A Note on Bias Optimality in Controlled Queueing Systems*†

Mark E. Lewis Martin L. Puterman

Faculty of Commerce and Business Administration

University of British Columbia, 2053 Main Mall, Vancouver, BC Canada V6T 1Z2

Submitted January 5, 1999; Revised May 17, 1999

Abstract

The use of *bias optimality* to distinguish among gain optimal policies was recently studied by Haviv and Puterman [1] and extended in Lewis, et al. [2]. In [1], upon arrival to an $M/M/1$ queue, customers offer the gatekeeper a reward R. If accepted, the gatekeeper immediately receives the reward, but is charged a holding cost, $c(s)$, depending on, the number of customers in the system. The gatekeeper, whose objective is to "maximize" rewards, must decide whether to admit the customer. If the customer is accepted, the customer joins the queue and awaits service. Haviv and Puterman [1] showed there can only be two Markovian, stationary, deterministic gain optimal policies and that only the policy which uses the *larger* control limit is bias optimal. This showed the usefulness of bias optimality to distinguish between gain optimal policies. In the same paper, they conjectured that if the gatekeeper receives the reward upon *completion* of a job instead of upon entry, the bias optimal policy will be the lower control limit. This note confirms that conjecture.

1 Introduction

The use of *bias optimality* to distinguish among gain optimal policies was recently studied by Haviv and Puterman [1] and extended in Lewis, et al. [2]. In Haviv and Puterman's model, upon arrival to an $M/M/1$ queue, customers offer the gatekeeper a reward R. If accepted, the gatekeeper immediately receives the reward, but is charged a holding cost, $c(s)$, depending on, the number of customers in the system. The gatekeeper, whose objective is to "maximize" rewards, must decide whether to admit the customer. If the customer is accepted, the customer joins the queue and awaits service. Haviv and Puterman [1] showed that when considering long-run average reward, or *gain*, there can only be two Markovian, stationary, deterministic optimal policies. In fact, these policies are of control limit form and

AMS Subject Classifications: **primary-90C40:** Markov and semi-Markov decision processes, **secondary-60K25:** queueing theory

†*IAOR Subject Classifications:* **primary-3160:** Markov processes, **secondary-3390:** queues: theory

they occur consecutively. Further, only the policy which uses the *larger* control limit is bias optimal. This showed the usefulness of bias optimality to distinguish between gain optimal policies.

Intuitively, among the policies which are long-run average optimal, bias opimality finds the policy that maximizes the total reward up to steady state. In the last section of Haviv and Puterman's analysis, the authors explain that this is so because the bias can be thought of as a limit of discount optimal policies. Since the holding cost is state dependent, and therefore is actually spread out over subsequent periods it is in some sense discounted. By this argument, they conjectured that if the gatekeeper receives the reward upon completion of a job instead of upon entry, the bias optimal policy will be the lower control limit. This note will confirm this conjecture.

Example 1 *Suppose the arrival and service rates of customers are* $1/3$ *and* $2/3$*, respectively. Further suppose that the holding cost per unit time in state* s *is* $c(s) = 2s$*. Let the system capacity be* 10*. The gains of control limit policies* 2 *and* 3 *are maximal and are both* 2.5*. However, letting* h_2 *and* h_3 *denote the bias vectors of control limit policies* 2 *and* 3 *respectively, we have*

$$\begin{aligned} h_2 &= \{-3.64286, 3.85714, 6.85714, 6.85714, 3.85714, -2.14286, \\ &\quad -11.1429, -23.1429, -38.1429, -56.1429, -77.1429\}, \\ h_3 &= \{-4.1, 3.4, 6.4, 6.4, 3.4, -2.6, -11.6, -23.6, -38.6, -56.6, -77.6\} \end{aligned}$$

Notice that $h_2 \geq h_3$*. We show that this holds in general for the problem considered.*

2 Definitions and Preliminaries

Let λ be the arrival rate of a Poisson arrival process and let μ be the exponential service rate. To ensure stability, assume that $\lambda < \mu$. Although the model is actually an infinite state queueing model, we consider a finite state truncation. Haviv and Puterman (Section 3) [1] present an argument to justify this provision which we will not reiterate here. Hence, let $S = \{0, 1, \ldots, U\}$ be the state space. Suppose $c(s)$ is the holding cost to the system per unit time when there are s customers in the system. We assume that this cost is nondecreasing and convex in the state. Note that as an alternative to the stability condition ($\lambda < \mu$), we could assume that the cost function is strictly increasing. Under this assumption, the state space truncation would again be justified as the cost of an arriving customer would eventually overtake the benefit of admission. Stidham [4] showed the existence of an optimal policy of control limit form for the model we consider. When a job is completed the customer pays the gatekeeper a reward R and leaves the system. Customers that are rejected are lost. We have defined a finite state, infinite horizon, continuous-time Markov decision process. As has become standard we actually consider the discrete time uniformized version of this process. Uniformization was first introduced in this context by Lippman [3].

