Adv. Appl. Prob. **23**, 429–430 (1991)
Printed in N. Ireland
© *Applied Probability Trust* 1991

LETTERS TO THE EDITOR

ADDENDUM TO 'ON AN INDEX POLICY FOR RESTLESS BANDITS'

RICHARD R. WEBER,* *University of Cambridge*
GIDEON WEISS,** *Georgia Institute of Technology*

Abstract

We show that the fluid approximation to Whittle's index policy for restless bandits has a globally asymptotically stable equilibrium point when the bandits move on just three states. It follows that in this case the index policy is asymptotic optimal.

In [2] we investigated properties of an index policy for restless bandits that had been the subject of an interesting paper by Whittle [3]. We showed that if the fluid approximation to his index policy has a globally asymptotically stable equilibrium point then it is asymptotically optimal, for the problem of choosing which m out of n bandits to make active, as $m, n \to \infty$, with $m/n = \alpha$. We observed that the existence of such a point is guaranteed when the bandits move on just $k = 2$ states. However, a counterexample with $k = 4$ states showed that this is not the case in general (though with very small suboptimality). The conjecture that the index policy might be asymptotically optimal when the bandits move on $k = 3$ states was left unanswered. The present note confirms that conjecture. In this note we use the notation of [2] and refer to formula and theorem numbers in that paper.

The state of the n arms (or bandits) under application of the index policy is expressed by a probability vector $z_n(t) = (z_{n1}(t), z_{n2}(t), z_{n3}(t))$. The fluid approximation to $z_n(t)$ is given by the solution to $\dot{z} = Q(z)z$ (10), where the $q_{ji}(z)$ are given by (9).

Lemma 1. *Assume the problem is indexable with index order* 1, 2, 3. *Then the fluid approximation for $z_n(t)$ is globally asymptotically stable.*

Proof. Imposing the condition that $z_1(t) + z_2(t) + z_3(t) = 1$ we eliminate $z_2(t)$ and the equation for $\dot{z}_2(t)$, and we partition the region $C = \{z_1(t), z_3(t) \geqq 0, z_1(t) + z_3(t) \leqq 1\}$ into regions $C_1 = \{z_1(t) \geqq 1 - \alpha\}$, $C_2 = \{z_1(t) \leqq 1 - \alpha, z_3(t) \leqq \alpha\}$, $C_3 = \{z_3(t) \geqq \alpha\}$. Here C_i is the region in which arms of index greater or less than i are made active or passive respectively, and a proportion of the arms of index i are made active. As in [2], let q^1_{ji} and q^2_{ji} be the transition rates from state i to j under the active and passive actions respectively. The equations (10) in region C_i are of the form

$$\begin{pmatrix} \dot{z}_1 \\ \dot{z}_3 \end{pmatrix} = b_i + A_i \begin{pmatrix} z_1 \\ z_3 \end{pmatrix}, \quad i = 1, 2, 3 \tag{1}$$

Received 27 November 1990; revision received 6 February 1991.

* Postal address: Management Studies Group, Department of Engineering, Mill Lane, Cambridge CB2 1RX, UK.

** Postal address: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0205, USA.

Research supported by NSF grant DDM–8914863.

https://doi.org/10.2307/1427757 Published online by Cambridge University Press

where

$$A_i = \begin{pmatrix} -q^k_{21} - q^k_{31} - q^k_{12} & q^l_{13} - q^l_{12} \\ q^k_{31} - q^k_{32} & -q^l_{13} - q^l_{23} - q^l_{32} \end{pmatrix}$$

and $(k, l) = (1, 1)$ for $i = 1$, $(k, l) = (2, 1)$ for $i = 2$, $(k, l) = (2, 2)$ for $i = 3$. The main thing to note is that A_i has negative diagonal elements for $i = 1, 2, 3$. Let us write

$$\dot{z}_1 = Z_1(z_1, z_3), \qquad \dot{z}_3 = Z_3(z_1, z_3).$$

Then Z_1, Z_3 are continuous throughout C, and are continuously differentiable within each region C_i, $i = 1, 2, 3$. Also,

$$\frac{\partial Z_1}{\partial z_1} + \frac{\partial Z_3}{\partial z_3}$$

is the sum of the diagonal elements of A_i for $z \in C_i$ and so is negative in each of C_1, C_2, C_3. Under these conditions, Bendixson's negative criterion [1] states that no solution to (1) in C can have limit cycles.

It is easy to verify that no solution can leave C. It follows from Theorem 2 that the stationary distribution of the relaxed policy is also the unique equilibrium point of (1) in C. Hence, by the Poincaré–Bendixson theorem [1], every solution of (1) in C converges to that equilibrium point. This proves the lemma.

Applying Theorem 2 also gives the following.

Corollary 2. *For $k = 3$, Whittle's index policy* [3] *is asymptotically optimal as $m, n \to \infty$, with $m/n = \alpha$.*

References

[1] Jordan, D. W. and Smith, P. (1987) *Nonlinear Ordinary Differential Equations*, 2nd edn. Clarendon Press, Oxford.

[2] Weber, R. R. and Weiss, G. (1990) On an index policy for restless bandits. *J. Appl. Prob.* **27,** 637–648.

[3] Whittle, P. (1988) Restless bandits: activity allocation in a changing world. In *A Celebration of Applied Probability,* ed. J. Gani, *J. Appl. Prob.* **25A,** 287–298.

https://doi.org/10.2307/1427757 Published online by Cambridge University Press