) | \mathcal{F}_{m-1})$$

$$= \overline{\lim}\ \frac{1}{n} \sum_{m=0}^{n-1} \sum_{j>J} p(x_m, j; z_m, \alpha_0) \leq \max_{\zeta \in Z} \sum_{i \in S} \pi(i,\zeta) \sum_{j>J}$$

$$p(i,j;\zeta(i),\alpha_0) \text{ a.s.} \tag{2.1}$$

The inequality in (2.1) follows from Lemma 2.10 by choosing $c(i,j,z) = \sum_{j>J} p(i,j;z,\alpha_0)$. As in the proof of Lemma 2.9 one can show that

$$\lim_{J} \max_{\zeta} \sum_{j>J} p(i,j;\zeta(i),\alpha_0) = 0 \quad \text{for all } i \text{ in } S \tag{2.2}$$

From Lemma 2.9 there exists $J_1$ such that

$$\sum_{i>J_1} \pi(i,\zeta) < \frac{1}{2}\varepsilon,\ \zeta \in Z.$$

From (2.2) there exists $J_2$ such that

$$\max_{i \leq J_1} \max_{\zeta} \sum_{j>J_2} p(i,j;\zeta(i),\alpha_0) \leq \frac{1}{2J_1} \varepsilon, \text{ so } \sum_{i \in S} \pi(i,\zeta) \sum_{j>J_2} \rho(i,j;\zeta(i),\alpha_0)$$

$$\leq \sum_{i \leq J_1} \pi(i,\zeta) \sum_{j>J_2} p(i,j;\zeta(i),\alpha_0) + \sum_{i>J_1} \pi(i,\zeta) < \varepsilon,\ \zeta \in Z.$$

Using this estimate in (2.1) gives

$$\overline{\lim}\ \frac{1}{n} \Sigma\ I(x_m = j, j > J_2) < \varepsilon \text{ a.s.}$$

Let $N_\varepsilon$ be the null set where this inequality fails, and let $N = \bigcup_{k=1}^{\infty} N_{\varepsilon/k}$. Then condition T holds outside of N. ¤

Lemma 2.12 Suppose A1-A5 and let N be the null set on which condition T fails. Let $\tilde{\omega} \notin N$. Suppose $\{n_k, k\geq 0\}$ is a subsequence of the integers such that

$$\overline{\lim}\ \frac{1}{n} \sum_{m=0}^{n-1} I(m = n_k \text{ for some } k) = \delta > 0. \tag{2.3}$$

Then there is $i \in S$ such that the sequence of events $\{m=n_k$ for some $k$ and $x_m=i,\ m\geq 0\}$ is frequent along $\tilde{\omega}$.

Proof Choose $0 < \tilde{\delta} < \delta$ and $\tilde{J} < \infty$ such that

$$\overline{\lim}\ \frac{1}{n} \Sigma\, I(x_m = j, j > \tilde{J}) < \tilde{\delta} \text{ for } \omega \notin N. \tag{2.4}$$

Suppose, contrary to the assertion, that

$$\frac{1}{n} \sum_{m=0}^{n-1} I(m = n_k \text{ for some } k,\ x_m \leq \tilde{J}) \to 0 \text{ along } \tilde{\omega}.$$

Then

$$\frac{1}{n} \sum_{m=0}^{n-1} I(m = n_k \text{ for some } k,\ x_m > \tilde{J}) = \delta > \tilde{\delta}, \text{ along } \tilde{\omega},$$

and, a fortiori, to

$$\overline{\lim}\ \frac{1}{n} \Sigma\, I(x_m = j, j > \tilde{J}) \geq \delta \text{ along } \tilde{\omega},$$

thereby contradicting (2.4). ¤

These results form the basis of the proofs in subsequent sections.

## 3. LIKELIHOOD RATIOS

We recall from [9] some facts about likelihood ratios. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\tilde{P}$ another probability on $(\Omega, \mathcal{F})$. Then there exists an integrable function $\Lambda \geq 0$ on $(\Omega, \mathcal{F}, P)$ called the likelihood ratio of $\tilde{P}$ to $P$ and a set $N$ with $P(N) = 0$ so that

$$\tilde{P}(A) = \int_A \Lambda dP + P(A \cap N),\ A \in \mathcal{F}.$$

$\Lambda$ is also denoted by $\frac{d\tilde{P}}{dP}$. $\Lambda$ is unique up to P-equivalence and N is unique up to $(P+\tilde{P})$-equivalence. Define the relations $\tilde{P} << P$, $P \perp P$ respectively, by $\tilde{P}(N) = 0$, $\tilde{P}(N) = 1$ (or, equivalently $\Lambda = 0$ P-a.s.). Let $\tilde{P} \equiv P$ mean $\tilde{P} << P$ and $P << \tilde{P}$.

Let $\{\mathcal{F}_n, n \geq 0\}$ be an increasing family of $\sigma$-fields such that $\mathcal{F} = \sigma(\cup \mathcal{F}_n)$. Let $P_n$, $\tilde{P}_n$ be the restrictions of P, $\tilde{P}$ to to $\mathcal{F}_n$ and let $\Lambda_n = \frac{d\tilde{P}_n}{dP_n}$. Then $\tilde{P} << P$ implies $\tilde{P}_n << P_n$. Also $\Lambda_n = E\{\Lambda | \mathcal{F}_n\}$P-a.s. . Thus $\{n, \mathcal{F}_n, P\}$ is a martingale and $\Lambda_n \to \Lambda$ P-a.s.

In our problem $\Omega$ is the set of all sequences $\{x_n\}$ in S and $\mathcal{F}_n = \sigma(x_0, \ldots, x_n)$. Each $\alpha \in A$ defines a probability measure $P^\alpha$. It is easy to check that

$$\Lambda_n(\alpha) = \frac{dP_n^\alpha}{dP_n^{\alpha_0}} = \sum_{m=0}^{n-1} \frac{p(x_m, x_{m+1}; z_m, \alpha)}{p(x_m, x_{m+1}; z_m, \alpha_0)} .$$

From the above-quoted facts it follows that for each $\alpha \in A$ there is a function $\Lambda(\alpha) = dP^\alpha | dP^{\alpha_0}$ such that $\Lambda_n(\alpha) \to \Lambda(\alpha)$ a.s. (with respect to $P^{\alpha_0}$).

<u>Lemma 3.1</u> On the set $\{\Lambda(\alpha) > 0\}$, $\frac{p(x_n, x_{n+1}; z_n, \alpha)}{p(x_n, x_{n+1}; z_n, \alpha_0)} \to 1$ a.s.

<u>Proof</u> Simply note that $\frac{\Lambda_{n+1}(\alpha)}{\Lambda_n(\alpha)} = \frac{p(x_n, x_{n+1}; z_n, \alpha)}{p(x_n, x_{n+1}; z_n, \alpha_0)}$. ⌑

<u>Corollary 3.1</u> Suppose the Markov chain has stationary transition probabilities and is recurrent. Then either $\Lambda(\alpha) = 0$ a.s. and $P^\alpha \perp P^{\alpha_0}$ or $\Lambda(\alpha) = 1$ a.s. and $P^\alpha \equiv P^{\alpha_0}$ in which case the transition probabilities under $\alpha$, $\alpha_0$ are identical.

Define $L_n(\alpha) = \ln \Lambda_n(\alpha)$. $L_n$ is the log-likelihood ratio.

<u>Definition 3.1</u> For each $\omega$ let $\alpha_n = \alpha_n(\omega) \in A$ be such that $L_n(\alpha_n) \geq L_n(\alpha)$, $\alpha \in A$. $\alpha_n$ is called the <u>maximum likelihood estimate</u>

(MLE) at time n. If the maximum value of $L_n(\alpha)$ is achieved at more than one value, we assume that only one of these is chosen according to some prescribed rule which ensures that $\alpha_n$ is $\mathcal{F}_n$-measurable.

