Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

SOME DESIGNS FOR MULTICRITERIA BANDITS

BY P. W. JONES, A. M. LEWIS AND R. HARTLEY

Keele University

Abstract

The Bernoulli two-armed bandit with finite horizon and independent beta priors is considered from a multicriteria perspective. Several sequential designs are suggested and their characteristics are derived.

1. Introduction. Consider the Bernoulli two armed bandit with probabilities p_1, p_2 of obtaining a reward and $(1-p_1)$, $(1-p_2)$ of obtaining nothing when pulling arms 1 and 2, respectively. It is assumed that the maximum number of pulls is N. The original two armed bandit problem concentrated on designs which maximize the expected return $E(R)$ [see, for example, Jones (1975)], this paper deals with two additional criteria, $P(CS)$, which is the probability of correctly selecting the superior arm at the termination of the process, and $E(N_{(1)})$, which is the expected number of pulls on the poorer arm. In the context of clinical trials where the arms are treatments and the number of patients is N , these additional criteria have ethical and statistical importance and need to be taken into account in evaluating designs. There are two ways of looking at this problem in this context, either as an optimization problem for this particular set of N patients, or as a decision problem where a recommendation on the superior treatment will be made for future patients. The two criteria, $E(R)$ and $E(N_{(1)})$ are of the first type and ethical considerations suggest that these should be

Received October 1992; revised April 1993.
AMS 1991 subject classification. Primary 62L05; secondary 62L15.
Key words and phrases. Backward induction, Bayesian methods, group sequential methods, sequential designs, stopping problems.

optimized; and $P(CS)$ of the second. A stopping rule and terminal decision rule are also introduced and the expected sample size, $E(M)$, is obtained, ethical considerations suggest that this should be minimized.

Numerical comparisons are made between fully sequential, group sequential, fixed sample size designs and a class called group fully sequential designs.

2. Backward Induction. It is assumed that p_1, p_2 are a priori independent and are assigned beta priors with integer parameters (a_i, b_i), or beta (a_i, b_i), and density proportional to

$$p_i^{a_i-1} q_i^{b_i-a_i-1}, \qquad 1 \leq a_i \leq b_i - 1 \qquad i = 1, 2.$$

After s_i successes in m_i trials on arm i, the posterior density of p_i is beta (r_i, n_i) with $r_i = (a_i + s_i)$, $n_i = (b_i + m_i)$, $i = 1, 2$. The posterior expectation of p_i is $\hat{p}_i = r_i/n_i$, which is also the posterior predictive probability of the next trial on arm i resulting in a reward.

The stopping rule and terminal decision rule used in this paper both depend on the relative values of $\hat{p}_i$. Sampling stops when $|\hat{p}_1 - \hat{p}_2| \geq \triangle$, where $\triangle$ is preassigned, and after stopping, the arm with the larger $\hat{p}_i$ is chosen.

The following backward induction equations for determining the characteristics of any sequential design with single observations are given in Jones (1992) for both stopping and non-stopping problems. These recursive equations do not depend on the assumption of beta priors; $\hat{p}_i$ and later the posterior probability that $p_1 < p_2$, may be obtained using other prior distributions.

1. $E(R)$ is the expected return over all N pulls, where at termination the better arm is pulled for the remaining trials. An alternative interpretation is that a two-stage design is used. This takes a sequential sample in the first stage, which consists of a random number of observations, M, and takes $N - M$ observations on the better arm in the second. Let $A(r_1, n_1, r_2, n_2)$ be the expected return when the sampling process is in state (r_1, n_1, r_2, n_2); also let $A_1(r_1, n_1, r_2, n_2)$ be the expected return if arm 1 is pulled at the next trial and $A_2(r_1, n_1, r_2, n_2)$ be the expected return if arm 2 is pulled. Then

$$A(r_1, n_1, r_2, n_2) = [A_1(r_1, n_1, r_2, n_2), A_2(r_1, n_1, r_2, n_2)],$$

where

$$A_1(r_1, n_1, r_2, n_2) = \widehat{p}_1 \left[1 + A(r_1 + 1, n_1 + 1, r_2, n_2)\right]$$

$$+(1 - \widehat{p}_1) \left[A(r_1, n_1 + 1, r_2, n_2)\right];$$

$A_2(r_1, n_1, r_2, n_2)$ is defined in a similar way; which arm is pulled is determined by the design. At termination,

$$A(r_1, n_1, r_2, n_2) = (N - m_1 - m_2) \max\left[\widehat{p}_1, \widehat{p}_2\right].$$

