Restless Bandits: Activity Allocation in a Changing World

P. WHITTLE

Abstract

We consider a population of n projects which in general continue to evolve whether in operation or not (although by different rules). It is desired to choose the projects in operation at each instant of time so as to maximise the expected rate of reward, under a constraint upon the expected number of projects in operation. The Lagrange multiplier associated with this constraint defines an index which reduces to the Gittins index when projects not being operated are static. If one is constrained to operate m projects exactly then arguments are advanced to support the conjecture that, for m and n large in constant ratio, the policy of operating the m projects of largest current index is nearly optimal. The index is evaluated for some particular projects.

GITTINS INDEX; MULTI-ARMED BANDITS; SEQUENTIAL SCHEDULING; STIMULATING PRICES; INDEXABILITY

1. Introduction

The multi-armed bandit problem is a classic version of the problem of optimal allocation of activity under certainty. One can phrase it by saying that one has n projects, the state of project i being denoted by x_i (or by $x_i(t)$ if one wishes to emphasise its dependence on time, t). One can operate only one project at once: if one operates project i then one receives reward $g_i(x_i(t))$ in the time-interval $(t, t+1)$ and the transition $x_i(t) \to x_i(t+1))$ follows a Markov rule specific to project i. The unused projects neither yield reward nor change state; current states of all projects are known at any time. The problem is to so choose the project at each moment that the expected discounted reward over an infinite future is maximal.

The problem was proposed first during the Second World War, and had been generally regarded as a challenge difficult to the point of hopelessness as recently as 1980. However, the problem had essentially been solved by Gittins by 1970 (see Gittins and Jones (1974), Gittins (1979)). Gittins showed that to each project one could attach an *index* $v_i(x_i)$, a function of the project label i and the project state x_i alone, and that the optimal policy had the simple

characterisation: operate the project of currently greatest index. It was by no means evident that the optimal policy would take the form of such an *index policy,* and certainly not how the index should be calculated. Gittins covered these points, and gave a subtle characterisation of the appropriate index, which he called a 'dynamic allocation index', but which is now, with greater convenience and justice, referred to as the *Gittins index.*

Gittins' proof of optimality was difficult to follow. Whittle (1980) gave a dynamic programming proof which relied on a different point of view, evaluated the value function, and revealed its relation to the index.

Gittins substantially generalised his initial solution by dealing, for example, with cases in which one operated m projects at a time rather than just one, or cases in which the choice of projects is restricted by, for example, precedence constraints. It is also possible to analyse the 'open' version of the problem, in which new projects may arrive on the scene (see Whittle (1981)).

However, one class of cases seemed to remain unamenable to either Gittins' original approach or Whittle's alternative view. The cases are those for which projects continue to change state even when they are not being operated (although by different transition rules, in general). For example, the 'projects' may be alternative medical treatments for some condition, and 'operation' implies use of a particular treatment. The change of state during operation corresponds to change of information state: the efficacy of the treatment itself does not change, but one's state of knowledge concerning its efficacy does change. Under the usual assumptions, this state of knowledge cannot change while the treatment is not being applied (i.e. the project is not in operation) because neither is one gaining information nor are the actual characteristics of the treatment changing in any way.

However, suppose the characteristics of all treatments are changing, because, for example, the virus which the treatments are intended to combat is mutating. Then the state of every project is indeed changing, whether or not it is in use.

For another, suppose m aircraft are trying to track the positions of n enemy submarines, where $m < n$, so that aircraft must change task from time to time if all submarines are to be monitored. We regard this as a case of the operation of exactly m projects out of n, in that exactly m submarines out of the n are under surveillance at a given time. The problem is to allocate this surveillance. For this problem, the bandits are restless in the most literal sense. While a submarine is under observation, information on its position, etc., is being gained. While it is not, information is usually being lost, because the submarine will certainly be taking unpredictable evasive action.

