

# A Bisection/Successive Approximation Method for Computing Gittins Indices<sup>1</sup>

By A. Ben-Israel<sup>2</sup> and S. D. Flåm<sup>3</sup>

*Abstract:* An iterative method, combining bisections and successive approximations, is proposed for computing intervals containing the Gittins indices. The intervals could be of a specified maximum length, or be merely disjoint. The first option gives approximations of the Gittins indices. The second option gives a ranking of indices, which in many applications is sufficient.

*Zusammenfassung:* Es wird ein iteratives Verfahren von Schranken von Gittins-Indizes vorgestellt. Der Algorithmus besteht aus Bisektion und einigen Schritten sukzessiver Approximation.

*Keywords:* Gittins indices, bisection, successive approximation.

## 1 Introduction

The terminology and notation of [14] are used throughout. Consider a “single project bandit process” with states  $i = 1, \dots, N$ , (bounded) rewards  $R(i)$ , transition probabilities  $P_{ik}$ , and discount factor  $\alpha$ ,  $0 \leq \alpha < 1$ . At each time  $t = 0, 1, \dots$  the decision (after observing the current state, say  $i$ ) is to retire and collect the terminal reward  $M$ , or continue, obtain  $R(i)$  and move to one of the states  $j = 1, \dots, N$  with probabilities  $P_{ij}$ .

The optimal expected  $\alpha$ -discounted return with initial state  $i$ ,  $V(i : M)$ , then satisfies the optimality equation

$$V(i : M) = \max \left\{ M, R(i) + \alpha \sum_k P_{ik} V(k : M) \right\} \quad (1)$$

The Gittins index ([6], [7], [8], [18]) of state  $i$ ,  $M(i)$ , is

---

<sup>1</sup> Supported by the Royal Norwegian Council for Industrial and Scientific Research and by the National Science Foundation.

<sup>2</sup> RUTCOR-Rutgers Center of Operations Research and School of Business, Rutgers University, New Brunswick, NJ 08903, USA.

<sup>3</sup> Institute of Economics, University of Bergen, N-5000 Bergen, Norway.

$$M(i) = \min\{M : M = V(i : M)\} \quad (2)$$

Gittins indices are useful in “multi-project bandit problems”. In typical applications several projects are given with possibly different state spaces, rewards and transition probabilities (see the example below on exploration for resource deposits). In any event, the actual states of the projects may differ, project  $j$  being in state  $i_j$ . The problem, at any stage, is to decide whether further operations will bring any benefits. If so, one should be directed to work, and be told for how long. The nature of the best solution is remarkably simple and appealing: *Continue* if some uncompleted project is still worth our while, i.e. its value exceeds our reservation price/wage given a priori and exogeneously. In that case work (first and exclusively) on the project which currently enjoys the highest Gittins index. In particular, switch project when (at possibly random times) the index of the engaged project drops below that of an inactive one. *Retire* as soon as all projects are found in states with indices dominated by the possibly updated reservation price.

Because of the assumption that projects are unchanged while they are not worked on, the Gittins indices are particularly relevant to problems of oil exploration ([4], [5]). Here is one example in this vein, see Weitzman [16]. A company owns an oil deposit with a certain net value  $v$ . It is offered one (and only one) substitute among several unexplored deposits. The value  $X_d$  of deposit  $d$  is uncertain ex ante, but ex post, after exploring during random time  $T_d$ , incurring thereby an instantaneous cost  $c_d$ , it is perfectly revealed and available. All random variables are assumed to be independent. There is no learning. Here the Gittins index  $M_d$  of deposit  $d$ , when still unexplored, satisfies the equation

$$M_d = -c_d + E\{\alpha^{T_d}\}E\max\{M_d, X_d\}$$

which can be solved numerically. If some index exceeds  $v$ , the company should continue exploring the site with highest index, and stop when all unexplored fields have indices below the best value detected so far.

Computational methods for the Gittins indices have been given by Chen and Katehakis [2], Gittins and Jones [7], Kallenberg [9], Katehakis and Veinott [10], Kramp [11] and Varaiya, Walrand and Buyukkoc [15]. Theoretical results have been extended to so called open bandit problems by Whittle [19] and Lai and Ying [12]. Other extensions of the theory to more general stochastic processes are found in Varaiya et al [15] and Mandelbaum [13].

In this note we use the observation that in applications it often suffices to have a ranking of the Gittins indices, and their exact values are not needed (except possibly for few states).