2. $E(N_{(1)})$ is the expected number of pulls on the poorer arm. Using similar notation to that above, B is the expected number of pulls on the poorer arm at (r_1, n_1, r_2, n_2) and B_i refers to the expectation when arm $i = 1, 2$ is pulled at the next trial. For typographical ease the arguments are dropped; the recurrence equations are,

$$B = [B_1, B_2],$$

$$B_1 = P^* + B_{1E},$$

$$B_2 = (1 - P^*) + B_{2E},$$

where P^* is the posterior probability that $p_1 < p_2$, given by

$$P^* = \sum_{j=r_1}^{n_1-1} \frac{[(n_1 + n_2 - 1)\beta(r_2 + j + 1, n_1 + n_2 - r_2 - j - 1)]}{[j(r_2 + j)\beta(r_2, n_2 - r_2)\beta(j, n_1 - j)]},$$

where $\beta(\cdot, \cdot)$ is a beta function and

$$B_{1E} \equiv \widehat{p}_1 B(r_1 + 1, n_1 + 1, r_2, n_2) + (1 - \widehat{p}_1) B(r_1, n_1 + 1, r_2, n_2);$$

B_{2E} is similarly defined. At termination,

$$B = \begin{cases} (N - m_1 - m_2)P^* & \text{if } \widehat{p}_1 > \widehat{p}_2, \\ (N - m_1 - m_2)(1 - P^*) & \text{if } \widehat{p}_1 \leq \widehat{p}_2. \end{cases}$$

3. $P(CS)$ is the probability of correct selection. It is likely to be more affected by introducing a stopping rule than are $E(R)$ and $E(N_{(1)})$ because it is only based on the number of pulls to termination. Let C be the probability of correct selection at the point (r_1, n_1, r_2, n_2); then

$$C = [C_{1E}, C_{2E}],$$

where C_{1E} and C_{2E} are the expected probabilities of correct selection if arms 1 or 2, respectively, are pulled at the next trial. These are defined in a similar way to B_{1E} and B_{2E} above, and

$$\begin{aligned} C &= (1-P^*) \quad \text{if} \quad \widehat{p}_1 > \widehat{p}_2, \\ &\quad P^* \qquad\quad \text{if} \quad \widehat{p}_1 \leq \widehat{p}_2. \end{aligned}$$

4. To obtain $E(M)$, the expected sample size, let D be the expected number of observations remaining to termination when the process is in state (r_1, n_1, r_2, n_2) and let D_1 and D_2 depend on whether arm 1 or 2 is pulled at the next trial. Then

$$D = [D_1, D_2];$$

$$D_1 = 1 + D_{1E};$$

$$D_2 = 1 + D_{2E};$$

$$D = 0 \text{ at termination.}$$

If more than one observation is taken at a time on a single arm, then the above recurrence relations may be modified easily by using the probability of reaching all points together with the values A, B, C or D at those points. If the process is in state (r_1, n_1, r_2, n_2), then the posterior predictive probability of obtaining r successes in a fixed number of further trials n on arm 1 is given by

$$P_1 = \binom{n}{r} \frac{\beta(r_1 + r, n_1 + n - r_1 - r)}{\beta(r_1, n_1 - r_1)},$$

with P_2 defined similarly. In group sequential sampling, there is no choice between the arms, and strictly speaking, the resulting procedure is not a design, but a stopping problem. The characteristics in this case may again be obtained by a simple modification of the recurrence relations. Here there is just a choice between stopping or continuing, so at each point visited, the value of the characteristic

$$\big(E(R), E(N_{(1)}), P(CS)\big)$$

is either the terminal value or the expected return from taking a further batch of observations.

3. Optimal designs. In this section, several sequential procedures are compared for the three characteristics $E(R), E(N_{(1)}), P(CS)$ using exact numerical results obtained from the recurrence relations in the previous section. These are optimal within the method of sampling used and with respect to the stoppling rule and criterion chosen. For example, the fully sequential design takes single observations and the group sequential design takes observations in batches, and the optimal sampling rule is found for both. The recurrence relations above may also be used to evaluate non-optimal designs. Results when all N pulls are used are also presented; in the group sequential case, this gives a fixed sample size scheme. In all computational results it is assumed that the p_i's are assigned uniform or beta(1,2) priors, $N = 40$, and for the stopping rule, $\triangle = 0.4$. This value of $\triangle$ was chosen to give reasonable values for the characteristics, especially $P(CS)$; it was not chosen in any optimal manner. Obvious alternatives are possible; for example, the value of P^*, defined earlier, could be used.