Lemma 3.2 $\Lambda_n(\alpha_n)$ is a positive submartingale. If $\alpha_0 \in A$, then $\Lambda_n(\alpha_n) \geq 1$ a.s. and

$$P^{\alpha_0}\{[\Lambda_n(\alpha_n) \to \infty] \cup [\Lambda_n(\alpha_n) \text{ converges to a finite value}]\} = 1.$$

Proof The first statement can be verified directly. The second follows from well-known convergence properties of positive submartingales (see, e.g. [5, pp. 89-91]). ¤

For the remainder of this section assume that A1-A5 hold and $\alpha_0 \in A$.

Lemma 3.3 $\frac{1}{n} L_n(\alpha_n) \to 0$ a.s.

Proof $L_n(\alpha_n) \geq L_n(\alpha_0) = 0$. Hence

$$\underline{\lim}\ \frac{1}{n} L_n(\alpha_n) \geq 0 \text{ a.s.} \tag{3.1}$$

From Lemma 2.10, and for any fixed $\alpha$ in A, an appropriate choice of the function c gives

$$\overline{\lim_n}\ \frac{1}{n} L_n(\alpha) \leq \max_\zeta \sum_i \pi(i,\zeta) \sum_k p(i,k;\zeta(i),\alpha_0) \ln \frac{p(i,k;\zeta(i),\alpha)}{p(i,k;\zeta(i),\alpha_0)} \text{ a.s.}$$

$$\leq 0 \text{ a.s., by Jensen's inequality.}$$

Let $\tilde{A}$ be a countable dense subset of A. By the preceding inequality there is a null set N outside of which $\overline{\lim}\ \frac{1}{n} L_n(\alpha) \leq 0$ for all $\alpha \in \tilde{A}$. By Lemma 2.11 condition T holds outside a null set which, we may assume, is included in N. Then for $\varepsilon > 0$ and $\omega \notin N$, there exists J such that

$$\overline{\lim}\ \frac{1}{n} \Sigma\ I(x_m > J) < \frac{\varepsilon}{8K} \tag{3.2}$$

where $K > 0$ is any number with $K > |\ln \bar{\varepsilon}|$ and $\bar{\varepsilon}$ is as in A3. Fix $\alpha$ in A. By continuity there exists $\tilde{\alpha} \in \tilde{A}$ such that

$$\left|\ln \frac{p(i,k;z,\alpha)}{p(i,k;,z,\alpha_0)} - \ln \frac{p(i,k;z,\tilde{\alpha})}{p(i,k;z,\alpha_0)}\right| < \frac{\varepsilon}{2}, \tag{3.3}$$

for all $i,k \leq J$, and $z \in Z(i)$. Hence

$$\left|\frac{1}{n} L_n(\alpha) - \frac{1}{n} L_n(\tilde{\alpha})\right| \leq \frac{1}{n}\sum_{m=0}^{n-1} \left|\ln \frac{p(x_m,x_{m+1};z_m,\alpha)}{p(x_m,x_{m+1},z_m,\alpha_0)} - \ln \frac{p(x_m,x_{m+1};z_m,\tilde{\alpha})}{p(x_m,x_{m+1};z_m,\alpha_0)}\right|$$

$$= \frac{1}{n}\Sigma \left|\ln \frac{p(x_m,x_{m+1};z_m,\alpha)}{p(x_m,x_{m+1};z_m,\alpha_0)} - \ln \frac{p(x_m,x_{m+1};z_m,\tilde{\alpha})}{p(x_m,x_{m+1};z_m,\alpha_0)}\right| [I(x_m < J \text{ or } x_{m+1} > J)$$

$$+ I(x_m \leq J \text{ and } x_{m+1} \leq J)]\} \leq \varepsilon$$

using (3.2), (3.3). Since $\varepsilon > 0$ is arbitrary and $\overline{\lim}\, \frac{1}{n} L_n(\tilde{\alpha}) \leq 0$, $\omega \notin N$ it follows that

$$\overline{\lim}\ \frac{1}{n} L_n(\alpha) \leq 0 \text{ for all } \alpha \in A \text{ and } \omega \notin N. \tag{3.4}$$

Now suppose there is $\omega \notin N$ and $\varepsilon > 0$ and a subsequence $\{\alpha_{n_k}\}$ of $\{\alpha_n(\omega)\}$ such that

$$\frac{1}{n_k} L_{n_k}(\alpha_{n_k}) \geq \varepsilon \quad \text{for all } k.$$

Since A is compact we may suppose $\alpha_{n_k} \to \alpha^*$. Then, an argument similar to the above may be employed to show that

$$\overline{\lim_{k}}\ \frac{1}{n_k} L_{n_k}(\alpha^*) \geq \varepsilon,$$

thereby contradicting (3.4). Hence it must be that $\overline{\lim}\, \frac{1}{n} L_n(\alpha_n) \leq 0$ for $\omega \notin N$ which, with (3.1), proves the assertion. ⌑

<u>Theorem 3.1</u> There is a null set N such that for all $\omega \notin N$ and any limit point $\alpha^*$ of $\{\alpha_n(\omega)\}$, $\frac{1}{n} L_n(\alpha^*) \to 0$.

<u>Proof</u> By Lemma 3.3 there is a null set $N_1$ outside which $n^{-1}L_n(\alpha_n) \to 0$. For each rational number $r > 0$ and integer J cover A by open sets $0_{r-1}(J),\ldots,0_{rm_r}(J)$ such that if $\alpha$, $\alpha'$ are in $0_{r\ell}(J)$ then

$$\left| \ln \frac{p(i,j;z,\alpha)}{p(i,j;z,\alpha_0)} - \ln \frac{p(i,j;z,\alpha')}{p(i,j;z,\alpha_0)} \right| < \frac{r}{2} \text{ for } i,j \leq J,\ z \in Z(i) \tag{3.5}$$

(Here the left hand side is taken to be 0 if $p(i,j;z,\alpha_0) = 0$.) Following Lemma 3.2, $\sup\{\Lambda_n(\alpha) | \alpha \in O_{r\ell}(J)\}$ is a submartingale which converges to a finite value or to $+\infty$ outside of a null set $N_{r\ell}(J)$. Let $N_2 = \bigcup_{r\ell J} N_{r\ell}(J)$. By Lemma 2.11, condition T holds outside a null set $N_3$. Let $N = N_1 \cup N_2 \cup N_3$. Let $\omega \notin N$ and $\alpha^*$ be a limit point of $\{\alpha_n(\omega)\}$.

Let $Q_n(\alpha) = \ln \dfrac{p(x_n,x_{n+1};z_n,\alpha)}{p(x_n,x_{n+1};z_n,\alpha_0)}$. Then $L_n(\alpha) = \sum^{n-1} Q_m(\alpha)$, and the and the assertion would be proved if $Q_n(\alpha^*) \to 0$. Suppose, in contradiction, that there are $\varepsilon > 0$ and a subsequence $\{n_k, k > 0\}$ with $|Q_{n_k}(\alpha^*)| > \varepsilon$ for all k. There are two possibilities.

<u>Case 1</u> There exists $J_1$ such that $\max(x_{n_k}, x_{n_k+1}) \leq J_1$ i.o.

<u>Case 2</u> For every $\varepsilon > 0$ and subsequence $\{n_k\}$ with $|Q_{n_k}(\alpha^*)| > \varepsilon$, for every $J$, $\max(x_{n_k}, x_{n_k+1}) > J$ eventually.