As a final example, suppose that one has a pool of n employees of whom exactly m are to be set to work at a given time. This is again a situation of

operating m projects out of n, with the problem of choosing which m to operate. One can imagine that employees who are working produce, but at a decreasing rate as they tire. Employees who are resting do not produce, but recover. The 'project' (the employee) is thus changing state whether or not he is at work.

We shall speak of the phases when a project is in operation or not as *active* and *passive* phases. The traditional assumption has been that a project is static in its passive phase. As we have seen, for many problems this is not true: the active and passive phases produce contrary movements in state space. For submarine surveillance the two cases correspond to gain and loss of information respectively. For the labour force the two cases correspond to tiring and recovery.

2. Populations of projects with required work rates

Let us use P to denote a Markov transition operator, so that, if $x(t)$ is the Markov variable concerned, then for a function $\phi(x)$

$$(P\phi)(x) = E[\phi(x(t+1)) \mid x(t) = x]. \tag{1}$$

Suppose that project i has state variable x_i, that it has transition operators P_{i1} and P_{i2} in the active and passive phases, and yields immediate rewards $g_{i1}(x_i)$ and $g_{i2}(x_i)$. Suppose also that one wishes to maximise average reward over an infinite horizon. If one could operate project i without constraint then it would yield an average reward γ_i determined by

$$\gamma_i + f_i(x_i) = \max_{k=1,2} [g_{ik}(x_i) + (P_{ik} f_i)(x_i)] \tag{2}$$

where the maximising alternative indicates the optimal action to be taken. We suppose for simplicity that the states of project i communicate well enough that γ_i is state-independent. The function $f_i(x_i)$ is the differential reward caused by the transient effect of starting from state x_i rather than from an equilibrium situation. We shall write (2) more compactly as

$$\gamma_i + f_i = \max [L_{i1} f_i, L_{i2} f_i] \quad (i = 1, 2, \cdots, n). \tag{3}$$

Let $m(t)$ be the number of projects which are active at time t. Then a generalisation of the classic multi-armed bandit problem would be to seek an optimal policy under the constraint $m(t) = m$: that exactly m projects should be active at all times. Let $R_{\text{opt}}(m)$ be the optimal average return (from the whole population of n projects) under this constraint.

However, a more relaxed demand would be simply to require that

$$Em(t) = m \tag{4}$$

where the expectation is that for the long-term average: the equilibrium distribution under the policy adopted. Essentially, then, we wish to maximise $E(\sum_i r_i)$ subject to $E(\sum_i 1_i) = n - m$, where r_i is the reward yielded by project i (dependent on project state and phase) and 1_i is 1 or 0 according as to whether project i is in the passive or the active phase. But this is a constraint we could take account of by maximising $E(\sum r_i + \nu \sum 1_i)$, where ν is a Lagrangian multiplier. We are thus effectively solving the modified version of (3):

$$\gamma_i(\nu) + f_i = \max [L_{i1} f_i,\ \nu + L_{i2} f_i] \quad (i = 1, 2, \cdots, n) \tag{5}$$

where f_i is a function $f_i(x_i, \nu)$ of x_i and ν.

An economist would view ν as a 'subsidy for passivity', pitched at just the level (which might be positive or negative) which ensures that m projects are active on average. Note that the subsidy is independent of the project: the constraint (4) is one on *total* activity, not on individual project activity.

A negative subsidy would usually be termed a *tax*. We shall use the term 'subsidy' under all circumstances, however, and shall refer to the policy induced by the optimality equations (5) as the *subsidy policy*. This is a policy optimal under the averaged constraint (4). If we wish to be specific about the value of ν we shall refer to the policy as the *ν-subsidy policy*. For definiteness, we shall close the passive set. That is, if x_i is such that $L_{i1} f_i = \nu + L_{i2} f_i$, then project i is to be rested.

