The Annals of Statistics
1992, Vol. 20, No. 4, 2111–2123

STAY-WITH-A-WINNER RULE FOR DEPENDENT BERNOULLI BANDITS[1]

BY K. SAMARANAYAKE

University of Idaho

The k-armed bandit problem on the Bernoulli dependent arms is discussed. Order relations on the prior distributions of the Bernoulli parameters using moments of the posterior are used to prove a monotonicity property of the value function. When $k = 2$, a stay-with-a-winner rule is derived for negatively correlated arms and for a certain class of positively correlated arms. These results are extensions of those given in Berry and Fristedt for independent Bernoulli arms. They also generalize the results of Benzing, Hinderer and Kolonko and Kolonko and Benzing.

1. Introduction. The bandit problem of independent Bernoulli arms has been investigated extensively. Many of the important results in this area are collected in a monograph by Berry and Fristedt (1985). This paper extends some of the results of Berry and Fristedt to include specific dependent Bernoulli problems. In particular, Section 2 contains extensions of the relations among distribution measures defined in Berry and Fristedt (1985). These extensions are used in Section 3 to formulate a theorem for monotonicity of the value function of a k-armed dependent Bernoulli bandit. In Section 4 recursive formulas for the advantage of one arm over the other will be provided when $k = 2$. The stay-with-a-winner rule is then extended to dependent arms.

This study uses the Bayesian approach. The Bernoulli parameters are $\theta = (\theta_1, \dots, \theta_k)$ and the prior distribution of θ is G. For a discount sequence $\mathbf{A} = (\alpha_1, \alpha_2, \dots)$ and distribution G, a strategy τ specifies the arm to be used at each stage for each possible history of observations. The worth of strategy τ in the $(G;\mathbf{A})$ bandit is denoted by $W(G;\mathbf{A};\tau)$. The value of the $(G;\mathbf{A})$ bandit $V(G;\mathbf{A})$ is the maximum possible worth. A strategy for which $V(G;\mathbf{A})$ is attained is called optimal.

Theorem 2.5.2 of Berry and Fristedt (1985) shows the existence of an optimal strategy for any $(G;\mathbf{A})$ bandit provided $E(|x_{i1}||G) < \infty$ for all $1 \le i \le k$, where x_{ij} is the outcome of arm i at state j. They also show in their Theorem 2.5.1 that for any fixed prior G with the previous property the value function $V(G;\mathbf{A})$ is uniformly continuous in $\mathbf{A}$. In Bernoulli cases $E(|x_{i1}||G) \le 1$ for all $1 \le i \le k$, and so these results apply herein.

2. Order relations on prior distributions. Some properties of the value function are derived in this section. The primary use of this development is to compare the value function of a bandit for different prior distributions.

Received January 1990; revised February 1992.
[1]Partially supported by NSF Grant DMS-85-05023.
AMS 1980 *subject classifications*. Primary 62L05; secondary 90D15.
Key words and phrases. Bernoulli bandits, stay with a winner.

The following definition concerns an order relation defined on k-dimensional distribution measures. It is a direct extension of the corresponding order relation on univariate distributions used in the independent case as in Berry and Fristedt (1985) and will be used to derive a monotone property of the value function. The notation $\{\sigma_*^{s*}\sigma_*^{f*}\}G$ represents $\sigma_1^{s_1}\phi_1^{f_1}\cdots\sigma_k^{s_k}\phi_k^{f_k}G$, the posterior distribution of θ, given that arm i has chosen s_i+f_i times resulting s_i successes.

DEFINITION 2.1. Let $m\geq 0$ and $1\leq i\leq k$. For two k-dimensional distribution measures G_1 and G_2, G_1 is said to be *m-greater than* G_2 *with respect to the ith component* (written $G_1\geq_i^m G_2$), if $E(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G_1)\geq E(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G_2)$ whenever $\sum_{j=1}^k(s_j+f_j)\leq m$ and $\{\sigma_*^{s*}\phi_*^{f*}\}G_j$ for $j=1,2$ are all defined with s_i and f_i any nonnegative integers. In addition, if $E(\theta_i|G_1)>E(\theta_i|G_2)$, then G_1 is *strictly m-greater than* G_2 *with respect to the ith component* (written $G_1>_i^m G_2$). If $G_1\geq_i^m G_2$ for all $i=1,\dots,k$, then G_1 is said to be *m-greater than* G_2, without any reference to components (written $G_1\geq^m G_2$). If $G_1\geq^m[\geq_i^m]G_2$ for all $m\geq 0$, then G_1 is said to be *greater than* G_2 (*with respect to the ith component*) (written $G_1\geq^\infty[\geq_i^\infty]G_2$).

Similar definitions apply for *strictly m-greater* and *strictly greater*.

The case $m=\infty$ is included in the results and definitions that follow, unless otherwise noted.