Suppose Case 1 occurs. Let $r < \varepsilon$ be rational and $J_2$ such that

$$\overline{\lim}\ \frac{1}{n} \sum^{n-1} I(x_m > J_2) < \frac{r}{8K} \tag{3.6}$$

where $K \geq |\ln \bar{\varepsilon}|$ and $\bar{\varepsilon}$ is as in A3. Let $J = \max(J_1, J_2)$. Suppose $\alpha^* \in O_{r\ell}(J)$, and let $\tilde{\alpha}_n \in \overline{O_{r\ell}(J)}$ satisfy $L_n(\tilde{\alpha}_n) \geq L_n(\alpha)$, $\alpha \in O_{r\ell}(J)$. Then

$$\begin{aligned} Q_n(\tilde{\alpha}_n) &= L_{n+1}(\tilde{\alpha}_n) - L_n(\tilde{\alpha}_n) \leq L_{n+1}(\tilde{\alpha}_{n+1}) - L_n(\tilde{\alpha}_n) \\ &= Q_n(\tilde{\alpha}_{n+1}) + L_n(\tilde{\alpha}_{n+1}) - L_n(\tilde{\alpha}_n) \leq Q_n(\tilde{\alpha}_{n+1}), \end{aligned}$$

so

$$Q_n(\tilde{\alpha}_n) \leq L_{n+1}(\tilde{\alpha}_{n+1}) - L_n(\tilde{\alpha}_n) \leq Q_n(\tilde{\alpha}_{n+1}). \tag{3.7}$$

Since $|Q_{n_k}(\alpha^*)| > \varepsilon > r$, it follows from (3.5) that $Q_{n_k}(\tilde{\alpha}_{n_k}) > \frac{r}{2}$ or $Q_{n_k}(\tilde{\alpha}_{n_k}) < -\frac{r}{2}$. In either case it follows from (3.7) that $\{L_n(\tilde{\alpha}_n)\}$ cannot converge to a finite value and so $L_n(\tilde{\alpha}_n) \to \infty$. Hence

$\underline{\lim} \frac{1}{n} L_n(\tilde{\alpha}_n) \geq 0$. But $\overline{\lim} \frac{1}{n} L_n(\tilde{\alpha}_n) \leq \overline{\lim} \frac{1}{n} L_n(\alpha_n) = 0$, by Lemma 3.3. Therefore

$$\lim \frac{1}{n} L_n(\tilde{\alpha}_n) = 0. \tag{3.8}$$

Now

$$\overline{\lim} \left|\frac{1}{n} L_n(\tilde{\alpha}_n) - L_n(\alpha^*)\right| \leq \overline{\lim} \frac{1}{n} \sum^{n-1} |Q_m(\tilde{\alpha}_m) - Q_m(\alpha^*)| I(\max(x_m,x_{m+1}) > J)$$

$$+ \overline{\lim} \frac{1}{n} \sum^{n-1} |Q_m(\tilde{\alpha}_m) - Q_m(\alpha^*)| I(\max(x_m,x_{m+1}) \leq J) \leq r$$

from (3.6) and (3.5). Since r is arbitrary, it follows from (3.8) that $\frac{1}{n} L_n(\alpha^*) \to 0$ as required.

Now suppose Case 2 occurs. Let $\varepsilon > 0$ such that $|Q_n(\alpha^*)| > \varepsilon$ i.o. Let $0 < \delta < \varepsilon$ and $\{n_k, k \geq 0\}$ be the maximal subsequence such that $|Q_{n_k}(\alpha^*)| \geq \delta$. Let J be such that $\overline{\lim} \frac{1}{n} \sum^{n-1} I(x_m > J) < \delta$. By hypothesis there is only a finite number of k's such that $\max(x_{n_k}, x_{n_k+1}) \leq J$ and so $|Q_m(\alpha^*)| I(\max(x_m,x_{m+1}) \leq J) \leq \delta$ eventually. Hence

$$\overline{\lim} \frac{1}{n}|L_n(\alpha^*)| \leq \overline{\lim} \frac{1}{n} \sum^{n-1} |Q_m(\alpha^*)| I(\max(x_m,x_{m+1}) \leq J)$$

$$+ \overline{\lim} \frac{1}{n} \sum^{n-1} |Q_m(\alpha^*)| I(\max(x_m,x_{m+1}) > J)$$

$$\leq \delta + 2K \overline{\lim} \frac{1}{n} \sum^{n-1} I(\max(x_m,x_{m+1}) > J) \leq (1 + 2K)\delta,$$

where K is as before. It follows that $\lim \frac{1}{n} L_n(\alpha^*) = 0$. ¤

<u>Corollary 3.2</u> There exists a null set N such that for every $\omega \notin N$, $\varepsilon > 0$, and limit point $\alpha^*$ of $\{\alpha_n(\omega)\}$, the sequence $\left\{\left| \ln \frac{p(x_n,k;z_n,\alpha^*)}{p(x_n,k;z_n,\alpha_0)} \right| > \varepsilon \text{ for some } k\right\}$ is rare.

<u>Proof</u> For each $\alpha \in A$ let

$$G_n(\alpha) = \sum_{k \in S} p(x_n,k;z_n,\alpha_0) \ln \frac{p(x_n,k;z_n,\alpha)}{p(x_n,k;z_n,\alpha_0)}$$

(As usual, we set the last term to 0 if $p(x_n,k;z_n,\alpha_0) = 0$.) It is easy to show that $G_n(\alpha)$ is continuous in $\alpha$. Let

$$B_n(\alpha) = \sum^{n-1} G_m(\alpha).$$

By the Martingale Stability Theorem $\lim \frac{1}{n}[L_n(\alpha) - B_n(\alpha)] = 0$ a.s. Let $\tilde{A}$ be a countable dense subset of A and $N_1$ a null set outside which $\lim \frac{1}{n}[L_n(\tilde{\alpha}) - B_n(\tilde{\alpha})] = 0$ for every $\tilde{\alpha} \in \tilde{A}$. Let $N_2$ be the null set in the statement of Theorem 3.1 and $N_3$ the null set outside which condition T holds. Let $N = N_1 \cup N_2 \cup N_3$. Let $\omega \notin N$ and $\alpha^*$ a limit point of $\{\alpha_n(\omega)\}$.

Let $\delta > 0$. Let J be such that $\overline{\lim}\, \frac{1}{n} \sum^{n-1} I(x_m > J) < \delta$. Define

$$h(i,j;z,\alpha) = \ell n \frac{p(i,j;z,\alpha)}{p(i,j;z,\alpha_0)} - \sum_{k \in S} p(i,k;z,\alpha_0) \ell n \frac{p(i,k;z,\alpha)}{p(i,k;z,\alpha_0)}$$

It is easy to see that $h(i,j;z,\alpha)$ is continuous in $\alpha$, and let $\tilde{\alpha} \in \tilde{A}$ be such that $|h(i,j;z,\alpha^*) - h(i,j;z,\tilde{\alpha})| < \delta$ for all $i,j \leq J$ and $z \in Z(i)$. Note that

$$L_n(\alpha) - B_n(\alpha) = \sum^{n-1} h(x_m,x_{m+1};z_m,\alpha)$$

so

$$\begin{aligned}
\overline{\lim}\ \frac{1}{n}|L_n(\alpha^*) - B_n(\alpha^*) - [L_n(\tilde{\alpha}) - B_n(\tilde{\alpha})]| &= \overline{\lim}\, \frac{1}{n}\Big|\sum^{n-1} [h(x_m,x_{m+1};z_m,\alpha^*) \\
&\quad - h|(x_m,x_{m+1};z_n,\tilde{\alpha})]\Big| \leq \overline{\lim}\, \frac{1}{n} \Sigma\ |[h(x_m,x_{m+1};z_m,\alpha^*) \\
&\quad - h(x_m,x_{m+1};z_m,\tilde{\alpha})]I(\max(x_m,x_{m+1}) > J)| + \overline{\lim}\, \frac{1}{n} \Sigma\ |[h(x_m,x_{m+1};z_m,\alpha^*) \\
&\quad - h(x_m,x_{m+1};z_m,\tilde{\alpha})]I(\max(x_m,x_{m+1}) \leq J)| \\
&\leq 8K\ \overline{\lim}\, \frac{1}{n} \Sigma\ I(\max(x_m,x_{m+1}) > J) + \delta \leq (8K + 1)\delta,
\end{aligned}$$

where K is as in the proof of Theorem 3.1. $\delta > 0$ being arbitrary, it follows that

$$\lim \frac{1}{n}[L_n(\alpha^*) - B_n(\alpha^*)] = \lim \frac{1}{n}[L_n(\tilde{\alpha}) - B_n(\tilde{\alpha})] = 0.$$