Proposition 1. *The maximal average reward under constraint* (4) *is*

$$R(m) = \inf_{\nu} \left[\sum_i \gamma_i(\nu) - \nu(n - m) \right]. \tag{6}$$

Proof. This is a classic Lagrangian result, but best seen directly. The actual average reward received is indeed $\sum_i \gamma_i(\nu) - \nu(n - m)$, because the average subsidy paid must be subtracted. Since

$$\sum_i \gamma_i(\nu) = \sup_{\pi} E_{\pi} \left[\sum_i (r_i + \nu 1_i) \right] \tag{7}$$

(where π denotes policy) then

$$\frac{\partial}{\partial \nu} \sum_i \gamma_i(\nu) = E_{\pi} \sum 1_i = n - m \tag{8}$$

which relates m and ν, and yields the minimality condition in (6). The condition is indeed one of minimality, because $\sum_i \gamma_i(\nu)$ is convex increasing in ν. The function $R(m)$ is concave, and represents the attainable average reward for all m in $[0, n]$.

Now define the *index* $\nu_i(x_i)$ of project i when in state x_i as the value of ν

which makes $L_{i1}f_i = \nu + L_{i2}f_i$ in (5). In other words, it is the value of subsidy which should make the two phases equally attractive for project i in state x_i.

Proposition 2. *The index $\nu_i(x_i)$ reduces to the Gittins index in the case $P_{i2} = I$, $g_{i2} = 0$ (i.e. when the project is static and yields no reward in the passive phase).*

This is immediate, because our construction of the index is exactly the Gittins construction, transferred to the restless, undiscounted case. The novelty of our view, however, is that the subsidy ν (which Gittins introduced as a 'retirement income') is now seen as the multiplier associated with a constraint on average activity.

We have defined the index only for the undiscounted case. It could also be defined for the discounted case by using the appropriate discounted version of (6). However, constraint (4) would then have to be replaced by

$$E\sum_{0}^{\infty}\beta^t m(t) = \sum_{0}^{\infty}\beta^t m = \frac{m}{1-\beta} \tag{9}$$

where β is the discount factor, and ν chosen to assure this constraint. If the expectation in (9) is one conditional upon initial values $x(0) = \{x_i(0); i = 1, 2, \cdots, n\}$ then ν will also depend upon $x(0)$. The alternative is to introduce a distribution over initial state $x(0)$, so that the expectation in (9) is then over both initial state and subsequent evolution. However, by considering the undiscounted case we avoid these questions.

However, if the index $\nu(x)$ is to be meaningful, it must induce a consistent ordering of the projects, in that any project which is rested under a subsidy ν will also be rested under a subsidy $\nu' > \nu$. Let us formalise this condition.

Definition. Let $D_i(\nu)$ be the set of values of x_i for which project i would be rested under a ν-subsidy policy. Then the project is indexable *if $D_i(\nu)$ increases monotonically from $\varnothing$ to $\mathscr{X}_i$ as ν increases from $-\infty$ to $+\infty$.*

Here $\mathscr{X}_i$ is the full state space for project i.

Proposition 3. *If all projects are indexable, then the projects i which are active under a ν-subsidy policy are those for which $\nu_i(x_i) > \nu$.*

This assertion is an immediate consequence of the definition. The following is less immediate.

Proposition 4. *Projects are always indexable if $P_{2i} = I$ (i.e. if resting projects are static), even in the discounted case. They are not necessarily indexable otherwise.*

We defer the proof to the Appendix. The assertion shows why the question

of indexability simply did not arise in the Gittins case, when resting projects were always assumed static.

Our counterexample of the Appendix shows that indexability cannot be taken for granted. The property seems to be a natural one in the context, however, and we certainly verify it later for some particular cases of interest. One would very much like to have simple sufficient conditions for indexability; at the moment, none are known.

One might now consider the *index policy*: at all times operate the m projects of currently greatest index. This is then a policy which enforces the rigid constraint $m(t) = m$. Let us denote the average return from this policy by $R_{\text{ind}}(m)$.

