SYSTEMATIC SEARCH, BELATED INFORMATION, AND THE GITTINS' INDEX

Brian P. McCall
and
John J. McCall
UCLA

UCLA Department of Economics
Working Paper #214
August 1981

1. Introduction

Roberts and Weitzman have shown that the recent contributions to the theory of multi-armed bandit (MAB) processes have important applications in economics. This paper uses MAB methods to obtain a simple solution to a rather complicated problem in search theory. The problem is one in which the job searcher seeks employment at one of N different firms. The searcher possesses subjective beliefs about the wages attainable at each firm. Furthermore, each job is characterized by belated information in that some non-pecuniary aspect of the job is revealed only after the job has been tested for one period. At each decision point, the optimal action among the N alternatives is the one possessing the largest Gittins index, where this index is the MAB generalization of the reservation wage.*

The action associated with each index is either search, test, or work. Work is an absorbing state and may commence immediately when the searcher chooses to work a previously tested job or belatedly when a test reveals a favorable non-pecuniary aspect. The ordered set of N Gittins indices is revised as new information is accumulated by the systematic, belated search (with recall).

Section 2 briefly describes MAB processes and the Gittins index, which is the fundamental technique for solving these problems. In Section 3 a simple model of belated information is posed as a MAB problem. With this characterization the Gittins index and the associated optimal policy are easily calculated. Section 4 extends the belated information model to include systematic search within the MAB framework. The main result of the paper is obtained by specifying the set of Gittins indices

---

*See Roberts and Weitzman.

and thus the optimal solution to this problem. The concluding section contains some suggestions for future research.

2. MAB Processes and the Gittins Index*

A gambler is confronted with N slot machines (one armed bandits) that may be played repeatedly in any order. The i^{th} machine is characterized by a success probability θ_i which is unknown to the gambler. The Bernoulli sequence of successes and failures on the i^{th} machine provides information used to obtain a Bayesian estimate of θ_i. The object is to choose the sequence of plays to maximize the total discounted expected reward.**

During the past few years Gittins and his colleagues have solved this MAB problem. Simply stated the optimal policy has the following form: a policy is optimal if and only if at each decision point the selected bandit process is the one with the largest Gittins index.

More concretely, suppose there are N projects indexed by $n = 1,2,\ldots,N$.*** At any decision point exactly one project must be selected for further development. Suppose project n is in state i. If n is chosen a reward R_i^n is collected and project n moves from state i to state j with probability P_{ij}^n. The states of all other projects remain unchanged. The discount

---

*For the complete story see the papers by Gittins and Whittle. A scholarly discussion of the 2 armed bandit is presented in DeGroot.

**To our knowledge the first successful solution to the problem of balancing present expected gains against the expected return from additional information is contained in Fox.

***The notation follows Roberts and Weitzman. In this important paper Roberts and Weitzman demonstrate the intimate relation between MAB processes and search theory with Weitzman's box search providing the original inspiration.

factor is β_i^n. Let $\psi(s)$ denote the maximal expected discounted return when the current state is $s = \overset{N}{\underset{n=1}{X}} i(n)$. Then $\psi(s)$ satisfies

$$(1) \quad \psi(s) = \max_{1 \leq n \leq N} \{R_i^n + \beta_i^n \; P_{ij}^n \; \psi(s - [i(n)] + [j])\}$$

where $s - [i(n)] + [j]$ means state s with n in state j instead of i(n).

The solution to this problem is obtained by assigning each project a Gittens index and then choosing each period that project with the largest index. Intuitively, the Gittins index is the value of a fallback position which would make a decision maker indifferent between continuing with project n in state i and rejecting the project for the fallback position. For any fallback position Z the maximum expected discounted return $V_i^n(Z)$ obtainable given the choice of either continuing with project n (in state i) or stopping and getting Z satisfies

$$(2) \quad V_i^n(Z) = \max\{Z; R_i^n + \beta_i^n \; \Sigma P_{ij} \; V_j^n(Z)\}$$

Thus the Gittins index Z_i^n is given by

$$(3) \quad Z_i^n = V_i^n(Z_i^n) = R_i^n + \beta_i \; \Sigma P_{ij}^n \; V_j(Z_i^n)$$

<u>Proposition 1</u> (Roberts and Weitzman) Let $Z_{i(n)}^{n*} = \underset{n}{\text{Max}} \; Z_{i(n)}^n$. Then the policy that chooses n* when $s = \overset{N}{\underset{n=1}{X}} i(n)$ is optimal.