By Theorem 3.1, this implies

$$\lim \frac{1}{n} B_n(\alpha^*) = \lim \frac{1}{n} \sum_{m=0}^{n-1} G_m(\alpha^*) = 0$$

Now

$$-G_m(\alpha^*) = \sum_k p(x_m,k;z_n,\alpha) \frac{p(x_m,k;z_m,\alpha_0)}{p(x_m,k;z_m,\alpha)} \ell n \frac{p(x_m,k;z_m,\alpha_0)}{p(x_m,k;z_m,\alpha)} .$$

The result follows from the strict convexity of the function $x \ \ell n \ x$, Jensen's inequality and Lemma 2.1. ¤

The next result is an immediate corollary of the preceding result.

Theorem 3.2 Suppose the Markov chain $\{x_n\}$ has stationary transition probabilities and is positive recurrent. Then there is a null set N such that for every $\omega \notin N$, $\varepsilon > 0$ and limit $\alpha^*$ of $\{\alpha_n(\omega)\}$, the transition probabilities under $\alpha^*$ and $\alpha_0$ coincide.

The application of Theorem 3.1 and Corollary 3.2 to adaptive control appears in the next section.

## 4. ADAPTIVE CONTROL

Throughout this section it is assumed that A1-A5 hold and $\alpha_0 \in A$. We also assume given for each $\alpha$ in A a stationary law $\xi(\alpha) = [\xi(\alpha,1),\xi(\alpha,2),\ldots]$. The actual law is given by the adaptive law $z_n = \xi(\alpha_n,x_n)$, $n \geq 0$. The particular choice of $\xi(\alpha)$ is not relevant in most of the subsequent discussion, but we shall consider the interesting case when $\xi(\alpha)$ is chosen to minimize, assuming $\alpha$ is the true parameter, the average cost of the form $\overline{\lim} \frac{1}{n} \sum_{m=0}^{n-1} c(x_m,x_{m+1},z_m)$.

In [4] we considered the case where A, S are both finite and the following result was obtained which is stronger than what seems possible in the more general setting considered here.

Theorem 4.1 [4] Under the scheme above, there exist a random variable $\alpha^*$ and a random time $N < \infty$ a.s. such that for almost all $\omega$, $\alpha_n(\omega) = \alpha^*(\omega)$, $n \geq N(\omega)$ and $p(i,j;\xi(\alpha^*(\omega),i),\alpha^*(\omega)) = p(i,j;\xi(\alpha^*(\omega),i),\alpha_0)$ for i,j in S. Moreover, if $\xi(\alpha)$ minimizes the average cost under $\alpha$, then the true cost $\lim \frac{1}{n} \sum^{n-1} c(x_m,x_{m+1},z_m) = J(\alpha^*)$ a.s. where $J(\alpha)$ is the cost corresponding to the stationary law $\xi(\alpha)$.

Thus in the finite case the adaptive law in "stable" in the sense that the parameter estimate $\alpha_n$, and the average cost $\frac{1}{n} \sum^{n-1} c(x_m,x_{m+1},z_m)$ converge. However, the limiting cost $J(\alpha^*)$ may exceed $J(\alpha_0)$ which is the minimum possible cost. (For an example see [4].)

To see what is possible in the more general setting we need the following definition.

Definition 4.1 For a sample path $\omega$, a limit point $\alpha^*$ of $\{\alpha_n(\omega)\}$, and a state $i \in S$, the pair $(i,\alpha^*)$ is said to be frequent if for each neighborhood $0$ of $\alpha^*$, the sequence of events $\{x_n = i, \alpha_n \in 0\}$ is frequent along $\omega$.

Lemma 4.1 There is a null set N such that for every $\omega \notin N$ and limit point $\alpha^*$ of $\{\alpha_n(\omega)\}$, $\alpha^*$ is a frequent limit point if and only if there exist $i \in S$ such that $(i,\alpha^*)$ is frequent along $\omega$

Proof Sufficiency follows from Def. 4.1 and necessity from Lemma 2.12. ¤

Theorem 4.2 Suppose $\xi(\alpha,i)$ is continuous in $\alpha$ for each i. Then there exists a null set N such that for every $\omega \notin N$, limit point $\alpha^*$ of $\{\alpha_n(\omega)\}$, and $i \in S$ such that $(i,\alpha^*)$ is frequent along $\omega$, the following relation holds: for all limit points $\tilde{\alpha}$ of $\{\alpha_n(\omega)\}$ and $j \in S$,

$$p(i,j;\xi(\alpha^*,i),\tilde{\alpha}) = p(i,j;\xi(\alpha^*,i),\alpha_0).$$

<u>Proof</u> From Def. 4.1 and Corollary 3.2, for $\varepsilon > 0$ there is a subsequence $\{n_k\}$ such that $\alpha_{n_k} \to \alpha^*$, $x_{n_k} = i$, and

$$\left|\frac{\ln p(x_{n_k},j;\xi(\alpha_{n_k},x_{n_k}),\tilde{\alpha})}{\ln p(x_{n_k};j;\xi(\alpha_{n_k},x_{n_k}),\alpha_0)}\right| = \left|\frac{\ln p(i,j;\xi(\alpha_{n_k},i),\tilde{\alpha})}{\ln p(i,j;\xi(\alpha_{n_k},i),\alpha_0)}\right| \leq \varepsilon.$$

The result follows from the continuity of $\xi(\alpha,i)$. ¤

This result is clearly weak in comparison with Theorem 4.1. To obtain stronger conclusions it is necessary to modify the adaptive control law through randomization. We study two such schemes.

<u>Randomization of control values</u>

We impose another assumption.

<u>A6.1</u> For any $\alpha \neq \beta$ in A, there exists $i \in S$ such that for every open set $O \subset Z(i)$ there exists $z \in Z(i)$ for which

$$[p(i,1;z,\alpha),p(i,2;z,\alpha),\ldots] \neq [p(i,1;z,\beta),p(i,2;z,\beta),\ldots] \quad (4.1)$$

It is worth comparing this assumption with Mandl's identifiability assumption (1.1). Whereas the latter requires that (4.1) holds for all $z \in Z(i)$, A6.1 requires that it hold only for a dense subset of $Z(i)$. Suppose $Z(i)$ is subset of $R^n$ as is usually the case. Then for $\alpha \neq \beta$ equality will hold in (4.1) for a set of $z \in R^n$ of dimension less than n and then A6.1 is likely to hold even when (1.1) does not.