In § 2 we give bounds on Gittins indices which are based on successive approximations, and upper bounds on the number of successive approximations needed in such schemes.

The algorithm described in § 3 combines bisections (i.e. selection of midpoint  $M$  of an interval known to contain a Gittins index), and a finite number of successive approximations for each such  $M$ . Numerical results are reported.

## 2 Bounds for Gittins Indices

The maximal expected returns  $V(i : M)$  of (1) can be computed by successive approximations,  $i = 1, \dots, N$

$$\begin{aligned}
 V_0(i : M) &= M, \quad i = 1, \dots, N \\
 V_n(i : M) &= \max \left\{ M, R(i) + \alpha \sum_k P_{ik} V_{n-1}(k : M) \right\}, \quad i = 1, \dots, N
 \end{aligned} \tag{3}$$

where  $V_n(i : M)$  is the maximal expected  $\alpha$ -discounted return with initial state  $i$  and at most  $n$  stages before retiring. Indeed, by (3), for all  $n$ ,

$$V_{n+1}(i : M) \geq V_n(i : M) \quad \forall i, M \tag{4}$$

and

$$V(i : M) = \lim_{n \rightarrow \infty} V_n(i : M) \tag{5}$$

For given  $i, M$  (and not knowing  $M(i)$ ) we study conditions under which the terms

$$R(i) + \alpha \sum_k P_{ik} V_{n-1}(k : M) \tag{6}$$

in (3) allow us to decide whether

$$M < M(i) \tag{7}$$

Since  $\frac{\max_k R(k)}{1-\alpha}$  is an upper bound for all indices  $M(j), j = 1, \dots, N^4$  (see also (26)

---

<sup>4</sup> Recall that the capital value of a steady flow of 1\$ in each and every time period is

$$\frac{1}{1-\alpha} = \sum_{t=0}^{\infty} \alpha^t$$

Any retirement reward greater than  $\max_k R(k) / 1 - \alpha$  would therefore be preferred to continuation.

below), we assume

$$M < \frac{\max_k R(k)}{1 - \alpha} \quad (8)$$

From (2) and the fact that  $V(i : M) - M$  is a decreasing function of  $M$  ([14], Lemma 2.1, p. 132) it follows that (7) is equivalent to

$$V(i : M) > M \quad (9)$$

which in turn is equivalent to

$$V_n(i : M) > M, \quad \text{for some } n, \quad (10)$$

or to

$$R(i) + \alpha \sum_k P_{ik} V_{n-1}(k : M) > M, \quad \text{for some } n. \quad (11)$$

We give now an estimate for the  $n$  in (11).

*Lemma 1:* Given  $i, M$  satisfying (7), the inequality (11) holds for all  $n$  greater than or equal to

$$n_1 = \frac{\log \left( \frac{V(i : M) - M}{\frac{\max_k R(k)}{1 - \alpha} - M} \right)}{\log \alpha} \quad (12)$$

*Proof:* We note that (10) and (11) hold for the same  $n$ , and rewrite (10) as

$$\Delta_n(i : M) < V(i : M) - M, \quad \text{for some } n, \quad (13)$$

where

$$\Delta_n(i : M) = V(i : M) - V_n(i : M). \quad (14)$$

To estimate  $\Delta_n(i : M)$  we write

$$V(i : M) = R(i) + \alpha \sum_k P_{ik} V(k : M), \quad \text{by (9),}$$

and

$$V_n(i : M) \geq R(i) + \alpha \sum_k P_{ik} V_{n-1}(k : M), \text{ by definition.}$$

Therefore, for all  $n = 1, \dots$

$$\begin{aligned} \Delta_n(i : M) &\leq \alpha \sum_k P_{ik} \Delta_{n-1}(k : M) \\ &\leq \alpha \max_k \Delta_{n-1}(k : M) \\ &\leq \alpha^n \max_k \Delta_0(k : M) \\ &= \alpha^n \max_k \{V(k : M) - M\} \end{aligned} \quad (15)$$

We estimate now the last term in (15)

$$\begin{aligned} V(i : M) - M &= \max \left\{ 0, R(i) - (1 - \alpha)M + \alpha \sum_k P_{ik} [V(k : M) - M] \right\} \\ &\leq \max \left\{ 0, \max_k R(k) - (1 - \alpha)M + \alpha \max_k [V(k : M) - M] \right\} \end{aligned} \quad (16)$$