The main use of ordering distribution functions in a general $(G;\mathbf{A})$-bandit problem is when G_1 and G_2 in the preceding definition are chosen from G, $\sigma_1 G$, $\sigma_2 G$, $\phi_1 G$ and $\phi_2 G$. The relevant comparisons are given in Lemma 2.4.

The following example shows that any class of two point distributions with the same support is well ordered under $\geq$.

EXAMPLE 2.1. Consider the two-point distributions

$$G_t=p_t\delta_{(a_1,\dots,a_k)}+(1-p_t)\delta_{(b_1,\dots,b_k)}$$

for $t=1,2$. Without loss of generality assume $p_1\leq p_2$. For any j, $1\leq j\leq k$,

$$\sigma_j G_t=p_t^s\delta_{(a_1,\dots,a_k)}+(1-p_t^s)\delta_{(b_1,\dots,b_k)}$$

$$\phi_j G_t=p_t^f\delta_{(a_1,\dots,a_k)}+\left(1-p_t^f\right)\delta_{(b_1,\dots,b_k)}$$

for $t=1,2$, where

$$p_t^s=\frac{p_t a_j}{p_t a_i+(1-p_t)b_j},\qquad p_t^f=\frac{p_t(1-a_j)}{p_t(1-a_j)+(1-p_t)(1-b_j)}.$$

In either case $\sigma_j G_t$ or $\phi_j G_t$, the distribution takes the form

$$P_t\delta_{(a_1,\dots,a_k)}+(1-P_t)\delta_{(b_1,\dots,b_k)}$$

for $t=1,2$. That is, the resulting distributions are also two-point distributions with the same support. Furthermore, since for any fixed nonnegative x and y, the function $px/(px + (1 - p)y)$ is nondecreasing in p, we also have $P_1 \le P_2$. Thus we get the same ordering as between the p_t's.

Applying this result inductively, it can be easily shown that $a_i \le b_i$ if and only if $G_1 \ge_i^\infty G_2$.

The case $p_1 \ge p_2$ follows by symmetry, upon replacing p_t by $1 - p_t$.

This result can be summarized as: Suppose $a_i \le b_i$, then $p_1 \le p_2$ if and only if $G_1 \ge_i^\infty G_2$.

The next result follows immediately from Definition 2.1.

PROPOSITION 2.1. *For two distributions G_1 and G_2, $m \ge 1$ and $1 \le i \le k$, if $G_1 \ge_i^m G_2$, then $E(\theta_i|G_1) \ge E(\theta_i|G_2)$ (true for $m = 0$, too), $\sigma_j G_1 \ge_i^{m-1} \sigma_j G_2$ and $\phi_j G_1 \ge_i^{m-1} \phi_j G_2$ for any $1 \le j \le k$.*

REMARK. These results hold also if suffix i is absent throughout.

The next definition introduces covariance properties among arms in terms of posterior expectations of parameters.

DEFINITION 2.2. For two arms i and j $(1 \le i, j \le k)$ and $m \ge 1$, arm i is said to have *nonnegative* [*nonpositive*] *covariance with arm j up to stage m with respect to a distribution G* if Cov $(\theta_i, \theta_j|\{\sigma_*^{s_*}\phi_*^{f_*}G\}) \ge [\le]\ 0$ whenever $\sum_{i=1}^k (s_i + f_i) \le m$ and the various $\{\sigma_*^{s_*}\phi_*^{f_*}\}G$ are all defined. In addition, the property is *strict* if

$$\mathrm{Cov}(\theta_i, \theta_j|G) > [<]\ 0.$$

When $m = \infty$ we sometimes delete the term "up to stage m."

REMARK. In Definition 2.2, m cannot be zero. However, for convenience we sometimes let m take the value zero. A covariance at stage 0 is considered a tautology.

The following lemma provides an equivalence "monotone property" of the previous definition. The immediate use of this result is upcoming in Lemma 2.3, which in turn will be used to prove Theorem 3.1.

LEMMA 2.2. *Let $m > 1$. Within the domain of $\sum_{i=1}^k (s_i + f_j) \le m$ and whenever all the distributions $\{\sigma_*^{s_*}\phi_*^{f_*}\}G$ are well defined, the following statements are equivalent*:

(i) *$E(\theta_i|\{\sigma_*^{s_*}\phi_*^{f_*}\}G)$ is nondecreasing in s_j.*
(ii) *$E(\theta_i|\{\sigma_*^{s_*}\phi_*^{f_*}\}G)$ is nonincreasing in f_j.*
(iii) $\mathrm{Cov}(\theta_i, \theta_j|\{\sigma_*^{s_*}\phi_*^{f_*}\}G) \ge 0$.

