Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.2 pp 195 200

A Note on the Two Dependent Bernoulli Arms

Dal Ho Kim[1] · Young Joon Cha · Jae Man Lee[2]

Abstract

We consider the Bernoulli two-armed bandit problem. It is well known that the myoptic strategy is optimal when the prior distribution is concentrated at two points in the unit square. We investigate several cases in the unit square whether the myoptic strategy is optimal or not. In general, the myoptic strategy is not optimal when the prior distribution is not concentrated at two points in the unit square

Keywords : Bandit problem, Bernoulli, myopic, optimal., prior distribution, two-armed.

1. Two-Armed Bandit Problems

Consider two dependent Bernoulli arms (or experiments) with the prior for (θ_1, θ_2) only concentrated on two points in the unit square. So arm 1 generates i.i.d. Bernoulli random variables (generically denoted by X) with mean θ_1, and arm 2 generates i.i.d. Bernoulli random variables (generically denoted by Y) with mean θ_2. Furthermore, every X is independent of every Y. The discount sequence is the N-horizon uniform. Objective is to maximize the expected sum of N observations when N is fixed. This two-armed Bandit problem is well introduced in Berry and Fristedt(1985).

It is of interest to know under what conditions the myopic strategy is optimal: always select the arm with greater mean. It is called myopic because it "behaves" as if there were always just on more trial to be allocated. When the myopic strategy is optimal, it means that the optimal strategy does not depend on the number of trials remaining: it is time invariant, so to speak.

Feldman(1962) considered this problem in the case that (θ_1, θ_2) is either

1. Department of Statistics, Kyungpook National University, 702-701, Korea
E-mail : dalkim@knu.ac.kr
2. Department of Statistics, Andong National University, Kyeongbuk 760-749, Korea

(a, b) or (b, a) where $0 \leq a < b \leq 1$. So the distribution of (θ_1, θ_2) can be specified by the single number $\phi = P(\theta_1 = a) = P(\theta_2 = b)$. He considered the procedure which minimizes the expected number of "mistakes" made by the statistician during the procedure (or minimizes the expected number of the inferior arm). In fact, this is equivalent to maximizing the expected number of successes in this situation.

For $j = 1, \ldots, N$ and $0 \leq \phi \leq 1$, we shall consider the situation in which the total number of observations remaining to be taken is j and the distribution of θ_1 and θ_2 is specified by the probability $\phi = P(\theta_1 = a)$. In this situation, let δ_j^X denote the procedure which specifies that the first observation should be taken on X and then an optimal procedure should be adopted over the remaining $j-1$ observations, and let $m_j^X(\phi)$ denote the expected number of mistakes during the j observations for which the procedure δ_j^X is used. Similarly, let δ_j^Y denote the procedure which specifies that the first observation should be taken on Y and then an optimal procedure should be adopted over the remaining $j-1$ observations, and let $m_j^Y(\phi)$ denote the expected number of mistakes when the procedure δ_j^Y is used.

Furthermore, let δ_j^{XY} be the procedure which specifies that the first observation should be taken on X, the second observation should be taken on Y, and then an optimal procedure should be adopted over the remaining $j-2$ observations. Similarly, let δ_j^{YX} be the procedure which specifies that the first observation should be taken on Y, the second observation should be taken on X, and then an optimal procedure should be adopted over the remaining $j-2$ observations. Also, let $m_j^{XY}(\phi)$ and $m_j^{YX}(\phi)$ be the expected numbers of mistakes for these procedures.

As usual, for the prior probability ϕ, let $\phi(X)$, $\phi(Y)$, $\phi(X, Y)$, or $\phi(Y, X)$ denote the posterior probability when either X (or Y) is taken or (X, Y) (or (Y, X)) are taken in that order. The following two lemmas and Theorem 1 are given in DeGroot (1970).

Lemma 1. For and for $0 \leq \phi \leq 1$, $j = 2, 3, \ldots$ $m_j^{XY}(\phi) = m_j^{YX}(\phi)$.

Lemma 2. For each fixed value of t in the interval $0 \leq t \leq 1$, the probabilities $P\{\phi(X) \geq t\}$ and $P\{\phi(Y) \geq t\}$ are nondecreasing functions or ϕ $(0 \leq \phi \leq 1)$.

Theorem 1. Let δ^* be a procedure which specifies that an observation should be taken on X at any stage for which $\phi = P(\theta_1 = a) < 1/2$ and that an

observation should be taken on Y at any stage for which $\phi > 1/2$. Then δ^* is an optimal sequential procedure.