Proposition 5.

$$R_{\text{ind}}(m) \leqq R_{\text{opt}}(m) \leqq R(m). \tag{10}$$

Proof. The first inequality holds because $R_{\text{opt}}(m)$ is by definition the optimal average return under the constraint $m(t) = m$. The second holds because $R(m)$ is the optimal average return under the *relaxed* constraint $Em(t) = m$.

3. Optimality and asymptotic optimality

In the case when passive projects are both static and rewardless the index $\nu_i(x_i)$ reduces to the Gittins index. It is known that the Gittins index policy is optimal (so that equality holds in the first inequality of (10) if $m = 1$, even in the discounted case). It is known that, even in this static case, the Gittins index policy may be suboptimal for $m > 1$, although probably it is almost optimal, in that one can set a uniform upper bound on the amount of reward lost (see Weiss (1987)).

It may be too much to expect the index policy to be optimal in the restless case we have formulated. However, two points are notable. The first is that we have derived a natural index from the Lagrangian multiplier ν associated with prescription of average utilisation rate: this may well be the natural viewpoint and the natural recipe for deriving such indices. The second point is that this view yields an immediate upper bound on performance: the bound $R(m)$ of (10). Furthermore, it is very plausible that the index policy will give a yield approaching this upper bound under certain conditions.

Suppose that projects are of different *types* $j = 1, 2, \cdots, p$. All projects of the same type are identical in that their g_{ik} and P_{ik} are the same $(k = 1, 2)$, although of course their states x_i are in general distinct. We shall now use $\gamma_j(\nu)$ to denote the value of $\gamma_i(\nu)$ for a project i which is of type j. That is, quantities which are project-dependent only through the type will be distinguished by the

type label rather than the project label. We shall refer to a project of type j as a *j-project.*

Let us consider a population of projects to be a *population of fixed composition* if, as n is allowed to become large, the proportion of j-projects tends to a limit π_j $(j = 1, 2, \cdots, p)$, and the proportion of projects required to be active, m/n, tends to a limit α.

Proposition 6. *For a population of fixed composition $R(m)/n$ tends to the limit*

$$r(\alpha) = \inf_{\nu} \left[\sum_j \pi_j \gamma_j(\nu) - \nu(1-\alpha) \right] \tag{11}$$

as $n \to \infty$, and if $R_{\text{ind}}(m)/n$ has a limit $r_{\text{ind}}(\alpha)$, then

$$r_{\text{ind}}(\alpha) \leqq r(\alpha). \tag{12}$$

The assertions are immediate. It is not necessary that the limit $r_{\text{ind}}(\alpha)$ should exist: from (10) one can make the assertion $\limsup R_{\text{ind}}(m)/m \leqq r(\alpha)$. However, it is plausible that the limit does indeed exist: an increase in n corresponds to a relaxation of policies which in general means that reward rates must be montone increasing with n.

Conjecture. Suppose all projects indexable. Then $r_{\text{ind}}(\alpha)$ exists and equals $r(\alpha)$.

That is, we conjecture that the index policy is optimal in terms of average yield per project in the limit.

The basis of the conjecture is clear. The index policy operates exactly the $m = n\alpha$ projects of largest index. Under the assumption of indexability, the subsidy policy operates the $m(t)$ projects of largest index, where $m(t)/n$ deviates from α only by a term of probable order at most $n^{-\frac{1}{2}}$. It would seem to be easy to formalise this argument to get a definite proof of the conjecture. However, the difficulty is that operation or non-operation of a project affects its dynamics, and it is this which makes comparison of the two cases less than straightforward. It is likely that there is a 'nice' proof of the conjecture which exploits the extremal and Lagrangian aspects of the problem in a natural manner—work continues! We now evaluate the index in some cases of interest.

4. The Ehrenfest project

When we consider single projects we no longer need the labels i or j. Consider a project whose state x can take values $x = 0, 1, 2, \cdots, K$. This project is more easily formulated in continuous rather than discrete time,

which makes no difference. We suppose a reward rate of cx in the active phase and zero in the passive phase. In the active phase the only possible transitions are of the form $x \to x-1$ with intensity μx; in the passive phase the only possible transitions are of the form $x \to x+1$ with intensity $\lambda(K-x)$.

