SRC TR 87-168

Parameter Estimation Under Threshold Policies for a Simple Flow Control Problem

by

D. Ma and A. M. Makowski

Invited paper to the 26th IEEE Conference on Decision and Control,
Los Angeles, California, December 1987

PARAMETER ESTIMATION UNDER THRESHOLD POLICIES FOR A SIMPLE FLOW CONTROL PROBLEM

by

Dye-Jyun Ma[1] and Armand M. Makowski[2]

Electrical Engineering Department and Systems Research Center
University of Maryland, College Park, Maryland 20742

ABSTRACT

In [5], the authors showed that threshold policies solve an optimal flow control problem for discrete-time $M|M|1$ queues, where the decision-maker seeks to maximize the system throughput subject to a bound on the long-run average queue size. In this paper, attention focuses on a non-Bayesian adaptive version of this problem when the arrival and service rates are assumed to be unknown constants. By invoking the Certainty Equivalence Principle, adaptive threshold policies are generated by substituting maximum likelihood estimates for the rate parameters in the definition of the optimal threshold policies. Under such policies, the maximum likelihood estimates are shown to be strongly consistent through an indirect method of analysis that combines ideas from stochastic ordering, a study of the rates of convergence via the theory of Large Deviations and absolutely continuous changes of measures. The optimality of the adaptive threshold policies follows as a byproduct of this consistency result.

[1] The work of this author was supported partially through NSF Grant ECS-83-51836 and partially through NSF Grant NSFD CDR-85-00108.

[2] The work of this author was supported partially through ONR Grant N00014-84-K-0614 and partially through a grant from AT&T Bell Laboratories.

1. Introduction

In a companion paper [5], the authors recently considered a flow control model for discrete $M|M|1$ queues with infinite buffer capacity. The selection of flow control strategies with desirable performance properties was addressed by formulating an optimal flow control problem where the decision-maker seeks to maximize the throughput with a bound on the long-run average number of customers in the system. The corresponding constrained optimal flow control policy was identified to be of *threshold* type with a simple structure completely determined by a critical acceptance level L (in $\mathbb{N}$) and an acceptance probability η ($0 \leq \eta \leq 1$). Under a threshold policy (L, η), at the beginning of each time slot, a new customer is accepted (resp. rejected) if the buffer content is strictly below L (resp. strictly above L), while if there are *exactly* L customers in the buffer, this new customer is accepted (resp. rejected) with probability η (resp. $1 - \eta$).

The constrained optimal threshold policy (L, η) is determined solely by its acceptance parameters, which in turn are functions of the service and arrival rate parameters. However, in many applications, these model parameters are not available to the decision-maker and the optimal threshold policy cannot be implemented in its given form. This model uncertainty leads naturally to the formulation of *adaptive* versions of the optimal flow control problem [6,7].

In this paper, a *non*-Bayesian version is discussed in the event arrival and service rates are assumed to be *unknown constants.* By invoking the Certainty Equivalence design principle, an adaptive threshold policy is proposed as a possible implementation of the optimal threshold policy. More specifically, at the beginning of any time slot, a joint *maximum likelihood* estimate of the arrival and service rates is computed on the basis of the available information which includes past and present history of the arrival, service completion and control processes. The flow control action to be implemented for that slot is then generated in accordance with the optimal threshold policy where the estimate is substituted for the true parameter value.

The situation discussed here cannot be handled by earlier work [1,8] on the adaptive control of Markov chains, and it is the purpose of this paper to outline a brief discussion of the estimation and control properties of this adaptive threshold policy. For details and proofs, the interested reader is refered to [7] where several information patterns (in addition to the one defined earlier) are considered. It is noteworthy that although explicit expressions are

easily derived for the maximum likelihood estimates under the adopted information pattern, there does not seem to be simple arguments to establish the strong consistency of the these estimates under the adaptive threshold policy. This can be traced back to the tight interaction that arises between estimation and control in the system under consideration here.