Consider now the following random perturbation of the given adaptive law $\xi$. For each i let $\mu_i$ be a probability measure on $Z(i)$ which assigns positive values to every open set. Pick $\varepsilon_i > 0$ small, and for each $z \in Z(i)$ let $B(i,z)$ be the open ball of radius $\varepsilon_i$ and center z. Suppose at time n, $\alpha_n = \alpha$ is the MLE and $x_n = i$. Then the control $z_n$ is chosen from $B(\xi(\alpha,i))$ by an independent experiment corresponding to the restriction of $\mu_i$ to the set $B(\xi(\alpha,i))$. Let $\mathcal{G}_n = \sigma(x_0,z_0,\ldots,x_n,z_n)$ and $\mathcal{G}'_n = \sigma(x_0,z_0,\ldots,x_n,z_n,x_{n+1})$. Then

$$P(x_{n+1} = j|\mathcal{G}_n) = p(x_n,j;z_n,\alpha_0),\ P(z_{n+1} \in C|\mathcal{G}'_n) = \mu_{x_{n+1}}(C)$$
$$[\mu_{x_{n+1}}(B(\xi(\alpha_{n+1},x_{n+1})))]^{-1}$$

for every open set $C \subset B(\xi(\alpha_{n+1},x_{n+1}))$. The results obtained previously continue to hold if we use $\mathcal{G}_n$ in place of $\mathcal{F}_n$. The control law $\{z_n\}$ is called an <u>$\{\varepsilon_i\}$-randomization of $\xi$</u>.

<u>Theorem 4.3</u> Under any $\{\varepsilon_i\}$-randomization of $\xi$, $\alpha_n \to \alpha_0$ a.s.

<u>Proof</u> Let $\tilde{Z}(i)$ be a countable dense subset of $Z(i)$ and $\tilde{\mathcal{B}}_i$ the set of all open balls in $Z(i)$ with rational radii and center in $\tilde{Z}(i)$. By the Martingale Stability Theorem there exists a null set $N_1$ outside which

$$\lim \frac{1}{n} \sum^{n-1} [I(z_{m+1} \in B) - E\{I(z_{m+1} \in B)|\mathcal{G}'_m\}] = 0 \qquad (4.2)$$

for every $i$ and $\tilde{B} \in \tilde{\mathcal{B}}(i)$. By Lemmas 2.8, 2.11 there is a null set $N_2$ outside which every state is positive recurrent. Finally let $N_3$ be the null set in Corollary 3.2.

Let $\omega \notin N = N_1 \cup N_2 \cup N_3$ and let $\alpha^* \neq \alpha_0$ be a limit point of $\{\alpha_n(\omega)\}$. By A6.1 there is $i \in S$ such that for every open set $0 \subset Z(i)$ there is $z \in 0$ and $j \in S$ for which $p(i,j;z,\alpha^*) \neq p(i,j;z,\alpha_0)$. Let $\{n_k, k \geq 0\}$ be the maximal subsequence for which $x_{n_k}(\omega) = i$ for every $k$. By Lemma 2.8 the sequence $\{n = n_k \text{ for some } k\}$ is frequent and so, by Lemma 2.3, there is a subsequence $\{\tilde{n}_k\}$ of $\{n_k\}$ and $z^* \in Z(i)$ such that along $\omega$

$$\xi(x_{\tilde{n}_k},\alpha_{\tilde{n}_k}) = \xi(i,\alpha_{\tilde{n}_k}) \to z^*, \text{ and } \overline{\lim}\, \frac{1}{n} \sum_{m=0}^{n-1} I(m = \tilde{n}_k \text{ for some } k) > 0.$$

We can find $M < \infty$ and an open set $0 \subset Z(i)$ such that $z^* \in 0 \subset B(\xi(i,\alpha_{\tilde{n}_k}))$, $k > M$. By A6.1 there are $\tilde{z} \in 0$, $j \in S$ for which $p(i,j;\tilde{z},\alpha^*) \neq p(i,j;\tilde{z},\alpha_0)$. By A1 there exists $\tilde{B} \in \tilde{\mathcal{B}}_i$ such that $\tilde{z} \in \tilde{B} \subset 0$ and $|\ln \frac{p(i,j;z,\alpha^*)}{p(i,j;z,\alpha_0)}| > \delta > 0$ for $z \in \tilde{B}$. Now by (4.2)

$$\overline{\lim}\, \frac{1}{n} \sum_{m=0}^{n-1} I(z_m \in \tilde{B}) = \overline{\lim}\, \frac{1}{n} \Sigma\, E\{I(z_m \in \tilde{B}) | \mathcal{G}'_{m-1}\}$$

$$\geq \overline{\lim}\, \frac{1}{n} \Sigma\, \mu_i(\tilde{B})[\mu_i(B(\xi(i,\alpha_m)))]^{-1} I(x_m = i \text{ and } \tilde{B} \subset B(\xi(i,\alpha_m)))$$

$$\geq \mu_i(\tilde{B})\, \overline{\lim}\, \frac{1}{n} \sum_{m=0}^{n-1} I(m = \tilde{n}_k \text{ for some } k) > 0$$

Hence the sequence $\left\{ \left| \dfrac{\ln p(x_n,j;z_n,\alpha^*)}{\ln p(x_n,j;z_n,\alpha_0)} \right| > \delta \right\}$ is frequent along $\omega$, contradicting Corollary 3.2. ¤

Let $c(i,j,z)$ be a nonnegative bounded cost function which is continuous in $z \in Z(i)$ for each $i,j$. For every stationary $\zeta \in Z$ and $\alpha \in A$, let

$$V(\zeta,\alpha) = \sum_i \pi_\alpha(i,\zeta) \sum_j p(i,j;\zeta(i),\alpha) c(i,j,\zeta(i)),$$

where $\{\pi_\alpha(i,\zeta)\}$ are the stationary probabilites corresponding to the transition probabilities $\{p(i,j;\zeta(i),\alpha)\}$. Thus $V(\zeta,\alpha)$ is the cost incurred by the control law $z_n = \zeta(x_n)$ if $\alpha$ is the true parameter.

Lemma 4.2 $V(\zeta,\alpha)$ is continuous in $\zeta$ for each $\alpha$.

Proof As in the proof of Lemma 2.9 we see that $\pi_\alpha(i,\zeta)$ is continuous in $\zeta$ for each $i,\alpha$. Given $\delta > 0$, it follows from Lemma 2.9 and the boundedness of $c$ that there exists $J$ such that

$$V(\zeta,\alpha) \geq \sum_{i=1}^{J} \pi_\alpha(i,\zeta) \sum_{j=1}^{\infty} p(i,j;\zeta(i),\alpha) c(i,j,\zeta(i)) \geq V(\zeta,\alpha) - \delta$$

Now $\sum_{j=1}^{n} p(i,j;\zeta(i),\alpha)$ is continuous in $\zeta$ and converges monotonically, hence uniformly, to 1 as $n \to \infty$ and so, increasing $J$ if necessary, we get

$$V(\zeta,\alpha) \geq \sum_{i=1}^{J} \pi_\alpha(i,\zeta) \sum_{j=1}^{J} p(i,j;\zeta(i),\alpha)c(i,j,\zeta(i)) \geq V(\zeta,\alpha) - 2\delta$$

The term in the middle is continuous in $\zeta$ and since $\delta > 0$ is arbitrary it follows that $V(\zeta,\alpha)$ is continuous as well. ¤

Suppose now that the given adaptive law $\xi$ is such that for each $\alpha$,

$$V(\xi(\alpha),\alpha) = V(\alpha) = \min_{\zeta \in Z} V(\alpha,\zeta).$$

We wish to show that if $\{z_n\}$ is an $\{\varepsilon_i\}$-randomization of $\xi$ then its cost can be made arbitrarily close to $V(\alpha_0)$ by choosing $\varepsilon_i > 0$ sufficiently small.