The idea is, for example, that the 'project' may be an individual who has K blood corpuscles, of which x are oxygen-bearing. While working his effectiveness is proportional to the number of oxygen-bearing corpuscles, but these become depleted. While resting he produces nothing, but his corpuscles gain oxygen. The model thus represents in essential form the two phases of tiring and recovery, with a natural limit on state in both directions. The rules are similar to those for the independent movement of atoms between two states in the classic Ehrenfest urn model, whence the name given.

Proposition 7. *The Ehrenfest project is indexable. To within discreteness effects it has index*

$$\nu(x) = \frac{c}{\mu K}(\mu x^2 - \lambda(K-x)^2). \tag{13}$$

The form of the index is particularly simple in the symmetric case $\lambda = \mu$, when

$$\nu(x) = c(2x - K).$$

Proof. Equation (5) becomes

$$\gamma = \max\,(cx - \mu x h(x-1),\ \nu + \lambda(K-x)h(x)) \tag{14}$$

where

$$h(x) = f(x+1) - f(x).$$

Denote the optimal active and passive regions by C and D respectively. We now deduce from (14) that

$$h(x) \sim \begin{cases} \dfrac{cx-\gamma}{\mu x} & (C) \\ \dfrac{\gamma-\nu}{\lambda(K-x)} & (D) \end{cases} \tag{15}$$

where the approximation lies in the fact that we have not distinguished between x and $x-1$ in C. Matching requires that $h(x)$ be continuous on the C/D boundary. Optimality requires that, to within a discreteness effect, $h'(x)$ be continuous on this boundary. Writing down these two conditions we obtain a relation between γ, ν and an arbitrary boundary point x. Eliminating γ we obtain the relation (13) between ν and x. Relation (13) expresses ν in terms of x, and hence gives an expression for the index. Since the expression is

monotone in x, relation (13) will have only a single solution $\bar{x}$ for a given value of ν. The sets C, D are thus intervals, either $x > \bar{x}$, $x \leqq \bar{x}$ respectively or $x \geqq \bar{x}$, $x > \bar{x}$, and we readily verify the former to be the case. Since $\nu(x)$ is monotone increasing in x we can thus characterise C as the set $\nu(x) > \nu$: the project is indexable.

Somewhat more directly, indexability is indicated by the fact that only a single value of ν is compatible with a given boundary point x.

5. The one-dimensional deterministic project

Consider the project in continuous time for which x is a real vector satisfying

$$\dot{x} = a_k(x) \tag{16}$$

with reward rate $g_k(x)$, where $k = 1, 2$ again correspond to the active and passive phases respectively. Then (5) becomes

$$\gamma = \max\left[g_1 + \frac{\partial f}{\partial x}a_1,\ \nu + g_2 + \frac{\partial f}{\partial x}a_2\right] \tag{17}$$

where $\partial f/\partial x$ is the row vector of first derivatives.

Proposition 8. *In the one-dimensional case the deterministic project has index*

$$\nu(x) = g_1 - g_2 + \frac{(a_2 - a_1)(a_2 g_1' - a_1 g_2')}{a_2 a_1' - a_1 a_2'} \tag{18}$$

where all quantities on the right are evaluated at state value x.

Proof. In the one-dimensional case we can solve (17) for $\partial f/\partial x$, obtaining

$$\frac{\partial f}{\partial x} = \begin{cases} \dfrac{\gamma - g_1}{a_1} \\ \dfrac{\gamma - g_2 - \nu}{a_2} \end{cases}$$

in the optimal active and passive regions respectively. Matching and optimality conditions again require that this quantity and its derivative be continuous on the decision boundary. Following the same argument as in the last section we deduce evaluation (18), the primes denoting x-differentiation.