3. A Bandit Model of Belated Information*

There are N jobs available and their wages w^n, $n = 1,2,\ldots,N$, are known. However, each job's non-pecuniary aspect $\alpha^n V$, $\alpha^n > 0$ for all n, is observable only after the job has been "tested" for one period. For simplicity we assume $P(V = +1) = P(V = -1) = \frac{1}{2}$. The objective is to design a testing policy that maximizes expected discounted returns, where β is the discount factor.

This problem can be formulated as a bandit process with

$$P^n_{01} = P^n_{02} = \frac{1}{2}; \quad R^n_0 = w^n; \quad \beta^n_0 = \beta,$$

where state 0 is pretest state.

After the job is tested for one period the non-pecuniary aspect is either favorable (state 1) or unfavorable (state 2). Hence,

$$P^n_{22} = P^n_{11} = 1; \quad R^n_1 = \frac{w^n + \alpha^n}{1-\beta}; \quad R^n_2 = \frac{w^n - \alpha^n}{1-\beta}; \quad \beta^n_i = 0, \ i = 1,2.$$

The equation defining the Gittins index for the n^{th} job in state 0 is

$$(4) \qquad Z^n_0 = w^n + \frac{1}{2}\beta\{\mathrm{Max}(Z^n_o, \frac{w^n+\alpha^n}{1-\beta}) + \mathrm{Max}(Z^n_o, \frac{w^n+\alpha^n}{1-\beta})\}.$$

The optimal policy tests that job with the largest Gittins index and continues testing until the total remuneration of a tested job exceeds the Gittins indices of all untested jobs.

*The version described here is a special case of the model studied in Lippman and McCall (1981a).

Proposition 2

The Gittins index is given by

$$Z_o^n = \frac{w^n + \frac{1}{2}\beta(w^n+\alpha^n)/(1-\beta)}{1 - \frac{1}{2}\beta}. \tag{5}$$

Proof:

From (4) if $Z_o^n \geq \frac{w^n+\alpha^n}{1-\beta}$, $Z_o^n = w^n + \beta Z_o^n$, a contradiction since $\alpha^n > 0$. If $Z_o^n \leq \frac{w^n-\alpha^n}{1-\beta}$, $Z_o^n = \frac{w^n}{1-\beta}$, a contradiction. Thus $\frac{w^n-\alpha^n}{1-\beta} < Z_o^n < \frac{w^n+\alpha^n}{1-\beta}$, which implies

$$Z_o^n = w^n + \frac{1}{2}\beta\left(\frac{w^n+\alpha^n}{1-\beta} + Z_o^n\right).$$

Q.E.D.

4. Optimal Systematic Search With Belated Information*

The model is the same as in Section 3 except that the wage w_j^n at job n is a random variable with subjective probability distribution q_j^n, $j = 1,2,\ldots,J$. The searcher must pay a cost c to observe the true wage. Job n is in state 0 before it has been searched, state j after wage w_j has been observed and, finally, either state 2j ($+\alpha^n$) or state 3j ($-\alpha^n$) after the job has been tested. Thus

$$P_{oj}^n = q_j^n;\ P_{j2j} = P_{j3j} = \frac{1}{2};\ R_o^n = -c;\ R_j^n = w_j; R_{2j}^n = \frac{w_j+\alpha^n}{1-\beta};\ R_{3j}^n = \frac{w_j-\alpha^n}{1-\beta};$$

$$\beta_j^n = \begin{cases}\beta, & 0 \leq j \leq J \\ 0, & j > J.\end{cases}$$

The Gittins index for the n^{th} job in state 0 satisfies

*The first systematic search model was designed by Salop.

$$(6) \quad Z_o^n = -c + \beta\Sigma_{j=1}^J q_j^n \max\{Z_o^n, w_j^n + \frac{1}{2}\beta[\max(Z_o^n, \frac{w_j^n+\alpha^n}{1-\beta}) + \text{Max}(Z_o^n, \frac{w_j^n-\alpha^n}{1-\beta})]\}$$

From (6) it is clear that Z_o^n is non decreasing in α^n, i.e., more uncertainty is preferred to less.