Theorem 1 means that at every stage the statistician should taken an observation on the random variable X and Y for which there is the greater probability that the observed value will be 1. In other words, the optimal procedure is the myopic procedure under which the statistician makes a choice at each stage as if it were the final stage.

Kelley(1974) considered the case that the prior for (θ_1, θ_2) is concentrated on two points: (a, b) and (c, d) where $0 \leq a, b, c, d \leq 1$. For $n = 0, \ldots, N$, let $V_n(\xi)$ denote the optimal expected gain for the remaining n trials when ξ is the current prior distribution for (θ_1, θ_2). These functions are defined by the following recursive formulas.

$$V_0(\xi) = 0,$$

and

$$V_n(\xi) = \max \{E[X + V_{n-1}(\xi(X))], E[Y + V_{n-1}(\xi(Y))]\} \text{ for } n = 1, \ldots, N,$$

where $\xi(X)$ denotes the posterior distribution after an observation on X, and $\xi(Y)$ the posterior distribution after an observation on Y. It follows from above formulas that there exists functio $V_n(\xi) = \max \{F_n(\xi), G_n(\xi)\}$ns $F_n(\xi)$ and $G_n(\xi)$ such that

for $n = 1, \ldots, N$

These functions are also defined recursively. Let $F_0(\xi) = 0$ and $G_0(\xi) = 0$. Then for $n = 1, \ldots, N$,

$$F_n(\xi) = E[X + \max \{F_{n-1}(\xi(X)), G_{n-1}(\xi(X))\}],$$

and

$$G_n(\xi) = E[Y + \max \{F_{n-1}(\xi(Y)), G_{n-1}(\xi(Y))\}].$$

Let $D_n(\xi) = F_n(\xi) - G_n(\xi)$, the relative advantage of experiment 1 over experiment 2. Recursive formulas may be developed for defining $D_n(\xi)$. In fact,

$$D_1(\xi) = E(\theta_1) - E(\theta_2),$$

and for $n = 2, \ldots, N$,

$$D_n(\xi) = E[D_{n-1}(\xi(X))^+] + E[D_{n-1}(\xi(Y))^-]$$

where X^+ denotes $\max \{X, 0\}$ and X^- denotes $\min \{X, 0\}$.

Using these formulas, the optimal strategies can be characterized. If $a \geq b$and $c \geq d$, then $D_n(\xi) \leq 0$ for all ξ and for $n = 1, \ldots, N$. So the optimal strategy is to always use arm 2. Now assume that $b > a$ and $c > d$.

Theorem 2. For each $n = 1, 2, \ldots . N$, the following are true:

(**i**) $D_n(\xi)$ is a strictly increasing function of ξ.
(**ii**) $D_n(\xi)$ is a continuous function of ξ.
(**iii**) $D_n(\xi) < 0$ and $D_n(\xi) > 0$.
(**iv**) There exists an unique $\alpha_n \in (0, 1)$ such that $D_n(\alpha_n) = 0$.

Theorem 2 was proved by Kelly(1974). This theorem means that the optimal strategy is determined by an unique sequence of constants $\alpha_1, \alpha_2, , \ldots, \alpha_N$. From above results, it follows that whenever $a \leq b$ and $c \leq d$ the myopic strategy is optimal. Also, this is true whenever $a \geq b$ and $c \geq d$. When $b > a$ and $c > d$, from above theorem, the myopic strategy will be optimal if and only if $\alpha_1 = \alpha_2 = \cdots = \alpha_N = \alpha$ where $\alpha_1, \alpha_2, \ldots, \alpha_N$ are those unique constants determining the optimal strategy.

In search for conditions under which the myopic strategy is optimal, Kelley(1974) showed that for $N \geq 3$, except for some simple special cases, Feldman's assumption that $a = d$ and $b = c$ is necessary for the conclusion that myopic strategies are optimal.

Theorem 3. Suppose the prior distribution on (θ_1, θ_2) is concentrated at two points (a, b) and (c, d) in the unit square and that $N \geq 3$. The myopic strategy is optimal if and only if one of the following four conditions holds.

(**i**) $a \leq b$ and $c \leq d$,
(**ii**) $a \geq b$ and $c \geq d$,
(**iii**) $a + b = c + d = 1$,
(**iv**) $(c, d) = (b, a)$.

Theorem 3 was also proved by Kelly(1974). This theorem gives necessary and sufficient conditions for the optimality of the myoptic strategy in terms of a, b, c and d.

2. Main Results

Now we consider more general situations. Question: Even if the prior distribution is not concentrated at two points, does the myopic strategy remain optimal for some cases?