Let $\lambda + \mu$ be the uniformization constant and without loss of generality assume $\lambda + \mu = 1$. Let P_d be the uniformized transition matrix of the decision rule $d \subset D$, where D is the set of Markovian, deterministic decision rules. A stationary policy made up of such decision rules

will be denoted d^∞. A decision rule which accepts arriving customers as long as the number of customers in the system is less than a prespecified level after which it rejects arriving customers is called a control limit decision rule. We denote the control limit decision rule with limit L by L. Hence,

$$P_L(j|s) = \begin{cases} \lambda & j = s+1, 0 \le s < L; j = s, s \ge L, \\ \mu & j = 0, s = 0; j = s-1, 1 \le s, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

It is well-known that an optimal gain vector can be determined by solving the following set of equations:

$$g = \max_{d \in D} P_d g \quad (2)$$

and

$$h = \max_{d \in G(g)} \{r_d - g + P_d h\} \quad (3)$$

for g and h where $G(g) \subseteq D$ is the set of decision rules that attain the maximum in (2). We refer to the system of equations (2) and (3) as the *gain optimality equations*. Let $H(g,h) \subseteq G(g)$ be the set of decision rules that attain the maximum in (3). It was shown in Lewis et. al. [2] under the assumption that the embedded Markov chain is irreducible, that $H(g,h)$, the set of decision rules that attain the maximum in (3), is the same for all g and h that satisfy (2) and (3). When the model is unichain, and the transient states only have one possible action, the proof found there extends. Hence, we will supress the dependence on g and h and write H. Proposition 5.3 in Haviv and Puterman [1] shows among other results that the *bias optimality equation* (4) characterizes bias optimal policies.

Proposition 1 *Suppose g and h satisfy the gain optimality equations and that there exists a vector w for which*

$$w = \max_{d \in H} \{-h + P_d w\}. \quad (4)$$

Then, if $d^* \in H$ attains the maximum in (4), $(d^*)^\infty$ is bias optimal. Moreover, suppose that the process is unichain and that for some $d \in H$ where g and h satisfy the gain optimality equations, there exists a vector w with

$$w = -h + P_d w. \quad (5)$$

Then if

$$P_{d'} w \ge P_d w \quad (6)$$

for some $d' \in H$ with strict inequality in at least one recurrent state in the Markov chain generated by d', $h^{(d')^\infty} \ge h^{d^\infty}$ with strict inequality in at least one component.

Remark 1 We note that the statement of the result in [1] does not include the essential conditions that the process must be unichain and that the strict inequality in (6) must occur at a recurrent state in the Markov chain generated by d'.

3 Formulation of Payment on Exit Model

In the model that we propose the gatekeeper does not receive rewards upon entry. Instead, the reward is received upon departure. In effect, the cost associated with an arriving customer is accrued before the reward is received. Hence, it seems that the gatekeeper may be less willing to accept customers than in the previous model of Haviv and Puterman [1]. As before, when a customer arrives the gatekeeper can either accept or reject the customer. To avoid degenerate cases assume that it is optimal for the gatekeeper to accept customers when the system is empty. We further assume that $c(1) > c(0) = 0$. Note that for any stationary policy the induced Markov decision process is unichain. Since the gain of these policies is a constant vector the first gain optimality equation (2) is superfluous. Consider the second gain optimality equation, written in component form. If $s > 1$,

$$\begin{aligned} h(s) &= \max\{-c(s) - g + \lambda h(s+1) + \mu(R + h(s-1)), \\ &\quad -c(s) - g + \lambda h(s) + \mu(R + h(s-1)\}. \end{aligned} \tag{7}$$

If $s = 0$,

$$h(0) = -g + \lambda h(1) + \mu h(0). \tag{8}$$

Note that the first term in the maximization in (7) corresponds to accepting the next arrival and the second corresponds to rejecting it.