An indirect method of analysis is proposed for establishing strong consistency of the estimates. It combines ideas of *stochastic ordering*, a study of rates of convergence via the theory of *Large Deviations* and *absolutely continuous change of measures*. Roughly speaking, parameter identification is first established for a class of auxiliary policies. The probability measure induced by the adaptive threshold policy is then shown to be absolutely continuous to the one induced by one of the auxiliary policies, thus yielding the result. The proposed approach is believed to be of independent interest, and could prove useful in studying other problems of parameter estimation and adaptive control for Markov chains.

2. The model

The model of interest here is essentially the one introduced by the authors in the companion paper [5] to which the reader is refered for additional information concerning notation and terminology.

As indicated there, take the sample space to be $\Omega := I\!N \times \left(\{0,1\}^3\right)^\infty$ and define the information spaces $\{I\!H_n\}_0^\infty$ by $I\!H_{n+1} := I\!H_n \times \{0,1\}^3$ for all $n = 0, 1, \ldots$, with $I\!H_0 := I\!N$. An element ω of Ω is viewed as a sequence $(x, \omega_0, \omega_1, \ldots)$ with x in $I\!N$ and ω_n in $\{0,1\}^3$ for all $n = 0, 1, \ldots$. Each block component ω_n is written in the form (u_n, a_n, b_n), with u_n, a_n and b_n being all elements in $\{0,1\}$. An element h_n in $I\!H_n$ is uniquely associated with the sample ω by $h_n := (x, \omega_0, \ldots, \omega_{n-1})$ with $h_0 := x$. The random variables (RV) Ξ, $\{U(n)\}_0^\infty$, $\{A(n)\}_0^\infty$, $\{B(n)\}_0^\infty$ and $\{H(n)\}_0^\infty$ are then defined on the sample space Ω by posing $\Xi(\omega) := x$, $U(n,\omega) := u_n$, $A(n,\omega) := a_n$, $B(n,\omega) := b_n$ and $H(n,\omega) := h_n$ for all $n = 0, 1, \ldots$ and for every ω in Ω. These RV's take values in $I\!N$, $\{0,1\}$, $\{0,1\}$, $\{0,1\}$ and $I\!H_n$, respectively, whereas the queue sizes $\{X(n)\}_0^\infty$ are $I\!N$-valued RV's recursively defined by

$$X(n+1) = X(n) + U(n)A(n) - 1[X(n) \neq 0]B(n) \qquad n = 0, 1, \ldots \quad (2.1)$$

with $X(0) := \Xi$.

For each $n = 0, 1, \ldots$, let $\mathbb{F}_n$ be the σ-field generated by the RV $H(n)$ on the sample space Ω. Clearly, $\mathbb{F}_n \subset \mathbb{F}_{n+1}$, and with standard notation, $\mathbb{F} := \vee_{n=0}^{\infty} \mathbb{F}_n$ is simply the σ-field on Ω generated by the RV's Ξ and $\{U(n), A(n), B(n)\}_0^{\infty}$.

Since randomization is allowed, an admissible policy π is defined as any collection $\{\pi_n\}_0^{\infty}$ of mappings $\pi_n: \mathbb{H}_n \to [0, 1]$, with the interpretation that the potential arrival during the slot $[n, n+1)$ is admitted (resp. rejected) with probability $\pi_n(h_n)$ (resp. $1 - \pi_n(h_n)$) whenever the information h_n is available to the decision-maker. The collection of all such admissible policies is denoted by $\mathcal{P}$. With the terminology developed in [5], a stationary policy g is said to be of *threshold* type if there exists a pair (L, η), with L an integer in $\mathbb{N}$ and η in $[0, 1]$, such that

$$g(x) = \begin{cases} 1 & \text{if } 0 \leq x < L; \\ \eta & \text{if } x = L; \\ 0 & \text{if } x > L. \end{cases} \tag{2.2}$$

Such a threshold policy is denoted by (L, η), and by extension, the Markov stationary policy that admits every single customer, i.e., $g(x) = 1$ for all x in $\mathbb{N}$, is denoted by $(\infty, 1)$

Let $q(\bullet)$ be a probability distribution on $\mathbb{N}$, and let $\theta = (\lambda, \mu)$ be an element in $[0, 1]^2$. Given any policy π in $\mathcal{P}$, there exists an unique probability measure $P^{\theta,\pi}$ on $\mathbb{F}$, with corresponding expectation operator $E^{\theta,\pi}$, satisfying the requirements (R1)-(R3) below, where