PROOF. For $\{\sigma_*^{s*}\phi_*^{f*}\}G$ that satisfy $\sum_{i=1}^k(s_i + f_i) \le m - 1$,

$$E\big(\theta_i|\phi_j\{\sigma_*^{s*}\phi_*^{f*}\}G\big) \le E\big(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G\big) \tag{2.1}$$

if and only if

$$\frac{E\big(\theta_i(1-\theta_j)\big|\{\sigma_*^{s*}\phi_*^{f*}\}G\big)}{E(1-\theta_j)\big|\{\sigma_*^{s*}\phi_*^{f*}\}G\big)} \le E\big(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G\big)$$

since $E(\theta_j|\{\sigma_*^{s*}\phi_*^{f*}\}G) \ne 1$ as $\phi_j\{\sigma_*^{s*}\phi_*^{f*}\}G$ is well defined. Cross multiplying terms, this simplifies to

$$\operatorname{Cov}\big(\theta_i, \theta_j|\{\sigma_*^{s*}\phi_*^{f*}\}G\big) \ge 0. \tag{2.2}$$

Since $E(\theta_j|\{\sigma_*^{s*}\sigma_*^{f*}\}G) \ne 0$ as $\sigma_j\{\sigma_*^{s*}\phi_*^{f*}\}G$ is well defined, (2.4) can be written as

$$\frac{E\big(\theta_i\theta_j|\{\sigma_*^{s*}\phi_*^{f*}\}G\big)}{E\big(\theta_j|\{\sigma_*{}^{s*}\phi_*^{f*}\}G\big)} \ge E\big(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G\big),$$

which is equivalent to

$$E\big(\theta_i|\sigma_j\{\sigma_*^{s*}\phi_*^{f*}\}G\big) \ge E\big(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G\big). \tag{2.3}$$

Since the equivalence of (2.1), (2.2) and (2.3) holds for arbitrary $\{\sigma_*^{s*}\phi_*^{f*}\}G$, the proof follows. □

The following result follows similarly.

LEMMA 2.3. *Let $m > 1$. Within the domain of $\sum_{i=1}^k(s_i + f_j) \le m$ and whenever all the distributions $\{\sigma_*^{s*}\phi_*^{f*}\}G$ are well defined, the following statements are equivalent*:

(i) $E(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G)$ *is nonincreasing in* s_j.
(ii) $E(\theta_i|\{\sigma_*^{s*}\phi_*^{f*}\}G)$ *is nondecreasing in* f_j.
(iii) $\operatorname{Cov}(\theta_i, \theta_j|\{\sigma_*^{s*}\phi_*^{f*}\}G) \le 0$.

EXAMPLE 2.2. Consider the two-point distribution function

$$G = p\delta_{(a_1,\dots,a_k)} + (1-p)\delta_{(b_1,\dots,b_k)}.$$

For any $1 \le i, j \le k$, it is easy to show that

$$\operatorname{Cov}\big(\theta_i, \theta_j|G\big) = p(1-p)(a_i - b_i)(a_j - b_j).$$

The sign of this covariance is that of $(a_i - b_i)(a_j - b_j)$. In Example 2.1 we saw that at any given stage the posterior distribution $\{\sigma_*^{s*}\phi_*^{f*}\}G)$ has the same form as G, but with a different value of p. Hence, arm i has the nonnegative (nonpositive) covariance with arm j if and only if the line joining the two points (a_i, a_j) and (b_i, b_j) has nonnegative (nonpositive) slope.

In particular, consider the "classical bandit problem" of Feldman (1962) in which $G = p\delta_{(a,b)} + (1-p)\delta_{(b,a)}$. That is, (θ_1, θ_2) is known to be either (a, b)

or (b, a) a priori, but which is which is not known. Here, the arms are negatively correlated (up to stage ∞).

In the pervious example, the arms that are correlated positively or negatively have the same sign in the covariance at every stage afterwards. This is not necessarily true in general as the next example shows.

EXAMPLE 2.3. Consider a two-armed bandit with the prior distribution

$$G = \tfrac{1}{3}\delta_{(0.2,\,0.4)} + \tfrac{1}{3}\delta_{(0.5,\,0.6)} + \tfrac{1}{3}\delta_{(0.8,\,0.5)}.$$

We have $\mathrm{Cov}(\theta_1, \theta_2 | G) = 0.01$, so arms are positively correlated initially. Since

$$\sigma_1^2 G = \tfrac{4}{93}\delta_{(0.2,\,0.4)} + \tfrac{25}{93}\delta_{(0.5,\,0.6)} + \tfrac{64}{93}\delta_{(0.8,\,0.5)},$$

we get $\mathrm{Cov}(\theta_1, \theta_2 | \sigma_1^2 G) = -0.003$. Thus, after two successes on arm 1 the arms are negatively correlated.

The next definition extends the nonnegative covariance property given in Definition 2.2. The corresponding extension for nonpositive covariance for arm i requires θ_i to be degenerate and can be treated as a special case of nonnegative covariance.