Fully sequential procedures are designs which have a single observation at each stage; they are denoted FS. A hybrid design called grouped fully sequential (denoted G/F) is also considered; this is a design in which more than one observation is made on the better arm at each stage. Results are presented for the two observations at each stage.

Two group sequential procedures are considered. In one (denoted GS), pairs of observations are taken, one on each arm. The second one is a multistage sampling procedure (denoted GSV), in which the number taken at each stage is not uniform and decreases throughout the procedure. Results are presented for a four stage procedure with 10, 5, 3 and 2 observations on each arm at stages 1 to 4, respectively.

GSV is essentially a GS design in which the results of sampling are looked at after 20, 30, 36, and 40 observations, respectively, unless sampling has stopped at a previous stage. Hence the characteristics of the GSV design may be obtained by using the GS procedure when the stopping rule is only used at each of the four stages.

4. Results. The results for maximizing $E(R)$ are given in Table 1, minimizing $E(N_{(1)})$ in Table 2 and maximizing $P(CS)$ in Table 3. In each case, the expected sample sizes of the procedures that employ stopping rules are also given.

Table 1.
Maximum values of $E(R)$, uniform priors, $\triangle = 0.4$.

Procedure	$E(R)$	$E(M)$
FS (all N)	25.4469	
FS (stopping)	25.4138	23.9759
G/F (all N)	25.3129	
G/F (stopping)	25.3060	25.8134
GS	23.9266	22.5774
GSV	22.0522	32.4840
Fixed s.s.	20.0000	

Table 2.
Maximum values of $E(N_{(1)})$, uniform priors, $\triangle = 0.4$.

Procedure	$E(R)$	$E(M)$
FS (all N)	7.2584	
FS (stopping)	7.3926	23.2365
G/F (all N)	7.5680	
G/F (stopping)	7.6048	24.8883
GS	12.1791	22.5774
GSV	16.2718	32.4840
Fixed s.s.	20.0000	

Table 3.
Maximum values of $P(CS)$, uniform priors, $\triangle = 0.4$.

Procedure	$E(R)$	$E(M)$
FS (all N)	0.9051	
FS (stopping)	0.9039	29.4833
G/F (all N)	0.9048	
G/F (stopping)	0.9045	30.2532
GS	0.8945	22.5774
GSV	0.9027	32.4840
Fixed s.s.	0.9027	

As expected, the fully sequential procedures are the best performers when maximizing $E(R)$ with minimal reductions introduced by stopping. The superiority is more marked when minimizing $E(N_{(1)})$; however, one would not expect group sequential designs to do well here since, prior to stopping, equal numbers are taken on each arm. In Table 3, very little difference is noted in the values when maximizing $P(CS)$. The GS design with stopping requires a sample size considerably smaller than N to attain a value close to the optimal given by using the FS (all N) procedure.

5. Discussion. A fair comparison among the methods discussed above should take account of the complexity of the sampling procedure. Obviously the group sequential procedures are easier to use since only the stopping rule needs to be checked each time a batch of results is obtained. One way of accounting for this would be to introduce a cost structure into the problem where the reward is adjusted by a cost per observation and a cost per look at the results of sampling. Which procedure is chosen will depend on the primary objective of the experimenter or player. In a clinical trial, for example, the objective could be to make a terminal decision on the better treatment, hence maximizing $P(CS)$ could be the main objective. It is shown in the previous section that this seems fairly robust with respect to the choice of sampling procedure. However, ethical considerations could dictate that $E(R)$ and $E(N_{(1)})$ are more important. Alternatively, target levels could set for certain objectives or a weighted combination could be considered.

References

JONES, P.W. (1975). The two armed bandit. *Biometrika* **62** 523-524.

JONES, P.W. (1992). Multiobjective Bayesian bandits. *Bayesian Statistics 4* (Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., eds). Oxford: Clarendon Press, 689-695.

DEPARTMENT OF MATHEMATICS
KEELE UNIVERSITY
KEELE ST5 5BG
UNITED KINGDOM