(R1): For all x in $\mathbb{N}$,

$$P^{\theta,\pi}[\Xi = x] = q(x),$$

(R2): For all a and b in $\{0, 1\}$,

$$\begin{aligned} &P^{\theta,\pi}[A(n) = a, B(n) = b | \mathbb{F}_n \vee \sigma\{U(n)\}] \\ &= (a\lambda + (1-a)(1-\lambda))(b\mu + (1-b)(1-\mu)) \end{aligned} \qquad n = 0, 1, \ldots$$

and

(R3):

$$P^{\theta,\pi}[U(n) = 1 | \mathbb{F}_n] = \pi_n(H(n)). \qquad n = 0, 1, \ldots$$

Throughout this paper, it is assumed that for every π in $\mathcal{P}$,

$$E^{\theta,\pi}[\Xi] = \sum_{x=0}^{\infty} x q(x) < \infty. \tag{2.3}$$

3. A constrained optimal control problem

Let θ be a fixed element in $[0,1]^2$ held fixed throughout this section. For any admissible policy π in $\mathcal{P}$, pose

$$T(\pi,\theta) := \liminf_n \frac{1}{n+1} E^{\theta,\pi} \sum_{t=0}^{n} \mu 1[X(t) \neq 0] \tag{3.1}$$

and

$$N(\pi,\theta) := \limsup_n \frac{1}{n+1} E^{\theta,\pi} \sum_{t=0}^{n} X(t). \tag{3.2}$$

The quantities $T(\pi,\theta)$ and $N(\pi,\theta)$ have the interpretation of *throughput* and long-run average *queue size*, respectively, when the policy π is used.

Given $V > 0$, consider the following *constrained* optimization problem $(P_{V,\theta})$, where

$$(P_{V,\theta}): \quad \text{Maximize } T(\pi,\theta) \text{ over } \mathcal{P}_{V,\theta}$$

with

$$\mathcal{P}_{V,\theta} := \{\pi \in \mathcal{P} : N(\pi,\theta) \leq V\}. \tag{3.3}$$

The reader is refered to [5] for a complete solution to this problem, the main features of which are summarized below.

Theorem 3.1 *If $N((\infty,1),\theta) \leq V$, then the policy $(\infty,1)$ solves problem $(P_{V,\theta})$. If $N((\infty,1),\theta) > V$, then there exists a threshold policy $(L(V,\theta),\eta(V,\theta))$ which solves problem $(P_{V,\theta})$ and the pair $(L(V,\theta),\eta(V,\theta))$ is the only solution to the equation*

$$N((L,\eta),\theta) = V, \quad L = 0,1,\ldots \text{ and } 0 \leq \eta \leq 1 \tag{3.4}$$

4. The adaptive threshold policies

Since the solution to the constrained problem described in Theorem 3.1 is parametrized by θ, implementing it requires knowledge of the actual value of this parameter. In some situations, this information may not be available to the decision-maker who is then faced with an adaptive version of the constrained control problem. In this paper, a non-Bayesian version of the adaptive control problem of Section 3 is considered, with θ^* in $(0,1)^2$ denoting the

true value of the service rate vector, which is assumed to be a *fixed* constant *unknown* to the decision-maker.

This adaptive control problem can be approached through the Certainty Equivalence Principle by coupling the control process with a parameter estimation scheme. Here, the estimation scheme is based on the Principle of Maximum Likelihood (ML) and the corresponding adaptive threshold policy (denoted by α) is defined by substituting the ML estimates for the actual value of the parameter in the definition of the threshold policy described in Theorem 3.1.