For ease of notation, for a function $f(s)$ let $\Delta f(s) = f(s+1) - f(s)$. From (7) it follows that it is gain optimal to accept an arriving customer if $\Delta h(s) > 0$ and it optimal to reject if $\Delta h(s) < 0$. When equality holds we are indifferent. One interesting thing to note is that the optimality criterion does not explicitly depend on R. However, the assumption that it is optimal to accept in state zero, implies from (8) that $\Delta h(0) = g/\lambda \geq 0$ which is an assumption on R. Consider state 1. If it is optimal to accept an arriving customer then $\Delta h(1) \geq 0$ and we have

$$h(1) = -c(1) - g + \lambda h(2) + \mu(R + h(0)). \tag{9}$$

Subtracting (8) from (9) we have

$$\Delta h(1) = \frac{1}{\lambda}[(c(1) + \Delta h(0)) - \mu R] \geq 0. \tag{10}$$

We get a similar expression if it is optimal to reject in state 1.

We now state the main result of this note.

Theorem 1 *Suppose control limits L and $L+1$ are gain optimal in the payment on departure model. Then $h_{L+1} \leq h_L$ with strict inequality in at least one component.*

Proof: Since the transition matrix is exactly the same as the model discussed in Haviv and Puterman [1] we apply Proposition 1 in the same manner. Since decision rules L and $L+1$

only differ in state L, we restrict our attention to that state. The L^{th} component of the vector w which satisfies the bias optimality equations is given by

$$\begin{aligned} w(L) &= \max\{-h_L(L) + \lambda w(L+1) + \mu w((L-1)^+), && (11)\\ &\quad -h_L(L) + \lambda w(L) + \mu w((L-1)^+)\}, && (12)\end{aligned}$$

where the first quantity corresponds to admitting a customer and the second to rejecting. Hence, it is optimal to reject if

$$w(L) < w(L+1) \text{ or, equivalently } \Delta w(L) < 0. \tag{13}$$

Hence, if we establish (13), the result follows from the second part of Proposition 1. Following the analysis in Haviv and Puterman [1] we get

$$(I - T)\Delta w_L = z \tag{14}$$

where w_L is the vector which satisfies the $w_L = -h_L + P_L w_L$, z is defined by

$$z(s) = \begin{cases} -\Delta h_L(s) & 0 \le s < L \\ -\frac{\Delta h_L(L)}{\mu} & s = L, \end{cases} \tag{15}$$

and

$$T(j|s) = \begin{cases} \lambda & j = s+1, 0 \le s < L-1, \\ \mu & j = s-1, 0 < s \le L-1, \\ 1 & j = s-1, s = L \\ 0 & \text{otherwise} \end{cases}. \tag{16}$$

Noting that the zero$^{\text{th}}$ and $(L-1)^{\text{st}}$ rows have row sums strictly less than one, T can be viewed as a $(L+1)\times(L+1)$ transition submatrix of transient states for a Markov chain and

$$(I - T)^{-1} = \sum_{i=0}^{\infty} T^i \tag{17}$$

Hence we get,

$$\Delta w = (I - T)^{-1} z \tag{18}$$

Note that all of the elements of T are nonnegative and that each row has at least one strictly nonegative element. Since L is the lowest gain optimal control limit, $\Delta h_L(s) > 0$ for all $s < L$. Further, since $L+1$ is also gain optimal $\Delta h_L(L) = 0$. Hence, using the second order approximation of the power series in (17) we get,

$$\begin{aligned} \Delta w(L) &\le -\Delta h_L(L) - (T\Delta h_L)(L) \\ &< 0 \end{aligned}$$

from which the result follows. □

4 Further Comments

Finally, we note the following minor errors in [1]. The definition for $z(s)$ should read

$$z(s) = \begin{cases} -\Delta h_L(s) & 0 \leq s < L, \\ -\frac{\Delta h_L(L)}{1-p} & s = L. \end{cases} \tag{19}$$

Further, (54) should be an "s" instead of a "2".

This note confirms the prior conjecture that bias optimality captures the intuitive result that a decision-maker should be more restrictive if the cost of an arriving customer must be incurred before receiving the reward. Further, although bias optimality does not explicitly discount rewards, it leads to selection of policies which receive rewards earlier.

References

[1] Moshe Haviv and Martin L. Puterman. Bias optimality in controlled queueing systems. *Journal of Applied Probability*, 35:136–150, 1998.

[2] Mark E. Lewis, Hayriye Ayhan, and Robert D. Foley. Bias optimality in a queue with admission control. *Probability in the Engineering and Informational Sciences*, 1999. to appear.

[3] Steven A. Lippman. Applying a new device in the optimization of exponential queueing systems. *Operations Research*, 23(4):687–712, 1975.

[4] Shaler Stidham, Jr. Socially and individually optimal control of arrivals to a $GI/M/1$ queue. *Management Science*, 24:1598–1610, 1978.