In particular,

$$\max_k \{V(k : M) - M\} \leq \text{RHS (16)}$$

and therefore, by (8),

$$\max_k [V(k : M) - M] \leq \frac{\max_k R(k)}{1 - \alpha} - M \quad (17)$$

which, together with (15), gives

$$\Delta_n(i : M) \leq \alpha^n \left( \frac{\max_k R(k)}{1 - \alpha} - M \right) \quad (18)$$

Thus (13) holds for all  $n$  such that

$$\alpha^n \left( \frac{\max_k R(k)}{1 - \alpha} - M \right) \leq V(i : M) - M$$

giving the estimate  $n_1$  of (12).  $\square$

The operational meaning of Lemma 1 is that, for any  $M$  (satisfying (8)) and  $i$ , at most  $n_1$  successive approximations are needed to decide whether  $M(i) > M$ . Since the decision is independent of information obtained later,  $n_1$  is a so called planning horizon, e.g. [1]. The computation of  $n_1$  in Lemma 1 is similar to the computation of the planning horizon  $n_0$  in [3, Theorem 9, p. 176].

In actual computations,  $n_1$  turns out to be rather small. In hundreds of random problems, with  $\alpha = 0.9$ , the average  $n_1$  was 10.3 for  $N = 10$  and 10.5 for  $N = 20$ .

A (crude) upper bound on  $n_1$  is

$$n_2 = \frac{\log \left( \frac{\frac{\varepsilon}{\max_k R(k)} - M}{1 - \alpha} \right)}{\log \alpha} \quad (19)$$

where  $\varepsilon > 0$  is a lower bound on the (unknown) quality  $V(i : M) - M$  in (12). For example, if we denote  $A = \log \left( \frac{\max_k R(k)}{1 - \alpha} - M \right)$ , and if the tolerance  $\varepsilon = 10^{-2}$ , then

$$n_2 \approx \begin{cases} 6.6 + 3.3A & \text{if } \alpha = .5 \\ 12.9 + 6.4A & \text{if } \alpha = .7 \\ 43.7 + 21.8A & \text{if } \alpha = .9 \end{cases}$$

Using Lemma 1 we may decide, after  $n_1$  successive approximations in which (11) has not occurred, that

$$M \geq M(i) \quad (20)$$

We show now that (20) may be concluded sooner, if for any  $n$  the term (6) drops sufficiently below  $M$ .

*Lemma 2:* If for some  $n = 1, 2, \dots$

$$R(i) + \alpha \sum_k P_{ik} V_{n-1}(k : M) \leq M - \alpha^n \left( \frac{\max_k R(k)}{1 - \alpha} - M \right) \quad (21)$$

then  $M \geq M(i)$ .

*Proof:* Suppose  $M < M(i)$ , so that

$$V(i : M) = R(i) + \alpha \sum_k P_{ik} V(k : M) > M$$

Then by subtracting (21),

$$\alpha \sum_k P_{ik} \Delta_{n-1}(k : M) > \alpha^n \left( \frac{\max_k R(k)}{1 - \alpha} - M \right) \quad (22)$$

But by (15) and (17),

$$\begin{aligned} \text{LHS (22)} &\leq \alpha \max_k \Delta_{n-1}(k : M) \leq \alpha^n \max_k \Delta_0(k : M) \\ &\leq \text{RHS (22)}, \text{ a contradiction. } \quad \square \end{aligned}$$

In broad terms, Lemma 2 says the following: Suppose operating some project at least  $n$  periods yields a loss (compared to the readily available  $M$ ) that cannot be recovered in the future. Then the project should have been abandoned immediately. Thus (21) is a test for suboptimality designed to eliminate bad strategies, see [17].

### 3 The Algorithm

If in Lemma 1 and 2, the point  $M$  lies in an interval  $[L(i), R(i)]$ , then it is possible to obtain a smaller interval, (also containing  $M(i)$ ), with  $M$  as one of its endpoints.

This suggests using bisection, i.e. taking  $M$  to be the midpoint.

$$M = \frac{1}{2}(L(i) + U(i)) \quad (23)$$

and successive approximations to compute  $V_n(k : M)$  as needed.

Intervals  $\{[L(i), U(i)] : i = 1, \dots, N\}$ , containing the Gittins indices

$$L(i) \leq M(i) \leq U(i) \quad (24)$$

are given initially by

$$L(i) = R(i) + \frac{\alpha}{1 - \alpha} \min_j R(j) \quad (25)$$

$$U(i) = R(i) + \frac{\alpha}{1 - \alpha} \max_j R(i) \quad (26)$$

The lower bound (25) corresponds to moving from state  $i$  (after collecting  $R(i)$ ) to the “lowest paying state” and staying there. The higher bound (26) is similarly associated with the “highest paying state”.