The argument is formal, and conditions are certainly required for the calculations to make sense (e.g. of the depletive/recuperative character obeyed by the Ehrenfest model). However, expression (18) will give the index in the cases when an index exists.

As an example, consider the deterministic version of the Ehrenfest project,

for which $g_1 = cx$, $g_2 = 0$, $a_1 = -\mu x$, $a_2 = \lambda(K - x)$. With these choices index (18) gives exactly the stochastic evaluation (13).

One can extend evaluation (18) to the case when x follows one-dimensional diffusions in the two regions. However, this case does demand consideration of directions of drift, etc., in that one must know how the diffusion behaves at the boundaries of state space. We therefore leave detailed discussion to elsewhere.

6. A general method for the derivation of indices

The methods used in this paper are potentially of much more general application. In general, suppose one has a constraint (or constraints) which restricts selection, as the constraint $m(t) = m$ restricted selection of active projects in this case. One replaces the constraint by an averaged version (e.g. $Em(t) = m$), and allows for this averaged constraint by the introduction of a Lagrangian multiplier ν. This multiplier can then form the basis for definition of a selection index in cases where the original constraint is applied.

Another example of this thinking is provided by Whittle (1984), (1986), where solution of the optimal design problem for fixed routing in a network of queues suggested an adaptive routing.

Appendix: Proof of Proposition 4

Since we are dealing with a single project we can drop the i-subscript. Let us suppose a discount factor β. The dynamic programming equation for the subsidised project then becomes

$$F = \max(g_1 + \beta P_1 F,\ \nu + g_2 + \beta P_2 F) \tag{19}$$

where $F(x, \nu)$ is the maximal expected discounted total reward conditional on a start from state x. Let D be the resting set, for which the maximum is attained by the second expression in the brackets of (19). If

$$\frac{\partial}{\partial \nu}(\nu + g_2 + \beta P_2 F - g_1 - \beta P_1 F) \geqq 0 \tag{20}$$

for x in D then the project will be indexable, for (20) implies that a point of D remains in D as ν increases. Equation (20) can be written

$$1 + \beta P_2 F_\nu \geqq \beta P_1 F_\nu \quad (D). \tag{21}$$

But $F_\nu(x, \nu) = \partial F(x, \nu)/\partial \nu$ is exactly the expected discounted number of visits to D starting from x, under the optimal policy induced by (19). We have

$$F_\nu \leqq \frac{1}{1 - \beta}$$

with equality if $P_2 = I$ and $x \in D$ (for then D is never left). Condition (21) then becomes

$$\frac{1}{1-\beta} \geqq \frac{\beta}{1-\beta}$$

which is certainly valid. The project is thus indexable for all $\beta < 1$ if $P_2 = I$; a limit argument extends the conclusion also to the case $\beta = 1$.

To prove that a project need not be indexable otherwise, let us return to the undiscounted case. Let the state space be $\mathscr{X} = (1, 2, \mathscr{Y})$, where $1, 2$ are a pair of designated states, and $\mathscr{Y}$ the set of remaining states. Consider first returns to the subset $(1, 2)$. Suppose that, if state j and all states of $\mathscr{Y}$ are active then, when starting from state j, one returns to state k with probability p_{jk}, incurs an expected cost before return of a_j, and takes an expected time m_j $(j, k = 1, 2)$. Let the corresponding quantities be q_{jk}, b_j and n_j if state j is passive and all states of $\mathscr{Y}$ are active.

Let us set $p_{21} = \lambda_1$, $p_{12} = \lambda_2$, $q_{21} = \mu_1$, $q_{12} = \mu_2$.