Theorem 4.4 Suppose $V(\xi(\alpha_0),\alpha_0) < V(\zeta,\alpha_0)$ when $\zeta \neq \xi(\alpha_0)$ i.e. $\xi(\alpha_0)$ is the unique optimal stationary control law. For any $\delta > 0$, there exists $\varepsilon > 0$ such that if $\{z_n\}$ is an $\{\varepsilon\}$-randomization of $\xi$, then

$$V(\alpha_0) \leq \overline{\lim} \frac{1}{n} \sum_{m=0}^{n-1} c(x_m,x_{m+1},z_m) \leq V(\alpha_0) + \delta \text{ a.s.}$$

Proof By Lemma 4.2 there exists an open set $O$ in $Z$ with $\xi(\alpha_0) \in O$ such that $V(\alpha_0) \leq V(\alpha_0,\zeta) < V(\alpha_0) + \delta$ for $\zeta \in O$. Since $Z = \prod_i Z(i)$ has the product topology we may suppose that $O$ has the form

$$O = \prod_{i=1}^{m} B(i) \times \prod_{i=m+1}^{\infty} Z(i)$$

for some $m < \infty$ and where $B(i)$ is a ball of radius $\varepsilon$ and center $\xi(\alpha_0,i)$. Let $\{z_n\}$ be an $\{\varepsilon\}$-randomization of $\xi$. By Theorem 4.3, $\alpha_n \to \alpha_0$ a.s. and so, from Lemma 4.2 and the uniqueness of $\xi(\alpha_0)$ it follows that $\xi(\alpha_n) \to \xi(\alpha_0)$ a.s. Hence there exists a random time $N < \infty$ a.s. such that

$$\xi(\alpha_n) \in O, \; n \geq N \text{ a.s.}$$

By Lemma 2.10 it follows that

$$\overline{\lim}\ \frac{1}{n}\sum^{n-1} c(x_m,x_{m+1},z_m) \le \sup_{\zeta\in 0} V(\alpha_0,\zeta) \le V(\alpha_0) + \delta. \qquad ¤$$

Randomization of parameter estimates

We consider an alternative perturbation of the given adaptive law $\xi$. For the finite case a similar randomization is proposed in [10]. We replace A6.1 by the following.

A6.2 For every $\alpha \ne \beta$ in A and i in S and every neighborhood 0 of $\alpha$ there is an open set $\tilde{0} \subset 0$ such that for every $\tilde{\alpha} \in \tilde{0}$,

$$[p(i,1;\xi(\tilde{\alpha},1),\alpha),p(i,2;\xi(\tilde{\alpha},2),\alpha),\ldots] \ne [p(i,1;\xi(\tilde{\alpha},1),\beta),p(i,2;\xi(\tilde{\alpha},2),\beta),\ldots]$$

Let $\nu$ be a probability measure on A which assigns positive values to every open set. Pick $\gamma > 0$ small and let $B(\alpha)$ denote the ball of radius $\gamma$ and center $\alpha$. Let $\alpha_n$ be the MLE and $x_n$ the state at time n. Then the control $z_n$ is chosen to be $z_n = \xi(\tilde{\alpha}_n,x_n)$ where $\tilde{\alpha}_n$ is selected from $B(\alpha_n)$ by an independent experiment corresponding to the restriction of $\nu$ to $B(\alpha_n)$. We call the control law $\{z_n\}$ a $\gamma$-randomization of $\xi$.

Theorem 4.5 Under $\{z_n\}$, $\alpha_0$ is the only frequent limit point of $\{\alpha_n\}$ almost surely.

Proof Let N be the null set in Lemma 2.12 and Corollary 3.2, and let $\omega \notin N$. Suppose $\alpha^* \ne \alpha_0$ is a frequent limit point of $\{\alpha_n(\omega)\}$. Then the sequence $\{\alpha_n(\omega) \in B(\alpha^*)\}$ is frequent and let $\{n_k, k \ge 0\}$ be the maximal subsequence such that $\alpha_{n_k} \in B(\alpha^*)$ for all k. By Lemma 2.12 there is $i \in S$ and a subsequence $\{\tilde{n}_k\}$ of $\{n_k\}$ such that along $\omega$

$$x_{\tilde{n}_k} = i, \text{ and } \overline{\lim}\ \frac{1}{n}\sum_{m=0}^{n-1} I(m = \tilde{n}_k \text{ for some } k) > 0$$

Proceeding from here on as in the proof of Theorem 4.2 we can show that for some $\delta > 0$ and $j \in S$, the sequence $\left\{\left|\frac{\ln p(x_n,j;z_n,\alpha^*)}{\ln p(x_n,j;z_n,\alpha_0}\right| > \delta\right\}$ is frequent along $\omega$, contradicting Corollary 3.2. ¤

Suppose now that $V(\xi(\alpha),\alpha) = V(\alpha)$ as in Theorem 4.4.

<u>Theorem 4.6</u> Suppose $\xi(\alpha_0)$ is the unique optimal stationary control law under $\alpha_0$. For any $\delta > 0$ there exists $\gamma > 0$ such that if $\{z_n\}$ is a $\gamma$-randomization of $\xi$, then

$$V(\alpha_0) \leq \overline{\lim}\, \frac{1}{n} \sum_{m=0}^{n-1} c(x_m, x_{m+1}, z_m) \leq V(\alpha_0) + \delta$$

<u>Proof</u> The proof is virtually identical to that of Theorem 4.4 with Theorem 4.5 taking the place of Theorem 4.3, and with the additional feature that the contribution to the average cost due to rare limit points of $\{\alpha_n(\omega)\}$ vanish asympotically due to Lemma 2.2 (iv) and the converse in Lemma 2.1. ¤

We close this section with the remark that A6.1 may be replaced by A6.3 which is more similar to A6.2.

<u>A6.3</u> For every $\alpha \neq \beta$ in A and i in S and every neighborhood 0 of $\xi(\alpha,i)$ there exists $z \in Z(i)$ for which (4.1) holds.

A heuristic discussion of A6.1-A6.3 is deferred till Section 6.

## 5. INADEQUATE PARAMETER SETS

So far most of the results were derived under the assumption $\alpha_0 \in A$. In Section 3, a crucial role is played by this assumption in the proof of Lemma 3.2 where we use the fact that $\wedge_n(\alpha_n) \geq \wedge_n(\alpha_0) = 1 > 0$. This observation motivates the following definition.

<u>Definition 5.1</u> The parameter set A is <u>adequate</u> if $P\{\wedge(\alpha) = 0$ for all $\alpha\} = 0$, or equivalently, if $P\{\sup_\alpha \wedge(\alpha) > 0\} = 1$; otherwise A is <u>inadequate</u>.

Evidently if $\alpha_0 \in A$ then A is adequate. The following results are proved along the same lines as in Section 3.

<u>Theorem 5.1</u> If A is adequate then Theorem 3.1 and Corollary 3.2 continus to hold.

Corollary 5.1 If A is inadequate then the conclusions of Theorem 3.1 and Corollary 3.2 hold outside the set $N = \{\omega | \Lambda(\alpha,\omega) = 0 \text{ for all } \alpha\}$. (Note that $P(N) > 0$.)

From now on we consider the case when A is inadequate. Suppose initially that there are no control parameters so that $\alpha$ merely indexes a stationary transition probability $p(i,j;\alpha)$. Suppose further that under each $\alpha$ all states are positive recurrent. Then it is easy to establish the following result which states that the MLE $\alpha_n$ converges to a subset of A consisting of parameter values which are "closest" to $\alpha_0$ in a well-defined sense.

Theorem 5.2 $\alpha_n$ converges almost surely to the subset of parameter values which maximize

$$D(\alpha) = \sum_i \pi_{\alpha_0}(i) \sum_j p(i,j;\alpha_0) \ell n \frac{p(i,j;\alpha)}{p(i,j;\alpha_0)}$$

where $\{\pi_{\alpha_0}(i)\}$ are the invariant probabilities corresponding to $\{p(i,j;\alpha_0)\}$.

