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 → ∞, with m/n = α. 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 zn(t) = (z1(t), z2(t), z3(t)). The fluid approximation to zn(t) is given by the solution to \( \dot{z} = Q(z)z \) (10), where the qμ(z) are given by (9).

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

Proof. Imposing the condition that z1(t) + z2(t) + z3(t) = 1 we eliminate z3(t) and the equation for z2(t), and we partition the region C = {z1(t), z3(t) ≥ 0, z1(t) + z3(t) ≤ 1} into regions C1 = {z1(t) ≥ 1 − α}, C2 = {z1(t) ≥ 1 − α, z3(t) ≤ α}, C3 = {z1(t) ≥ α}. Here C1 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 qμ1 and qμ2 be the transition rates from state i to j under the active and passive actions respectively. The equations (10) in region Ci are of the form

\[
\begin{pmatrix}
\dot{z}_1 \\
\dot{z}_2
\end{pmatrix} = b_i + A_i \begin{pmatrix}
z_1 \\
z_2
\end{pmatrix}, \quad i = 1, 2, 3
\]

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.
where

\[ A_i = \begin{pmatrix}
-q^k_{21} - q^k_{31} - q^l_{12} & q^l_{13} - q^l_{12} \\
q^k_{51} - q^k_{32} & -q^l_{13} - q^l_{23} - q^l_{32}
\end{pmatrix} \]

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

\[ \dot{z}_1 = Z_1(z_1, z_3), \quad \dot{z}_3 = Z_3(z_1, z_3). \]

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**