This is now formalized as follows. The information available to the decision-maker is encoded in the RV's $\{\tilde{H}(n)\}_0^\infty$ defined recursively by

$$\tilde{H}(n+1) = (\tilde{H}(n), U(n), A(n), D(n)) \qquad n = 0, 1, \ldots (4.1)$$

with $\tilde{H}(0) = \Xi$, where the *departure* (or *service completion*) sequence $\{D(n)\}_0^\infty$ is given by

$$D(n) := 1[X(n) \neq 0]B(n). \qquad n = 0, 1, \ldots (4.2)$$

At each time n, the ML estimate $\theta(n) := (\lambda(n), \mu(n))$ of the true parameter $\theta^* = (\lambda^*, \mu^*)$ is used in the interval $[n, n+1)$ and is generated from the information $\tilde{H}(n)$ by maximizing a *likelihood functional* evaluated on this *observed* data trajectory. The corresponding Certainty Equivalence policy α in $\mathcal{P}$ is then defined by

$$\alpha_n(H(n)) = \begin{cases} 1 & \text{if } 0 \leq X(n) < L(n); \\ \eta(n) & \text{if } X(n) = L(n); \\ 0 & \text{if } X(n) > L(n) \end{cases} \qquad n = 0, 1, \ldots (4.3)$$

with

$$L(n) = L(V, \theta(n)) \quad \text{and} \quad \eta(n) = \eta(V, \theta(n)). \qquad n = 0, 1, \ldots (4.4)$$

The ML estimate $\theta(n) = (\lambda(n), \mu(n))$ of the service rate vector is then determined by

$$\theta(n) = \arg \max_{\theta \in [0,1]^2} P^{\theta,\alpha}[\tilde{H}(n) = h_n]_{h_n = \tilde{H}(n)} \qquad n = 0, 1, \ldots (4.5)$$

with a tie-breaker if necessary. The procedure described by (4.3)-(4.5) is well defined when performed sequentially starting with $\theta(0)$ arbitrary in $[0,1]^2$ ($\theta(0) = \arg \max_{\theta \in [0,1]^2} q(\Xi)$).

The Certainty Equivalence policy α defined above constitutes an implementation of the threshold policy $(L(V,\theta^*),\eta(V,\theta^*))$ based on the estimates $\{\theta(n)\}_0^\infty$ for θ^*. The RV's $\{L(n)\}_0^\infty$ and $\{\eta(n)\}_0^\infty$ can be viewed as estimates of the threshold parameters $L(V,\theta^*)$ and $\eta(V,\theta^*)$.

5. The main results

Fix $V > 0$ and from now on denote the threshold parameters $L(V,\theta^*)$ and $\eta(V,\theta^*)$ by L^* and η^*, respectively. In order to simplify the presentation, the probability measure $P^{\theta^*,\pi}$ and its expectation operator $E^{\theta^*,\pi}$, associated with any admissible policy π in P, are denoted by P^π and E^π, respectively.

The results reported here all hinge on the following key strong consistency result for the parameter estimates $\{\theta(n)\}_0^\infty$ under the policy α.

Theorem 5.1 - Parameter identification. *The convergence*

$$\lim_n \lambda(n) = \lambda^* \quad \text{and} \quad \lim_n \mu(n) = \mu^* \qquad P^\alpha - a.s. \tag{5.1}$$

takes place.

The proof of Theorem 5.1 is outlined in Sections 6 and 7, and represents the main technical contribution of this paper since the strong consistency reported here is not subsumed by earlier results on the non-Bayesian adaptive control of Markov chains [1,8]. The specific arguments given here are of independent interest and could probably be tailored to other situations as well [7].

The definition of the threshold parameters L^* and η^* through (3.4) easily yields the following result on the asymptotic agreement of the policies (L^*,η^*) and α.

Theorem 5.2 - Control identification. *The convergence*

$$\lim_n |\alpha_n(H(n)) - (L^*,\eta^*)(X(n))| = 0 \qquad P^\alpha - a.s. \tag{5.2}$$

takes place, and in particular, if $0 < \eta^* < 1$, *then*

$$\lim_n L(n) = L^* \quad \text{and} \quad \lim_n \eta(n) = \eta^*. \qquad P^\alpha - a.s. \tag{5.3}$$

The optimality of the adaptive policy α for the problem (P_{V,θ^*}) is now a consequence of Theorem 5.2 and of an argument originally proposed by Mandl [8,9].