DEFINITION 2.3. For a given distribution G, if arm i has nonnegative covariance with arm j up to stage m for all $j = 1, \dots, k$, then arm i is said to have *nonnegative covariances up to degree m with respect to G*. If all the arms have nonnegative covariances up to stage m with respect to G, then G is said to have the *nonnegative covariances up to stage m*.

The next result follows immediately from Lemma 2.2. It will be used to prove the upcoming Theorem 3.1.

LEMMA 2.4. *Let $1 \le m \le \infty$ and G be a given joint distribution. Then*:

(a) *If arm i has nonnegative covariance with arm j up to stage m with respect to G, then $\sigma_j G \ge_i^{m-1} G \ge_i^{m-1} \phi_j G$.*
(b) *If arm i has nonpositive covariance with arm j up to stage m with respect to G, then $\phi_j G \ge_i^{m-1} G \ge_i^{m-1} \sigma_j G$.*

REMARK. The strict version of Lemma 2.4 holds if the covariance property is strict.

3. Monotonicity of value function. The next theorem is the main result concerning monotonicity of the value function. Its proof follows the same steps as in the independent Bernoulli case given in Theorem 4.1.6 of Berry and Fristedt (1985). A discount sequence $\mathbf{A} = (\alpha_1, \alpha_2, \dots)$, with $\alpha_i = 0$ for all $i > n$ is said to have horizon n.

THEOREM 3.1. *Suppose for $1 \leq n \leq \infty$, joint distributions G_1 and G_2 each have nonnegative covariances up to stage $n-1$ and $G_1 \geq^{n-1} G_2$. Then for any discount sequence with horizon n,*

$$V(G_1; \mathbf{A}) \geq V(G_2; \mathbf{A}).$$

The following corollary follows from Lemmas 2.2, 2.3 and Theorem 3.1.

COROLLARY 3.2. *Let $n \geq 1$. Suppose G has the nonnegative covariances up to stage m and the horizon of the discount sequence $\mathbf{A}$ is m. Then for any i where $1 \leq i \leq k$, $V(\{\sigma_*^{s_*}\phi_*^{f_*}\}G; \mathbf{A})$ is nondecreasing with respect to s_i and nonincreasing with respect to f_i whenever $\Sigma_{i=1}^k(s_i + f_j) \leq n$.*

A result similar to Corollary 3.2 where $\mathbf{A}$ is geometric or truncated geometric is proven by Hengartner, Kalin and Theodorescu (1981) for two arms; this was extended to k arbitrary arms by Benzing, Hinderer and Kolonko (1984).

Extending results from finite to infinite horizon require continuity of the value function. For example, the proof of Theorem 3.1 for $n = \infty$ uses the fact that $V(G; \mathbf{A})$ is continuous to $\mathbf{A}$. Berry and Fristedt [(1985), Theorem 2.5.1] prove that $V(G; \mathbf{A})$ is uniformly continuous in $\mathbf{A}$ and give an example (Example 2.5.1) in which $V(G; \mathbf{A})$ is not continuous in the distribution G. However when the arms are Bernoulli, $V(G; \mathbf{A})$ is jointly continuous in $(G, \mathbf{A})$ and their Theorem 4.1.1 gives a method of proof when the arms are independent. It can be shown [Samaranayake (1988), Theorem 2.3.1] that this joint continuity of $V(G; \mathbf{A})$ is true in the present setting of dependent arms as well.

4. The advantage of one arm over another. This and subsequent sections consider $k = 2$. Define the advantage of arm 1 over arm 2 as

$$\Delta(G; \mathbf{A}) = V^{(1)}(G, \mathbf{A}) - V^{(2)}(G, \mathbf{A}),$$

where $\mathbf{V}^{(i)}(G, \mathbf{A})$ is the value of a strategy that starts with arm i and continues optimally. Using the arguments of Berry and Fristedt (1985), their Lemma 4.2.1 given for independent arms applies as well in the current setting of dependent arms:

LEMMA 4.1. *For the bandit $(G; \mathbf{A})$*

$$\begin{aligned} \Delta(G; \mathbf{A}) = {} & (\alpha_1 - \alpha_2)\left[E(\theta_1|G) - E(\theta_2|G)\right] \\ & + E(\theta_1|G)\Delta^+\left(\theta_1 G; \mathbf{A}^{(1)}\right) + E(1-\theta_1|G)\Delta^+\left(\phi_1 G; \mathbf{A}^{(1)}\right) \\ & - E(\theta_2|G)\Delta^-\left(\theta_2 G; \mathbf{A}^{(1)}\right) + E(1-\theta_2|G)\Delta^-\left(\phi_2 G; \mathbf{A}^{(1)}\right), \end{aligned} \tag{4.1}$$

where $\Delta^+ = \max(0, \Delta)$ *and* $\Delta = \max(0, -\Delta)$.

REMARK. Some of these Δ's are not defined when the support of either marginal distribution of G is concentrated on $\{0, 1\}$. Take the corresponding term in (4.1) to be equal to zero since the multiplier is zero.