The case when the transition probabilities do depend on a control parameter, discussed next, is considerably more complicated. For simplicity we assume that $Z(i) = \{z_1^i, z_2^i, \ldots, z_L^i\}$ contains L elements. The general case can be worked out in a similar way though the details are cumbersome.

Let A1-A5 hold and let N be the set on which condition T fails. N is null by Lemma 2.11. Let $\omega \notin N$ and let

$$q_n(i,j,\ell) = \frac{1}{n} \sum_{m=0}^{n-1} I(x_m = i, x_{m+1} = j, z_m = z_\ell^i).$$

Then $0 \leq q_n \leq 1$ and $\sum\limits_{i,j,\ell} q_n(i,j,\ell) = 1$. Thus $q_n$ defines a (random) probability measures on triples $(i,j,\ell)$. Let $\rho_n = \{q_n(i,j,\ell)\}$ denote the vector with components $q_n(i,j,\ell)$. We think of $\rho_n$ as an element of the normed space $\ell_\infty$.

Lemma 5.1 Suppose $\omega \notin N$. Then the $\ell_\infty$ closure of the sequence $\{\rho_n(\omega)\}$ is compact, and each of its limit points is itself a probability. The set of frequent limit points of $\{\rho_n(\omega)\}$ is compact and for any open neighborhood $O$ of this set the sequence $\{\rho_n(\omega) \notin O\}$ is rare.

Proof Considered as a subset of $[0,1]^\infty$ with product topology the set $\{\rho_n\}$ has compact closure. Let $\rho^* = \{q^*(i,j,\ell)\}$ be such that

$$\lim_k q_{n_k}(i,j,\ell) = q^*(i,j,\ell) \text{ for each } i,j,\ell.$$

Clearly $q^*(i,j,\ell) \geq 0$. By Lemma 2.11 for $\varepsilon > 0$ there exists $J_\varepsilon$ and $K_\varepsilon$ such that, along $\omega$,

$$\frac{1}{n_k} \sum_{m=0}^{n_k} I(x_m > J_\varepsilon) < \varepsilon, \; k > K_\varepsilon$$

and so, for $k > K_\varepsilon$,

$$1 \geq \sum_{ij=1}^{J_\varepsilon} \sum_{\ell=1}^{L} q_{n_k}(i,j,\ell) \geq 1-\varepsilon \tag{5.1}$$

From this it follows that

$$1 \geq \sum_{ij=1}^{J_\varepsilon} \sum_{\ell=1}^{L} q^*(i,j,\ell) \geq 1-\varepsilon \tag{5.2}$$

(5.1), (5.2) imply that $\rho_{n_k}$ converges to $\rho^*$ in $\ell_\infty$ and that $\rho^*$ is a probability. The remaining assertions now follow from Lemma 2.2. ¤

Corollary 5.2 If $c(i,j,\ell)$ is any bounded function then

$$\lim_k \sum_{ij\ell} c(i,j,\ell) q_{n_k}(i,j,\ell) = \sum_{ij\ell} c(i,j,\ell) q^*(i,j,\ell).$$

As before let $\{\pi(i,\zeta)\}$ denote the invariant probabilities under the stationary law $\zeta \in Z$. Let $q_\zeta(i,j,\ell) = \pi(i,\zeta)p(i,j;\zeta(i),\alpha_0)I(\zeta(i) = z_\ell^i)$. Let $q_\zeta = \{q_\zeta(i,j,\ell)\}$. Consider $G = \{q_\zeta | \zeta \in Z\}$ as a subset of $\ell_\infty$.

Lemma 5.2 There exists a null set $N$ such that if $\omega \notin N$ then every limit point of $\{\rho_n\}$ belongs to the convex hull of $G$.

*Proof* Let $c_r(i,j,\ell)$, $r = 1,2,\ldots$ be a countable dense set of functions in $\ell_\infty$ and N a null set such that for $\omega \notin N$, and every r

$$\overline{\lim_n} \frac{1}{n} \sum_{m=0}^{n-1} c_r(x_m,x_{m+1},z_m) \le \max_{\zeta\in Z} \sum_i \pi(i,\zeta) \sum_j p(i,j;\zeta(i),\alpha_0)c_r(i,j,\zeta(i)). \tag{5.3}$$

Such a null set N exists by Lemma 2.10. Augment N if necessary so that Lemma 5.1 applies. Let $\omega \notin N$ and suppose $\rho_{n_k}$ converges to $\rho^*$ along $\omega$. Then by Corollary 5.2, for every r

$$\sum_{ij\ell} c_r(i,j,\ell)q^*(i,j,\ell) = \lim_k \sum_{ij\ell} c_r(i,j,\ell)q_{n_k}(i,j,\ell)$$

$$= \lim_k \frac{1}{n_k} \sum_{k=0}^{n_k-1} c_r(x_m,x_{m+1},z_m) \le \max_\zeta \sum_i \pi(i,\zeta) \sum_j p(i,j,\zeta(i),\alpha_0)$$

$$\cdot\, c_r(i,j,\zeta(i)) = \max_{g\in G} \sum_{ij\ell} c_r(i,j,\ell)g(i,j,\ell) \tag{5.4}$$

Now G is a compact subset of $\ell_\infty$ and hence so is its convex hull. If $\{q^*(i,j,\ell)\}$ does not lie in the convex hull of G, by the separation theorem there exists $c_r$ such that for all q

$$\sum_{ij\ell} c_r(i,j,\ell)[q^*(i,j,\ell) - q(i,j,\ell)] > 0$$

contradicting (5.3). ⌑

*Lemma 5.3* Let N be as in Lemma 5.1 and $\omega \notin N$. Suppose $\rho_{n_k}$ converges to $\rho^*$ along $\omega$. Let $\rho_{\tilde{n}_k}$, $k \ge 0$ be another subsequence such that $|n_k-\tilde{n}_k| \le M < \infty$ for all k. Then $\rho_{\tilde{n}_k}$ also converges to $\rho^*$ along $\omega$.

*Proof* For any $i,j,\ell$, if $n_k \ge \tilde{n}_k$,

$$|q_{n_k}(i,j,\ell) - q_{n_k}(i,j,\ell)| \le \frac{1}{\tilde{n}_k} I(x_m = i, x_{m+1} = j, z_m = z^i_\ell) \le \frac{M}{\tilde{n}_k}\,.$$

Thus $|q_{n_k}(i,j,\ell) - q_{\tilde{n}_k}(i,j,\ell)| \le M(\frac{1}{n_k} + \frac{1}{\tilde{n}_k}) \to 0$ as $k \to \infty$. ⌑

To describe the asymptotic behavior of $\{\rho_n\}$ we need another concept. Let $\{a_n, n \ge 0\}$ be a sequence in a metric space. Let $\tilde{O}$, $O$ be open sets with $\tilde{O} \subset \bar{\tilde{O}} \subset O$. Let

$$m_k(\tilde{O}) = \min\{n > m_{k-1}(\tilde{O}) \mid a_{n-1} \notin \tilde{O}, a_n \in \tilde{O}\}$$

be the kth time $a_n$ enters $\tilde{O}$ after leaving it. Let

$$n_k(O) = \min\{n > m_k(\tilde{O}) \mid a_n \notin O\},$$

$$\ell_k(O) = \max\{n \leq m_k(\tilde{O}) - 1 \mid a_n \notin O\}$$

We say that $\{a_n\}$ <u>drifts slowly</u> if for any open sets $O$, $\tilde{O}$ with $\tilde{O} \subset \bar{\tilde{O}} \subset O$, the sequences $\{n = n_k(O)$ for some $k\}$ and $\{n = \ell_k(O)$ for some $k\}$ are both rare.