Theorem 5.3 - Cost identification. *The adaptive policy α solves the problem (P_{V,θ^*}). In particular, if $N((\infty,1),\theta^*) \leq V$, then $T(\alpha,\theta^*) = T((\infty,1),\theta^*) = \lambda^*$, while if $N((\infty,1),\theta^*) > V$, $T(\alpha,\theta^*) = T((L^*,\eta^*),\theta^*)$ and $N(\alpha,\theta^*) = N((L^*,\eta^*),\theta^*) = V$.*

6. The ML estimates

Explicit expressions can be derived for the ML estimates $\{\theta(n)\}_0^\infty$ defined through (4.5). In order to state the results, it is notationally convenient to define the $\mathbb{N}$-valued RV's $\{N(n)\}_0^\infty$ by

$$N(n) := \sum_{t=0}^{n-1} 1[X(t) \neq 0] \qquad n = 1,2,\ldots (6.1)$$

with $N(0) = 0$; the RV $N(n)$ counts the number of slots over $[0,n)$ during which the queue is non-empty.

Theorem 6.1 *The ML estimates $\{\theta(n)\}_0^\infty$ defined through (4.5) are given by*

$$\lambda(n) = \frac{1}{n}\sum_{t=0}^{n-1} A(t) \qquad n = 1,2,\ldots (6.2)$$

and

$$\mu(n) = \begin{cases} \frac{1}{N(n)}\sum_0^{n-1} D(t) & \text{if} \quad N(n) > 0; \\ \textit{arbitrary in } [0,1] & \text{if} \quad N(n) = 0, \end{cases} \qquad n = 1,2,\ldots (6.3)$$

with $\theta(0) = (\lambda(0),\mu(0))$ arbitrary in $[0,1]^2$.

The *strong consistency* of the ML estimates $\{\lambda(n)\}_0^\infty$ under P^α is an immediate consequence of the Law of Large Numbers, since the RV's $\{A(n)\}_0^\infty$ form a Bernoulli sequence with rate λ^* under P^α. However, establishing the strong consistency of the estimates $\{\mu(n)\}_0^\infty$ is a much more challenging task to which the remainder of this paper is devoted. The next proposition provides a preliminary yet useful characterization for this convergence.

Theorem 6.2 *For any policy π in $\mathcal{P}$, with the estimates $\{\mu(n)\}_0^\infty$ defined by (6.3), the convergence*

$$\lim_n \mu(n) = \mu^* \qquad P^\pi - a.s. (6.4)$$

takes place whenever the condition

$$\liminf_n \frac{N(n)}{n} > 0 \qquad P^\pi - a.s. (6.5)$$

holds.

Proof. On the set $\Omega_0 = [\liminf_n \frac{N(n)}{n} > 0]$, the quantity $N(n)$ becomes positive for n sufficiently large, at which time the relation

$$\mu(n) - \mu^* = \frac{1}{n} \sum_0^{n-1} 1[X(t) \neq 0][B(t) - \mu^*] \cdot [\frac{N(n)}{n}]^{-1}. \tag{6.6}$$

holds. It is plain from (R1)-(R3) that for any π in $\mathcal{P}$, the RV's $\{V(t)[B(t) - \mu^0]\}_0^\infty$ form an *uncorrelated zero-mean* sequence under P^π, and their variance satisfies the bound

$$E^\pi \left[|1[X(t) \neq 0][B(t) - \mu^*]|^2 \right] \leq \mu^*(1 - \mu^*). \qquad n = 0, 1, \ldots \tag{6.7}$$

A version of the Law of Large Numbers [2] now implies the convergence

$$\lim_n \frac{1}{n} \sum_{t=0}^{n-1} 1[X(t) \neq 0][B(t) - \mu^*] = 0 \qquad P^\pi - a.s. \tag{6.8}$$

and the result follows from (6.6) and the definition of Ω_0. □

7. Outline of the proof of Theorem 5.1

In view of Theorem 6.2, strong consistency of the estimates $\{\mu(n)\}_0^\infty$ can be established by showing that $P^\alpha[\Omega_0] = 1$. However, the tight interaction between the estimation and control processes makes it difficult to show this fact in a straightforward way. An indirect approach is thus proposed below that uses auxiliary policies and combines ideas of *stochastic ordering*, rate of convergence via *Large Deviations* and *absolutely continuous* change of measures.