Since  $V(i : M) > M$ , we can write

$$V(i : M) \geq \max \left\{ M, R(i) + \alpha \sum_k P_{ik} V(k : M) \right\} \geq \max \{ M, R(i) + \alpha M \}$$

Thus, if  $L(i)$ , as defined in (25), does not already majorize  $R(i) + \alpha L(i)$ , it should be replaced by  $R(i) / (1 - \alpha)$ .

An iteration follows (given initial intervals  $[L(i), U(i)]$  and their midpoints).

1.  $n \leftarrow 0$   
 $V(j : M) \leftarrow M \quad (j = 1, \dots, N)$
  2. for  $j = 1, \dots, N$ 

$$\text{SUM}(j) \leftarrow R(j) + \alpha \sum_k P_{jk} V(k : M)$$

**if**  $\text{SUM}(j) > M$  **then**  $L(j) \leftarrow \max \{L(j), \text{SUM}(j)\}$

**if**  $\text{SUM}(j) \leq M - \alpha^n \left( \frac{\max_k R(k)}{1 - \alpha} \right)$  **then**

$U(j) \leftarrow \min \{U(j), M\}$
  3. **if** approximation criterion = ‘FALSE’ **then** update the values  
 $n \leftarrow n + 1$   
 $V(j : M) \leftarrow \max \{M, \text{SUM}(j)\}, (j = 1, \dots, N)$   
**go to 2.**
  4. **if** stopping criterion = ‘FALSE’ **then** bisect  
**select**  $i$   
 $M \leftarrow \frac{1}{2}(L(i) + U(i))$   
**go to 1.**
- else stop.**

An iteration (steps 1 through 4) contains a number of successive approximations (step 2), a number which is determined by the approximation criterion used in step 3. The stopping criterion, used in step 4, determines the number of iterations.

A possible approximation criterion is, when the selected interval has been bisected, which by Lemma 1 requires at most  $n_1$  approximations.

A successive approximation requires  $N^2$  operations, corresponding to the matrix-vector multiplication in (6). Therefore the total computational effort is about  $N^2$  times the number of successive approximations. Numerical experiments with randomly generated problems show that the average total number of successive approximations is minimized if the number of successive approximations per iteration is bounded a priori, say by  $n_3$ . The accuracy is compromised if  $n_3$  is taken too small. For problems of sizes  $N = 10$  to  $20$ , our experiments indicate  $n_3$  between 5 and 10.

Possible stopping criteria, in terms of the intervals

$$\{[L(i), U(i)] : i = 1, \dots, N\}$$

are

$$\text{the intervals are disjoint,} \tag{27}$$

or

$$U(i) - L(i) < \delta, \forall i, \tag{28}$$

where  $\delta > 0$  is a specified tolerance, or

$$U(i) - L(i) < \delta, \text{ for some } i. \tag{29}$$

The stopping criterion (27) is appropriate for ranking the Gittins indices, while criteria (28) or (29) serve to approximate all, or some, indices.

The selection of  $i$  in step 4 is dictated by the stopping criterion. For example, under criterion (28) select  $i$  with largest  $U(i) - L(i)$ .

We observe that, unless the intervals  $[L(i), U(i)]$  are already disjoint, it is possible for several of them to shrink in step 2, in addition to the interval being bisected.

The operations count per iteration (with at most  $n_1$  repetitions of step 2) is  $O(N^2)$ . With stopping criterion (29), the number of iterations is independent of  $N$ : To approximate with tolerance  $\delta$  one Gittins index  $M(i)$ , at most

$$\frac{\log \frac{U(i) - L(i)}{\delta}}{\log 2} \tag{30}$$

midpoints  $M$  need be considered, where  $[L(i), U(i)]$  is the initial interval. For

criterion (28) multiply (30) by  $N$ , showing this to be an  $O(N^3)$  algorithm for approximating all the Gittins indices<sup>5</sup>.

With criterion (27), i.e. when only ranking of indices is required, a much smaller number of bisections is needed.