The next definition introduces another monotone property that compares two joint distributions.

DEFINITION 4.1. Let $m \geq 0$. For joint distributions G_1 and G_2, G_1 is said to be *m-greater than* G_2 *in the* $i \to j$ *direction* (written as $G_1 \geq_{i \to j}^{m} G_2$) if

$$E\big[(\theta_i - \theta_j)\big|\{\sigma_*^{s_*}\phi_*^{f_*}\}G_1\big] \geq E\big[(\theta_i - \theta_j)\big|\{\sigma_*^{s_*}\phi_*^{f_*}\}G_2\big]$$

whenever $\Sigma_{t=1}^{2}(s_t + f_t) \leq m$ and $E[(\theta_i - \theta_j)|\{\sigma_*^{s_*}\phi_*^{f_*}\}G_2]$ for $t = 1, 2$ are all defined. In addition if $E[\theta_i - \theta_j|G_1] > E[\theta_i - \theta_j|G_2]$, then G_1 is *strictly m-greater than* G_2 *in the* $i \to j$ *direction* (written $G_1 \geq_{i \to j}^{m} G_2$). As in Definition 2.1, $m = \infty$ is possible.

EXAMPLE 4.1. Consider the two distribution functions in Example 2.1. As shown there, for any nonnegative integers (s_1, f_1, s_2, f_2),

$$\{\sigma_*^{s_*}\phi_*^{f_*}\}G_t = P_t^*\delta_{(a_1, a_2)} + (1 - P_t^*)\delta_{(b_1, b_2)}$$

for $t = 1, 2$. Furthermore, $P_1^* \leq P_2^*$ if and only if $p_1 \leq p_2$. For $a_i < b_i$,

$$E\big(\theta_i - \theta_j|\{\sigma_*^{s_*}\phi_*^{f_*}\}G_1\big) - E\big[(\theta_i - \theta_j)\big|\{\sigma_*^{s_*}\phi_*^{f_*}\}G_2\big]$$

$$= (P_1^* - P_2^*)\left[\frac{(b_j - a_j)}{(b_i - a_i)} - 1\right](b_i - a_i).$$

Hence if $p_1 \geq p_2$, depending on whether the slope of the line joining the two supporting points of G is at least 1 or at most 1, we have $G_1 \geq_{i \to j}^{\infty} G_2$ or $G_2 \geq_{i \to j}^{\infty} G_1$. The cases $p_1 \leq p_2$ and $a_i \geq b_i$ follow by symmetry.

DEFINITION 4.2. Let $m \geq 1$. Then direction $i \to j$ of a joint distribution G is said to have the *nonnegative* [*nonpositive*] *association with arm t up to stage m* if $\mathrm{Cov}(\theta_i - \theta_j, \theta_t|\{\sigma_*^{s_*}\phi_*^{f_*}\}G) \geq [\leq]\ 0$ whenever $\Sigma_{u=1}^{2}(s_u + f_u) \leq m$. In addition if $\mathrm{Cov}(\theta_i - \theta_j, \theta_t|G) > [<]\ 0$, the property is said to be *strict*. Furthermore in the above definition the phrase "with arm t" is omitted if the direction $i \to j$ of G has the nonnegative [nonpositive] association with every arm.

As in Section 2, we allow $m = 0$, but an association property of degree zero is not an extra condition. Also, both these definitions extend to the case $m = \infty$ in the obvious way, and $m = \infty$ is deleted in the notation. The following results include the case $m = \infty$.

As in the covariance properties defined in Section 2, Definition 4.2 has a "monotone property" structural equivalence.

LEMMA 4.2. *Let* $m \geq 1$. *Then within the domain of* $\Sigma_{u=1}^{2}(s_u + f_u) \leq m$ *and whenever all the distributions* $\{\sigma_*^{s_*}\phi_*^{f_*}\}G$ *are defined*, $E[\theta_i - \theta_j|\{\sigma_*^{s_*}\phi_*^{f_*}\}G]$ *is* *nondecreasing* [*nonincreasing*] *in* s_t

$$\textit{if and only if } E\big[\theta_i - \theta_j | \{\sigma_*^{s_*}\phi_*^{f_*}\}G\big] \textit{ is nonincreasing } [\textit{nondecreasing}] \textit{ in } f_t,$$

$$\textit{if and only if } \operatorname{Cov}\big(\theta_i - \theta_j, \theta_t | \{\sigma_*^{s_*}\phi_*^{f_*}\}G\big) \geq [\leq]\ 0.$$

PROOF. The proof of Lemma 4.2 follows along the lines of the proof of Lemma 2.2. In particular, substituting $\theta_i - \theta_j$ for θ_i, σ_t and ϕ_t for σ_j and ϕ_j and θ_t for θ_j we can rewrite the equivalence of statements (2.1), (2.2) and (2.3) to obtain the lemma. □