7.1. An auxiliary policy

The form of (6.3) suggests an interpretation of the estimates $\{\mu(n)\}_0^\infty$ as rate estimates for the Bernoulli sequence $\{B(t)\}_0^\infty$ when *sampled* only at the times where the system is not empty. In fact, Theorem 6.2 clearly shows that (6.5) will fail if there are not enough sampling instants and therefore not enough information collected in the long run. However, there does not seem to be any obvious argument to guarantee that under α the queue is not empty often

enough over the infinite horizon. To remedy to this difficulty, consider the auxiliary policies α^ϵ in $\mathcal{P}$ for ϵ in the interval $(0, \frac{1-\mu^*}{\lambda^*}]$, where

$$\alpha_n^\epsilon(H(n)) = \begin{cases} 1 & \text{if } 0 \leq X(n) < L(n); \\ \eta^\epsilon(n) & \text{if } X(n) = L(n); \\ 0 & \text{if } X(n) > L(n) \end{cases} \qquad n = 0, 1, \ldots \quad (7.1)$$

with

$$\eta^\epsilon(n) = \begin{cases} \epsilon \vee \eta(n) & \text{if } L(n) = 0; \\ \\ \eta(n) & \text{if } L(n) \geq 1. \end{cases} \qquad n = 0, 1, \ldots \quad (7.2)$$

In contrast with what is happening under α, in the worst possible case, namely $L(n) = X(n) = 0$, there is a (uniformly) positive probability that the learning process will be activated under α^ϵ, and strong consistency of the parameter estimates should thus be expected under this auxiliary policy. This is made precise in what follows.

Since $(0, \epsilon)_n(H(n)) \leq \alpha_n^\epsilon(H(n))$ and $\lambda^*(0, \epsilon)_n(0) = \lambda^*\epsilon \leq 1 - \mu^*$, for all $n = 0, 1, \ldots$, it follows from the comparison results developed in [5, Thm. A.3] that

$$(\{X(n)\}_0^\infty, P^{(0,\epsilon)}) \leq_{st} (\{X(n)\}_0^\infty, P^{\alpha^\epsilon}). \quad (7.3)$$

It is now easy to see from (7.3) [10, Thm. 4.1.2, pp. 61] that

$$1 = P^{(0,\epsilon)}[\Omega_0] \leq P^{\alpha^\epsilon}[\Omega_0] \leq 1 \quad (7.4)$$

and therefore $P^{\alpha^\epsilon}[\Omega_0] = 1$ for all ϵ in the interval $(0, \frac{1-\mu^*}{\lambda^*}]$. The equality in (7.4) is a simple consequence of the fact that under $P^{(0,\epsilon)}$, the RV's $\{X(n)\}_0^\infty$ form a Markov chain with $\{0, 1\}$ as its single positive recurrent class and therefore $\lim_n \frac{N(n)}{n} > 0$ $P^{(0,\epsilon)} - a.s.$ by the strong Ergodic Theorem for Markov chains.

7.2. An absolutely continuous change of measures

In view of the fact that $P^{\alpha^\epsilon}[\Omega_0] = 1$ for every ϵ in the interval $(0, \frac{1-\mu^*}{\lambda^*}]$, the condition $P^\alpha[\Omega_0] = 1$ necessarily holds if it could be shown that for some such ϵ,

$$P^\alpha << P^{\alpha^\epsilon} \quad \text{on } \mathcal{F}. \quad (7.5)$$

To proceed with this line of arguments, first notice that for every ϵ in $(0, \frac{1-\mu^*}{\lambda^*}]$,

$$P^{\alpha} << P^{\alpha^{\epsilon}} \quad \text{on } \mathbb{F}_n \qquad n = 1, 2, \ldots (7.6)$$

while routine calculations show that the corresponding Radon-Nykodym derivative $L^{\epsilon}(n)$ satisfies the inequality

$$L^{\epsilon}(n) \leq \prod_{k=0}^{n-1} \Big\{ 1 + B|\eta(k) - \eta^{\epsilon}(k)| \Big\} \qquad n = 1, 2, \ldots (7.7)$$

for some positive constant B independent of ϵ.