The following tables report numerical tests for  $N = 10$  and tolerance  $\epsilon = 0.01$ . One hundred random problems were solved for each of the values  $\alpha = 0.7, 0.8, 0.9$ , using the three stopping criteria given above, with values  $\delta = 0.1, 0.5, 1.0$  in criteria (28) and (29). In each case, the number of successive approximation per iteration has been bounded by  $n_3 = 5$ . The three numbers in each cell are

average  
maximum (encountered)  
standard deviation

of the number of successive approximations per problem. An estimate of the average operations count is  $N^2$  (here 100) times the average number of successive approximations.

For criterion (27), ranking of indices, the results are:

$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
56.2	55.0	57.5
88	88	82
8.4	9.2	7.9

showing insensitivity to  $\alpha$ . However, for  $\alpha = 0.9$  the number of successive approximations in step 2 tended to equal its a priori bound  $n_3 = 5$ , giving a smaller variance.

For criterion (28), approximating all indices, the results are:

	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
$\delta = 0.1$	250.6	268.2	304.6
	274	311	350
	10.7	15.3	9.4
$\delta = 0.5$	151.1	173.8	216.9
	183	192	239
	15.4	7.3	7.6
$\delta = 1.0$	109.3	131.1	173.0
	137	146	193
	13.9	6.2	7.1

And finally, for criterion (29), approximating a single index:

---

<sup>5</sup> Not so interesting in applications where  $N$  is typically small.

	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
$\delta = 0.1$	34.6	39.0	43.7
	40	45	50
	6.9	7.8	8.9
$\delta = 0.5$	26.3	30.7	35.4
	30	35	40
	4.7	5.6	6.3
$\delta = 1.0$	22.0	26.5	31.2
	25	30	35
	3.6	4.5	5.2

## References

- [1] Bean JC, Smith RL (1984) "Conditions for the existence of planning horizons", *Math of OR* 9: 391–401
  - [2] Chen YR, Katehakis MN (1986) "Linear programming for the finite state multi-armed bandit problems", *Math of OR* 11: 180–183
  - [3] Denardo E (1982) *Dynamic Programming*, Prentice Hall
  - [4] Flåm SD, Stensland G (1985) "Exploration and taxation: Some normative issues" *Energy Economics*: 234–240
  - [5] Flåm SD, Olsen TE (1985) "Scheduling and taxation of resource deposits" *The Energy Journal* 6: 137–143
  - [6] Gittins JC (1980) "Bandit processes and dynamic allocation indices" *J Roy Statist Soc Ser B* 42: 143–149
  - [7] Gittins JC, Jones DM (1974) "A dynamic allocation index for the sequential design of experiments" in J Gani (editor) *Progress in Statistics*, North Holland, Amsterdam
  - [8] Gittins JC, Glazerbrook KD (1977) "On Bayesian models in stochastic scheduling", *J Appl Prob* 14: 556–565
  - [9] Kallenberg LCM (1986) "A note on M.N. Katehakis' and Y.R. Chen's computation of the Gittins index", *Math of OR* 11: 184–186
  - [10] Katehakis MN, Veinott AF Jr (1987) "The multi-armed bandit problem: Decomposition and computation", *Math of OR* 12: 262–268
  - [11] Kramp M (May 1987) *Models of Switching: Multi-Purpose Fleets in Fisheries, Multi-Armed Bandits and Gittins Indices*, Ph D dissertation in OR, Department of Mathematical Sciences, University of Delaware
  - [12] Lai TL, Ying Z (1988) "Open bandit processes and optimal scheduling of queueing networks" *Adv Appl Prob* 20: 447–472
  - [13] Mandelbaum A (1986) "Discrete multi-armed bandits and multi-parameter processes", *Prob Theory Rel Fields* 71: 129–147
  - [14] Ross SM (1983) *Introduction to Stochastic Dynamic Programming*, Academic Press, New York
  - [15] Varaiya P, Warland C, Buyukkoc C (1985) "Extensions of the multi-armed bandit problem: The discounted case", *IEEE Trans Autom Control* AC-30: 426–439
  - [16] Weitzmann ML (1979) "Optimal search for the best alternative", *Econometria* 47: 641–654
-

- [17] White DJ (1987) "Elimination of non-optimal actions in Markov decisions processes" in Puterman (Editor), *Dynamic Programming and its Applications*, Academic Press, New York: 131-160
- [18] Whittle P (1980) "Multi-armed bandits and the Gittins index", *J Roy Statist Soc B42*: 143-149
- [19] Whittle P (1981) "Arm acquiring bandits", *Ann Prob 9*: 284-292
- [20] Whittle P (1982) *Optimization over time*, Vol I Chapter 14, J. Wiley, New York

Received June 1987

Revised version received November 1989