EXAMPLE 4.2. Consider the distribution function G in Example 2.2 with $k = 2$. For any i, j,

$$V(\theta_i|G) - \operatorname{Cov}\big(\theta_i, \theta_j|G\big) = p(1-p)(a_i - b_i)^2\left[1 - \frac{b_j - a_j}{b_i - a_i}\right].$$

In Example 2.1 we saw that for any given stage $\{\sigma_*^{s_*}\phi_*^{f_*}\}G$ also takes the form of G and hence $\operatorname{Var}(\theta_i|G) - \operatorname{Cov}(\theta_i, \theta_j|G)$ takes the sign of $(1 - (b_j - a_j)/(b_i - a_i))$. That is, the direction $i \to j$ of G has the nonnegative or nonpositive association with arm i depending on the slope of the line joining two points of support of G is at most 1 or at least 1.

EXAMPLE 4.3. Consider the two-dimensional distribution G_p with θ_2 defined by

$$\theta_2 = \begin{cases} \theta_1, & \text{with probability } p, \\ 1 - \theta_1, & \text{with probability } 1 - p. \end{cases}$$

Then we have $\operatorname{Cov}(\theta_1, \theta_2|G_p) = (2p - 1)\operatorname{Var}(\theta_1|G_p)$. So $\operatorname{Cov}(\theta_1, \theta_2|G_p) \leq \operatorname{Var}(\theta_1|G_p)$ for any $0 \leq p \leq 1$. Since this inequality is true for any G_p, we have

$$\operatorname{Cov}\big(\theta_1, \theta_2|\{\sigma_*^{s_*}\phi_*^{f_*}\}G_p\big) \leq V\big(\theta_1|\{\sigma_*^{s_*}\phi_*^{f_*}\}G_p\big)$$

for any s_* and f_*. Thus direction $1 \to 2$ of G_p has the nonnegative association (up to stage ∞) with arm 1. Note that in this example the prior distribution of θ_1 (or θ_2) is not specifically defined, and was not needed in the comparison.

The next two results are immediate from the definitions. They will be used to show monotonicity with respect to the distribution G of the $\Delta(G; \mathbf{A})$.

PROPOSITION 4.3. *Let* $m \geq 0$. *Then for two distribution functions* G_1 *and* G_2:

(a) $G_1 \geq_{i \to j}^m G_2$ *if and only if* $G_2 \geq_{j \to i}^m G_1$.
(b) *If* $G_1 \geq_i^m G_2$ *and* $G_2 \geq_j^m G_1$, *then* $G_1 \geq_{i \to j}^m G_2$.

PROPOSITION 4.4. *Let $m \geq 1$. Then for two distribution functions G_1 and G_2:*

(a) *If $G_1 \geq^{m}_{i \to j} G_2$, then for any arm t, $\sigma_t G_1 \geq^{m-1}_{i \to j} \sigma_t G_2$ and $\phi_t G_1 \geq^{m-1}_{i \to j} \phi_t G_2$.*

(b) *If direction $i \to j$ of G has the nonnegative association with arm t up to stage m, then*

$$\sigma_t G \geq^{m-1}_{i \to j} G \geq^{m-1}_{i \to j} \phi_t G.$$

(c) *If direction $i \to j$ of G has the nonpositive association with arm t up to stage m, then*

$$\phi_t G \geq^{m-1}_{i \to j} G \geq^{m-1}_{i \to j} \sigma_t G.$$

Furthermore, the strict versions hold in (b) *and* (c) *if the association properties are strict.*

The next result describes the relations possible between directional association property and covariance properties described in Section 2.

PROPOSITION 4.5. *With respect to a distribution G and up to degree m for any $m \geq 1$:*

(a) *If direction $i \to j$ of G either has nonnegative association with arm j or nonpositive association with arm i, then arm i has nonnegative covariance with arm j.*

(b) *If arm i has nonpositive covariance with arm j, then direction $i \to j$ of G has nonnegative association with arm i and the nonpositive association with arm j.*

(c) *If direction $i \to j$ of G has the nonnegative [nonpositive] association with arm j [i], then direction $i \to j$ of G has the nonnegative [nonpositive] association with arm i [j].*

PROOF. Parts (a) and (b) follow directly from Lemmas 2.2 and 4.2. Part (c) follows from Lemma 4.2 and from the fact that $\mathrm{Var}(X) \leq \mathrm{Cov}(X, Y)$ implies $\mathrm{Var}(Y) \geq \mathrm{Cov}(X, Y)$.

5. Monotonicity of delta functions. Now we state a lemma for comparing two Δ functions with respect to two distribution functions G_1 and G_2. This lemma will be used to prove staying with a winner properties.