Let γ_D be the average reward for an assigned resting set D. If $D = \varnothing$ (i.e. all states are active) then the equations determining γ and $f(x)$ are

$$m_j\gamma + f(j) = a_j + \sum_K p_{jk} f(k)$$

where j, k take the values $1, 2$. We find that γ then has the evaluation

$$\gamma_0 = \frac{\lambda_1 a_1 + \lambda_2 a_2}{\lambda_1 m_1 + \lambda_2 m_2}.$$

Correspondingly

$$\gamma_1 = \frac{\lambda_1(b_1 + \nu) + \mu_2 a_2}{\lambda_1 n_1 + \mu_2 m_2}$$

$$\gamma_2 = \frac{\mu_1 a_1 + \lambda_2(b_2 + \nu)}{\mu_1 m_1 + \lambda_2 n_2}$$

$$\gamma_{12} = \frac{\mu_1(b_1 + \nu) + \mu_2(b_2 + \nu)}{\mu_1 n_1 + \mu_2 n_2}.$$

We can now choose the rewards g_1 and g_2 and the transition matrices P_1 and P_2 so that the a_j and the b_j can have any values, the m_j and n_j can have any values not less than unity, and the optimal D does not include any state of $\mathscr{Y}$ until ν exceeds an arbitrary specified value $\bar{\nu}$,

Consider the particular choice of values $(\lambda_1, \lambda_2, \mu_1, \mu_2) \propto (1, 2, 1, 1)$, $a_1 = -1$, $a_2 = b_1 = b_2 = 0$, $m_1 = m_2 = n_2 = 1$, $n_1 = 5$. The expressions above then become

$$\gamma_0 = -\tfrac{1}{3}, \quad \gamma_1 = \frac{\nu}{6}, \quad \gamma_2 = \frac{2\nu - 1}{3}, \quad \gamma_{12} = \frac{\nu}{3}.$$

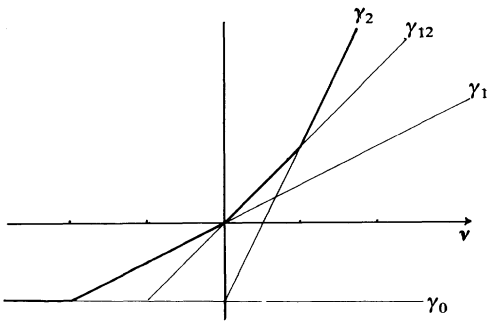

Figure 1. A graph of average reward γ_D for different resting sets D for the counterexample of the Appendix. This demonstrates that the optimal D need not necessarily be monotonic in ν, for as ν increases state 1 first enters the optimal D and then later leaves it.

We can see from the graph in Figure 1 that the optimal resting set is given by

$$D = \begin{cases} \varnothing & \nu < -2 \\ 1 & -2 \leqq \nu < 0 \\ 12 & 0 \leqq \nu < 1 \\ 2 & 1 \leqq \nu < \bar{\nu} \end{cases}.$$

That is, D is not monotonic increasing, for state 1 leaves it as ν increases through 1. The project is thus not indexable.

The reason for the non-monotonicity is the relatively large size of n_1. State 1 is the first to enter D as ν increases. However, the paths starting from 1 with 1 in D show, on average, long excursions from D. This implies a surrender of subsidy which becomes non-optimal once the subsidy becomes large enough.

References

GITTINS, J. C. (1979) Bandit processes and dynamic allocation indices. *J. R. Statist. Soc.* B **41**, 148–164.

GITTINS, J. C. AND JONES, D. M. (1974) A dynamic allocation index for the sequential design of experiments. In *Progress in Statistics* ed. J. Gani, North-Holland, Amsterdam, 241–266.

WEISS, G. (1987) Approximation in results in parallel machines stochastic scheduling. Presented at the Twelfth Symposium on Operations Research, Passau.

WHITTLE, P. (1980) Multi-armed bandits and the Gittins index. *J. R. Statist. Soc.* B **42**, 142–149.

WHITTLE, P. (1981) Arm-acquiring bandits. *Ann. Prob.* **9**, 284–292.

WHITTLE, P. (1984) Optimal routing in Jackson networks. *Asia-Pacific J. Operat. Res.* **1**, 32–37.

WHITTLE, P. (1986) *Systems in Stochastic Equilibrium.* Wiley, Chichester.