Let G be the distribution on $\{(\theta_1, \theta_2) \mid 0 \leq \theta_j \leq 1, i = 1, 2\}$, and let F_j be the corresponding marginal distribution of the θ_j $(i = 1, 2)$.

Case 1: $\{(\theta_1, \theta_2) \mid \theta_1 = 0\}$.

Since F_2 is to the right of F_1, we always use arm 2. So the myopic strategy is optimal.

Case 2: $\{(\theta_1, \theta_2) \mid \theta_2 \leq \theta_1\}$. Since $\theta_2 \leq \theta_1$ a.s., we always use arm 1. So the myopic strategy is optimal.

Case 3: $\{(\theta_1, \theta_2) \mid \theta_2 + \theta_1 = 1\}$. Let $v_1 = E[\theta_1]$ and $v_2 = E[\theta_1^2]$. So $E[\theta_2] = 1 - v_1$ and $E[\theta_2^2] = 1 - 2v_1 + v_2$. Let $n = 2$. Assume stay-on-a-winner rule. Consider

$$\Delta = v_1 + v_2 + (1 - v_1)[(v_1 - v_2)/(1 - v_1) \vee (1 - v_1)] - \{(1 - v_1) + (1 - 2v_1 + v_2) + v_1[(v_1 - v_2)/v_1 \vee v_1]\}$$

where $\vee$ denotes maximum. We show that if $v_1 \geq 1/2$, then $\Delta \geq 0$ for any v_2 such that $v_1^2 \leq v_2 \leq v_1$. Now

$$\Delta = 2(2v_1 - 1) + [(v_1 - v_2) \vee (1 - v_1)^2] - [(v_1 - v_2) \vee v_1^2].$$

i) $(v_1 - v_2) > v_1^2$: $\Delta = 2(2v_1 - 1) \geq 0$.

ii) $(1 - v_1^2) \leq (v_1 - v_2) \leq v_1^2$: $\Delta = 2(2v_1 - 1) + (v_1 - v_2) - v_1^2$

$= (2v_1 - 1) + [(v_1 - v_2) - (1 - v_1)^2] \geq 0$.

iii) $(v_1 - v_2) < (1 - v_1)^2$: $\Delta = 2(2v_1 - 1) + (1 - v_1)^2 - v_1^2 = 2v_1 - 1 \geq 0$.

Now consider

$$\Delta' = v_1 + v_1[v_2/v_1 \vee (1 - v_1)] + (1 - v_1)[(v_1 - v_2)/(1 - v_1) \vee (1 - v_1)] - \{(1 - v_1) + (1 - v_1)[(1 - 2v_1 + v_2)/(1 - v_1) \vee v_1] + v_1[(v_1 - v_2)/v_1 \vee v_1]\}$$

$$= (2v_1 - 1) + [v_2 \vee v_1(1 - v_1)] + [(v_1 - v_2) \vee (1 - v_1)^2] - [(1 - 2v_1 + v_2) \vee v_1(1 - v_1)] - [(v_1 - v_2) \vee v_1^2]$$

Since $[v_2 \vee v_1(1 - v_1)] \geq [(1 - 2v_1 + v_2) \vee v_1(1 - v_1)]$

and

$(2v_1 - 1) + [(v_1 - v_2) \vee (1 - v_1)^2] - [(v_1 - v_2) \vee v_1^2] \geq 0$ from Δ case, we can get easily $\Delta' \geq 0$.

Case 4: $\{(\theta_1, \theta_2) \mid 0 \leq \theta_1, \theta_2 \leq 1\}$ In general, myopic strategy is not optimal in the unit square. We have the following counter example. Assume $n = 2$. Consider $G = F_1 \times F_2$, $dF_1(\theta_1) \propto \theta_1^{-\varepsilon}(1 - \theta_1)^{-\varepsilon/2} d\theta_1$, and $dF_2(\theta_2) \propto d\theta_2$. Then $E(\theta_1 \mid F_1) = 1/2 - \delta$ for some $\delta > 0$ and $E(\theta_2 \mid F_2) = 1/2$. So in this case, myopic strategy is not optimal.

Reference

1. Berry, D. A. and Fristedt B. (1985) *Bandit Problems*, Chapman and Hall, New York.
2. DeGroot, M. H. (1970) *Optimal Statistical Decisions*, McGraw-Hill, New York.
3. Feldman, D. (1962) Contributions to the 'two-armed bandit' problem. *Ann. Math. Statist.* 33: 847-856.
4. Kelley, T. A. (1974) A note on the Bernoulli two-armed bandit problem. *Ann. Statist.* 2: 1056-1062.

[2002년 9월 접수, 2002년 10월 채택]