LEMMA 5.1. *Let $1 \leq n \leq \infty$. Suppose $\mathbf{A}$ is a nonincreasing discount sequence of horizon n. For two distributions G_1 and G_2,*

$$\Delta(G_1; \mathbf{A}) \geq \Delta(G_2; \mathbf{A}) \tag{5.1}$$

if either of the following holds:

(i) *Direction* $1 \to 2$ *of both* G_1 *and* G_2 *have nonnegative association with arm* 2 *up to stage* n, $G_1 \geq^{n-1} G_2$ *and* $G_1 \geq^{n-1}_{1\to 2} G_2$.

(ii) *Arm* 1 *has nonpositive covariance with arm* 2 *up to stage n with respect to both* G_1 *and* G_2, $G_1 \geq^{n-1}_1 G_2$ *and* $G_2 \geq^{n-1}_2 G_1$.

Furthermore (5.1) *is strict if in case* (i) $G_1 \geq^{n-1}_{1\to 2} G_2$ *and in case* (ii) *either* $G_1 \geq^{n-1}_1 G_2$ *or* $G_2 \geq^{n-1}_2 G_1$.

PROOF. (i) We will use induction on n. Let $P(n)$ be the statement, "For two distributions G_1 and G_2, if $G_1 \geq^{n-1}_{1\to 2} G_2$, $G_1 \geq^{n-1} G_2$, direction $1 \to 2$ of both G_1 and G_2 has the increment property with respect to arm 2 and $\mathbf{A}$ is any nonincreasing discount sequence of horizon n, then $\Delta(G_1; \mathbf{A}) \geq \Delta(G_2; \mathbf{A})$."

For $n = 1$, $\Delta(G_1; \mathbf{A}) - \Delta(G_2; \mathbf{A}) = \alpha_1[E(\theta_1|G_1) - E(\theta_2|G_1) - E(\theta_1|G_2) + E(\theta_2|G_2)]$ which is nonnegative since $G_1 \geq^{0}_{1\to 2} G_2$. So $P(1)$ is true.

Now suppose $n > 1$ and assume $P(n-1)$. Let G_1 and G_2 satisfy $G_1 \geq^{n-1}_{1\to 2} G_2$, $\mathbf{A}$ be a nonincreasing discount sequence with horizon n and assume condition (i). Using Lemma 4.1,

$$
\begin{aligned}
&\Delta(G_1; \mathbf{A}) - \Delta(G_2; \mathbf{A}) \\
&\quad = (\alpha_1 - \alpha_2)\left[E(\theta_1|G_1) - E(\theta_1|G_2) - E(\theta_2|G_1) + E(\theta_2|G_2)\right] \\
&\qquad + E(\theta_1|G_1)\left[\Delta^+(\sigma_1 G_1; \mathbf{A}^{(1)}) - \Delta^+(\sigma_1 G_2; \mathbf{A}^{(1)})\right] \\
&\qquad + E(1 - \theta_1|G_1)\left[\Delta^+(\phi_1 G_1; \mathbf{A}^{(1)}) - \Delta(\phi_1 G_2; \mathbf{A}^{(1)})\right] \\
&\qquad + \left[E(\theta_1|G_1) - E(\theta_1|G_2)\right]\left[\Delta^+(\sigma_1 G_2; \mathbf{A}^{(1)}) - \Delta^+(\phi_1 G_2; \mathbf{A}^{(1)})\right] \\
&\qquad + E(\theta_2|G_2)\left[\Delta^-(\sigma_2 G_2\, \mathbf{A}^{(1)}) - \Delta^-(\sigma_2 G_1; \mathbf{A}^{(1)})\right] \\
&\qquad + E(1 - \theta_2|G_2) - \left[\Delta^-(\phi_2 G_2; \mathbf{A}^{(1)}) - \Delta^-(\phi_2 G_1; \mathbf{A}^{(1)})\right] \\
&\qquad + \left[E(\theta_2|G_1) - (\theta_2|G_2)\right]\left[\Delta^-(\phi_2 G_1 \colon \mathbf{A}^{(1)}) - \Delta^-(\sigma_2 G_1; \mathbf{A}^{(1)})\right].
\end{aligned} \tag{5.2}
$$

In the right side of (5.2), the first term is nonnegative since $\mathbf{A}$ is nonincreasing and $G_1 \geq^{n-1}_{1\to 2} G_2$. The second, third, fifth and sixth terms are nonnegative by Propositions 2.1 and 4.4(c) in view of $P(n-1)$. In the fourth and last terms the differences in moments are nonnegative by Proposition 2.1 and the differences in Δ's are nonnegative by Proposition 4.5(a), 4.5(c), 4.4(b) and Lemma 2.3 in view of $P(n-1)$. So the right side of (5.2) is nonnegative and $P(n)$ holds.