If job k is searched, i.e., $Z_o^k \geq Z_{i(n)}^n$ for all n, then the job will be tested if and only if $Z_j^k \geq Z_k^*$ where $Z_k^* = \max_{n\neq k} Z_{i(n)}^n$. This observation gives

<u>Proposition 3</u>. If the k^{th} job is searched, there is a reservation wage ξ^k for the k^{th} job such that the k^{th} job will be tested if and only if $w_j^k \geq \xi^k$ where ξ^k is defined by

$$(7) \quad \xi^k = (1-\beta)\, Z_k^* - \frac{\frac{1}{2}\beta\alpha^k}{1-\frac{1}{2}\beta}$$

Proof: From (5),

$$Z_j^k = \frac{w_j^k + \frac{1}{2}\beta(w_j^k + \alpha^k)/(1-\beta)}{1-\frac{1}{2}\beta}$$

Consequently, the job is tested if and only if

$$Z_k^* \leq \frac{w_j^k + \frac{1}{2}\beta(w_j^k + \alpha^k)/(1-\beta)}{1-\frac{1}{2}\beta}$$

which means that ξ^k satisfies

$$Z_k^* = \frac{\xi^k + \frac{1}{2}\beta(\xi^k + \alpha^k)/(1-\beta)}{1-\frac{1}{2}\beta}$$

Q.E.D.

Notice that the set of Gittins indices is revised after each observation. For example, if $w_j^k < \xi^k$, Z_o^k is replaced by $E(\frac{w_j^k+\alpha^k v}{1-\beta})$ and ranked accordingly;

if $w_j^k \geq \xi^k$ and $V = -1$, then Z_j^k is replaced by $\frac{w_j^k - \alpha^k}{1-\beta}$ and ranked accordingly. Once a job is searched the revised set of Gittins indices allows the prospective employee to choose the best of his four options: (1) Test the job; (2) Continue search; (3) Return to work at previously tested job and (4) Test a previously searched job. If search continues to generate unacceptable offers, jobs that were initially lowly ranked become increasingly attractive. Thus Z_j^k decreases and the set of reservation wages $\{\xi^k\}$ declines over time

A complete description of the optimal policy provides a convenient summary of this section.

If the k^{th} job has the largest Gittins index, search it (if it has not been searched), or test it (if it has been searched but not tested), or work it (if it has been searched and tested).

After searching, if $w_j^k \geq \xi^k$, test k^{th} job; if $w_j^k < \xi^k$, search, test, or work that job with the largest Gittins index.

After testing the k^{th} job, work it if it has the largest Gittins index; otherwise search, test, or work the job that does have the largest index.

5. Conclusion

The ease with which the MAB framework allowed us to extend* the belated information model suggests that it might be a friendly environment in which to pose more difficult problems. Clearly, some employers (universities)

---

*Though we restricted our search to a finite number N of distinct jobs where Lippman and McCall and most search models consider an infinite number of independent, identically distributed jobs.

also engage in systematic search with belated information. Thus it may be possible to develop an equilibrium model where both employees and employers search systematically and receive belated information.* It also seems that this is the appropriate framework for studying migration, occupational choice and mobility in hierarchical organizations.

---

*This would extend the equilibrium model of Lippman and McCall (1981b).

REFERENCES

DeGroot, M.H., Optimal Statistical Decision, (New York: McGraw Hill, 1970).

Fox, B., An Adaptive Age Replacement Policy, (Berkeley: Operations Research Center, 1965).

Gittins, J.C., "Bandit Processes and Dynamic Allocation Indices," J. R. Statist. Soc., B 41 (1979), 148-177.

Lippman, S.A. and J.J. McCall, "The Economics of Belated Information," Inter. Econ. Rev., 22 (1981a), 135-146.

__________, "An Equilibrium Model of Turnover with Belated Information," UCLA, 1981b.

Roberts, K.S. and M.L. Weitzman,"On a General Approach to Search and Information Gathering," (1980) working paper, MIT.

Salop, S.C., "Systematic Job Search and Unemployment," Rev. of Econ. Stud., 40 (1973), 191-201.

Weitzman, M.L., "Optimal Search for the Best Alternative," Econometrica 47 (1979), 641-654.

Whittle, P., "Multi-armed Bandits and the Gittins Index," J. R. Statist. Soc. V 42 (1980), 143-149.