<u>Theorem 5.3</u> There exists a null set $N$ such that if $\omega \notin N$ then $\{\rho_n(\omega)\}$ drifts slowly considered as a sequence in $\ell_\infty$.

<u>Proof</u> Let $N$ be as in Lemma 5.1 and $\omega \notin N$. Let $O$, $\tilde{O}$ be open sets in $\ell_\infty$ with $\tilde{O} \subset \bar{\tilde{O}} \subset O$, and define $\ell_k$, $m_k$, $n_k$ as previously with $a_n = \rho_n(\omega)$. We shall only show that the sequence $\{n = \ell_k$ for some $k\}$ is rare, the proof of the other half of the assertion being similar. We may suppose $\rho_{m_k} \in \tilde{O}$ for infinitely many $k$ because otherwise there is nothing to prove. We claim first that

$$m_k - \ell_k \to \infty \text{ as } k \to \infty. \tag{5.5}$$

For, suppose in contradiction that there is a subsequence $k_i$, with

$$m_{k_i} - \ell_{k_i} \leq M < \infty \text{ for all } i.$$

Now, by Lemma 5.1, there exists a subsequence, denoted again by $\{k_i\}$ such that

$$\lim_i \rho_{m_{k_i}} = \rho^* \text{ and } \lim_i \rho_{\ell_{k_i}} = \tilde{\rho}.$$

Clearly $\rho^* \in \bar{\tilde{O}}$ and $\tilde{\rho} \notin O \supset \bar{\tilde{O}}$ so that $\rho^* \neq \tilde{\rho}$ contradicting Lemma 5.3. Thus (5.4) must hold.

Now $\ell_{k+1} > m_k > \ell_k$ and so (5.5) implies $\lim \frac{k}{\ell_k} = 0$. Let $k(m) = \max\{k \mid \ell_k \leq m\}$. Then

$$\frac{1}{n}\sum_{i=1}^{m} I(i = \ell_k \text{ for some } k) \leq \frac{1}{k(m)}\sum_{i=1}^{k(m)} I(i = \ell_k \text{ for some } k)$$

$$= \frac{k(m)}{\ell_{k(m)}} \to 0 \text{ as } m \to \infty. \quad \square$$

In summary the results of this section show that when $\alpha_0 \notin A$ (more precisely, when A is inadequate) and for any control law, the relative frequencies, $\rho_n$, of the various state and control combinations, converges to a tight set of probabilities which is the convex hull of the set G of invariant probabilities under all stationary control laws. The sequence $\rho_n$ may, however, drift slowly.

In [4] we have given an example which shows that $\rho_n$ may not converge almost surely. In that example, $z_n = \xi(\alpha_n, x_n)$ is an adaptive law constructed in such a way that, as $\rho_n$ begins to converge to some $\rho^*$ and hence $\alpha_n$ to some $\alpha^*$, the corresponding control values $z_n = \xi(\alpha^*, x_n)$ are such that the likelihood ratio is maximized at some other parameter value $\tilde{\alpha} \neq \alpha^*$ and so $\alpha_n$ begins to drift slowly to $\tilde{\alpha}$. But at $\tilde{\alpha}$, the control values $\xi(\tilde{\alpha}, x_n)$ are such that the likelihood ratio is maximized at $\alpha^*$. Thus the MLE $\alpha_n$ keeps switching more and more slowly between $\alpha^*$ and $\tilde{\alpha}$.

The following conjecture seems plausible

<u>Conjecture</u> - For $\alpha, \beta$ in A define

$$M(\alpha,\beta) = \sum_i \Pi(i,\zeta(\zeta)) \sum_j p(i,j;\zeta(\alpha,i),\alpha_0) \ln \frac{p(i,j;\zeta(\alpha,i),\beta)}{p(i,j;\zeta(\alpha,i),\alpha_0)}$$

Then a sufficient condition for the MLE $\alpha_n$ to converge a.s. to some A-valued random variable is

$$M(\alpha,\alpha) > M(\alpha,\beta) \tag{5.6}$$

for all $\alpha, \beta \in A$ such that $\alpha \neq \beta$.

If true, the above conjecture suggests the following choice of $\zeta(\alpha)$ for each $\alpha$. Choose $\zeta(\alpha)$, from among the strategies that are near-optimal under $\alpha$, to make (5.6) hold for as many $\beta \neq \alpha$ as possible. This is a manifestation of the trade-off between identification and optimality considerations in the choice of inputs. Thus the control in the adaptive scheme fulfills a dual purpose — to ensure good convergence properties for the estimates (with, of course, convergence to the true parameter value if possible) and to satisfy the optimality criteria as closely as possible. Clearly, this discussion is only heuristic and much needs to be done.

6. DISCUSSION

The approach adopted here puts greater emphasis on "time domain" or sample path behavior and many of the concepts introduced can be seen as analogs of certain ensemble concepts, viz., rare events are analogs of null sets, condition T is an analog of tightness etc. The reduced dependence on ensemble averages makes this approach more suitable for non-stationary processes which are asymptotically well-behaved.

Many of the concepts introduced in Section 2, such as recurrence and positive recurrence, can be extended to more general spaces such as an arbitrary Borel state space, and it seems reasonable to expect that similar results will hold.

Assumptions A6.1, A6.2 and A6.3 have the common objective of overcoming the limitations of Theorem 4.2 which says that at the limiting values of the parameter estimates, the frequent limits of control values are such that these control values cannot distinguish between different limits of the parameter estimates. This cannot occur if for each $\alpha \neq \alpha_0$ we use frequently a control to distinguish between $\alpha$ and $\alpha_0$. This can be achieved, as in A6.1 and A6.3, by a small randomization if for each $\alpha$ the set of control values which cannot distinguish between $\alpha$ and $\alpha_0$ is "thin" i.e. has empty interior. A6.2 permits an analogous randomization in parameter space. The latter appears more appealing for practical problems even though the result is slightly weaker.

The case when $\alpha_0 \notin A$ is practically important since models used for identification and control are approximations of the true system. The results presented here are very incomplete and much needs to be done.

REFERENCES

[1] K. Åström and B. Wittenmark, "On self-tuning regulators," Automatica 9, 185-199, 1973.

[2] P. Mandl, "Estimation and control in Markov chains," Adv. Appl. Prob. 6, 40-60, 1974.

[3] L. Ljung and B. Wittenmark, "Asymptotic properties of self-tuning regulators," Dept. Auto. Contr., Lund Inst. Technology, Lund, Sweden, Rep. 7407, Feb. 1974.

[4] V. Borkar and P. Varaiya, "Adaptive control of Markov chains, I: Finite parameter set," to appear in IEEE Trans. Auto. Contr., Dec. 1979.

[5] L. Breiman, Probability, Reading, Mass.: Addison-Wesley, 1968.

[6] M. Loeve, Probability Theory, Princeton, N. J.: Van Nostrand, second edition, 1960.

[7] S. Ross, Applied Probability Models with Optimization Applications, San Francisco, Ca.: Holden-Day, 1970.

[8] A. Federgruen, A. Hordijk and H. C. Tijms, "A note on simultaenous recurrence conditions on a set of denumerable stochastic matrices," Dept. Oprn. Res., Mathematisch Centrum, Amsterdam, Rept. BW 85/77, Nov. 1977 (to appear in J. Appl. Prob.).

[9] J. Neveu, Discrete-Parameter Martingales, Amsterdam: North-Holland, 1975.

[10] B. Doshi and S. E. Shreve, "Randomized self-tuning control of Markov chains," preprint, Dept. of Math. Sc., Univ. of Delaware, Newark, Del., April 1979 (to appear in J. Appl Prob.).