Fix ϵ in the interval $(0, \eta^* \wedge \frac{1-\mu^*}{\lambda^*})$. ¿From (7.4) and the characterization result of Theorem 6.2, the convergence $\lim_n \mu(n) = \mu^*$ takes place $P^{\alpha^{\epsilon}}$-a.s. and since $\eta^* > 0$, it is then easy to see that $\lim_n \eta(n) = \eta^*$ $P^{\alpha^{\epsilon}} - a.s.$ The condition on ϵ now implies that $\lim_n \eta^{\epsilon}(n) = \eta^*$ also $P^{\alpha^{\epsilon}} - a.s.$ and therefore $\eta(n) = \eta^{\epsilon}(n)$ after a number τ_{ϵ} slots given by

$$\tau_{\epsilon} := \max\{n \geq 0 : \eta^{\epsilon}(n) \neq \eta(n)\} \qquad (7.8)$$

with τ_{ϵ} being a.s. *finite* under $P^{\alpha^{\epsilon}}$.

All these remarks now combine with (7.7) to yield

$$L^{\epsilon}(n) \leq \prod_{k=0}^{\tau_{\epsilon}} \Big\{ 1 + B|\eta(k) - \eta^{\epsilon}(k)| \Big\} \leq \exp[\epsilon B(\tau_{\epsilon} + 1)] \qquad n = 1, 2, \ldots (7.9)$$

where the fact

$$\eta^{\epsilon}(n) - \eta(n) = 1[L(n) = 0, \eta(n) < \epsilon](\epsilon - \eta(n)) \qquad n = 0, 1, \ldots (7.10)$$

has been used.

7.3. Rate of convergence via Large Deviations Theory

The absolutely continuity (7.5) will be obtained for some ϵ in the interval $(0, \eta^* \wedge \frac{1-\mu^*}{\lambda^*})$ whenever the RV's $\{L^{\epsilon}(n)\}_0^{\infty}$ are *uniformly integrable* under $P^{\alpha^{\epsilon}}$. This will be the case if the integrability condition

$$E^{\alpha^{\epsilon}}\big[\exp[\epsilon B \tau_{\epsilon}]\big] < \infty \qquad (7.11)$$

holds, i.e., if the RV τ_ϵ has *exponential* moments. This will established by showing that the RV τ_ϵ has an exponential tail, namely that there exists positive constants a_ϵ and C_ϵ such that

$$P^{\alpha^\epsilon}[\tau_\epsilon \geq n] \leq a_\epsilon \exp[-nC_\epsilon] \qquad n = 0, 1, \ldots (7.12)$$

with $\lim_\epsilon C_\epsilon = \infty$ as $\epsilon \downarrow 0$.

This in turn can be established with the help of the Theory of Large Deviations by a careful analysis of the rate of convergence of the estimate sequences $\{\mu(n)\}_0^\infty$ and $\{\lambda(n)\}_0^\infty$. More precisely, for every $\delta > 0$ it can be shown that there exist positive constants a_1, a_2, $K_1(\delta)$ and $K_2(\delta)$ such that

$$P^{\alpha^\epsilon}[|\lambda(n) - \lambda^*| > \delta] \leq a_1 \exp[-nK_1(\delta)] \qquad n = 0, 1, \ldots (7.13a)$$

and

$$P^{\alpha^\epsilon}[|\mu(n) - \mu^*| > \delta] \leq a_2 \exp[-nK_2(\delta)] \qquad n = 0, 1, \ldots (7.13b)$$

It is easy to see from these exponential bounds that the sequence $\{\eta(n)\}_0^\infty$ also exhibits a similar exponential behavior and the result (7.12) readily follows.

The first exponential bound (7.13a) follows by a simple application of Cramer's Theorem [4,11] to the Bernoulli sequence $\{A(t)\}_0^\infty$. The derivation of the second exponential bound (7.13b) is more involved and takes as point of departure the fact that $\mu(n) =_{st} \frac{S(N(n))}{N(n)}$ where

$$S(t) := \sum_{i=0}^{t-1} B(i) \qquad t = 1, 2, \ldots (7.14)$$

with $S(0) = 0$. With this in mind, it is plain that for $0 < a < 1$,

$$\begin{aligned} P^{\alpha^\epsilon}[|\mu(n) - \mu^*| > \delta] \leq\ & P^{\alpha^\epsilon}\left[|\frac{S(N(n))}{N(n)} - \mu^*| > \delta, \frac{N(n)}{n} > a\right] \\ & + P^{\alpha^\epsilon}\left[\frac{N(n)}{n} \leq a\right] \end{aligned} \qquad n = 0, 1, \ldots (7.15)$$

The first term on the right handside of (7.15) is easily seen to decay exponentially fast to 0 as a result of Cramer's result applied to the Bernoulli sequence $\{B(t)\}_0^\infty$. For the second term, note from (7.3) that

$$P^{\alpha^\epsilon}[\frac{N(n)}{n} \leq a] \leq P^{(0,\epsilon)}[\frac{N(n)}{n} \leq a] \qquad n = 0, 1, \ldots (7.16)$$

and recall that the RV's $\{\frac{N(n)}{n}\}_0^\infty$ also obey a Large Deviations Principle as discussed by Donsker and Varadhan [3,4]. Therefore, if a is chosen small enough, namely $0 < a < \lim_n \frac{N(n)}{n}$ where the a.s limit is taken under $P^{(0,\epsilon)}$, the second term on the right handside of (7.15) can be shown to go to 0 exponentially fast. □

8. Comments on information patterns

The information pattern implicit in the RV's $\{\tilde{H}(n)\}_0^\infty$ is coarser than the information pattern associated with the information RV's $\{H(n)\}_0^\infty$ that was used in the formulation of the constrained flow control problems of Section 3. Indeed, knowledge of $\tilde{H}(n)$ contains knowledge of $D(n)$ but not necessarily of $B(n)$.

To understand why this coarser information pattern was selected, recall that the Bernoulli RV's $\{B(n)\}_0^\infty$ were introduced for modelling purpose and that only the actual departures $\{D(n)\}_0^\infty$ have a physical meaning and are thus observable. Moreover, it is also worth pointing out that the ML estimate $\underline{\theta}(n) = (\underline{\lambda}(n), \underline{\mu}(n))$ of the rate vector on the basis of the information $H(n)$ (instead of $\tilde{H}(n)$) is given by $\underline{\lambda}(n) = \lambda(n)$ and

$$\underline{\mu}(n) = \frac{1}{n}\sum_{t=0}^{n-1} B(t). \qquad n = 0, 1, \ldots (8.1)$$

Obviously the estimates $\{\underline{\theta}(n)\}_0^\infty$ are strongly consistent under P^α by virtue of the Law of Large Numbers, since both sequence of RV's $\{A(n)\}_0^\infty$ and $\{B(n)\}_0^\infty$ are Bernoulli sequences with rates λ^* and μ^* under P^α. This situation is clearly not too interesting.

References

[1] V. Borkar and P. Varaiya, "Identification and adaptive control of Markov chains," *SIAM J. Control Opt.*, Vol. 20 (4), pp. 470-489 (1982).

[2] K. L. Chung, *A Course in Probability Theory*, Second Edition, Academic Press, New York (1974).

[3] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time, I," *Comm. Pure Appl. Math.*, Vol. 27, pp. 1-47 (1975).

[4] R. S. Ellis, "Large deviations for a general class of random vectors," *Ann. Probab.*, Vol. 12, pp. 1-12 (1984).

[5] D.-J. Ma and A. M. Makowski, "Optimality results for a simple flow control problem," these Proceedings.

[6] D.-J. Ma and A. M. Makowski, "A simple problem of flow control II: Implementation of threshold policies via Stochastic Approximations," *IEEE Trans. Auto. Control*, submitted (1987).

[7] D.-J. Ma and A. M. Makowski, "A simple problem of flow control III: Parameter estimation under threshold policies," *IEEE Trans. Auto. Control*, submitted (1987).

[8] P. Mandl, "Estimation and control in Markov chains," *Adv. Appl. Prob.*, Vol. 6, pp. 40-60 (1974).

[9] A. Shwartz and A. M. Makowski, "Comparing policies in Markov decision processes: Mandl's lemma revisited," *Mathematics of Operations Research*, submitted (1987).

[10] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, John Wiley & Sons, New York (1983).

[11] S. R. S. Varadhan, *Large deviations and Applications*, CBMS-NSF Regional conference Series in Applied Mathematics, Vol. 46, SIAM, Philadelphia, Pennsylvania (1984).