The strict version of the result follows from the same proof by induction, and uses the fact that at least one of the following relations that are due to Proposition 4.4(a) and 4.4(b) must hold in its strict version:

$$
\sigma_1 G_1 \geq^{n-2}_{1\to 2} \sigma_1 G_2, \qquad \phi_1 G_1 \geq^{n-2}_{1\to 2} \phi_1 G_2 \quad \text{or} \quad \sigma_1 G_2 \geq^{n-2}_{1\to 2} \phi_1 G_2.
$$

Proof of case (ii) is similar to (i). □

The major application of Lemma 5.1 is the upcoming stay with a winner property. The following results characterize some situations concerning how the advantage of arm 1 over arm 2 changes with successes and failures.

PROPOSITION 5.2. *Suppose* $\mathbf{A}$ *is a nonincreasing discount sequence of horizon* n. *For the* $(G;\mathbf{A})$ *bandit*:

(a) *If direction* $1 \to 2$ *of* G *has nonnegative association up to stage* n *with respect to arm* 2, *then*

$$\Delta(\sigma_1 G;\mathbf{A}) \ge \Delta(G;\mathbf{A}) \ge \Delta(\phi_1 G;\mathbf{A}),$$
$$\Delta(\sigma_2 G;\mathbf{A}) \ge \Delta(G;\mathbf{A}) \ge \Delta(\phi_2 G;\mathbf{A}).$$

(b) *If arm* 1 *has nonpositive covariance with arm* 2 *up to stage* n *with respect to* G, *then*

$$\Delta(\sigma_1 G;\mathbf{A}) \ge \Delta(G;\mathbf{A}) \ge \Delta(\phi_1 G;\mathbf{A}),$$
$$\Delta(\sigma_2 G;\mathbf{A}) \le \Delta(G;\mathbf{A}) \le \Delta(\phi_2 G;\mathbf{A}).$$

(c) *If direction* $1 \to 2$ *of* G *has nonpositive association up to stage* n *with respect to arm* 1, *then*

$$\Delta(\sigma_1 G;\mathbf{A}) \le \Delta(G;\mathbf{A}) \le \Delta(\phi_1 G;\mathbf{A}),$$
$$\Delta(\sigma_2 G;\mathbf{A}) \le \Delta(G;\mathbf{A}) \le \Delta(\phi_2 G;\mathbf{A}).$$

Furthermore, if the support of one of the marginal distributions of G *has more than one point and in cases of* (a) *and* (c) *the association property is strict, all these inequalities involving the* Δ*'s are strict.*

6. Stay-with-a-winner rule. Now we state and prove the stay-with-a-winner rule for cases (i) and (ii) in Lemma 5.1. Its proof is an adaptation of that given for the independent case in Berry and Fristedt (1985, Theorem 4.3.8).

THEOREM 6.1. *Let* $2 \le n \le \infty$. *Suppose* $\mathbf{A}$ *is a nonincreasing discount sequence of horizon* n, *the support of one of the marginal distributions of* G *has more than one point and either* $\alpha_1 = \alpha_2$ *or* $E(\theta_1|G) \le E(\theta_2|G)$. *Then, if either* (i) *direction* $1 \to 2$ *of* G *has the strict nonnegative association with arm* 2 *up to stage* n, *or* (ii) *arm* 1 *has nonpositive covariance with arm* 2 *with respect to* G *up to stage* n, *then* $\Delta(G;\mathbf{A}) \ge 0$ *implies* $\Delta(\sigma_1 G;\mathbf{A}^{(1)}) \ge 0$.

PROOF. We will prove the result by contradiction assuming θ_2 is not concentrated at 0 or 1 cases in which the result is trivial.

The proof is detailed to accommodate both cases (i) and (ii) at the same time and also to be used in slightly different versions later. From Lemma 4.1, $\Delta(G;\mathbf{A}) \ge 0$ and $\alpha_1 = \alpha_2$ or $E(\theta_1|G) \le E(\theta_2|G)$ gives

$$\begin{aligned}0 \le {} & E(\theta_1|G)\Delta^+(\sigma_1 G;\mathbf{A}^{(1)}) + E(1-\theta_1|G)\Delta^+(\phi_1 G;\mathbf{A}^{(1)}) \\ & - E(\theta_2|G)\Delta^-(\sigma_2 G;\mathbf{A}^{(1)}) - E(1-\theta_2|G;\mathbf{A}^{(1)})\Delta^-(\phi_2 G;\mathbf{A}^{(1)}).\end{aligned} \tag{6.1}$$

Now suppose $\Delta(\sigma_1 G; \mathbf{A}) \le 0$. Then by Proposition 5.2(a) for case (i) or by Proposition 5.2(b) for case (ii), $\Delta(\phi_1 G; \mathbf{A}) \le 0$. Hence (6.1) reduces to

$$(6.2) \qquad 0 \le -E(\theta_2|G)\Delta^-(\sigma_2 G; \mathbf{A}^{(1)}) - E(1-\theta_2|G)\Delta^-(\phi_2 G; \mathbf{A}^{(1)}).$$

So $\Delta^-(\sigma_2 G; \mathbf{A}^{(1)}) = 0$ and $\Delta^-(\phi_2 G; \mathbf{A}^{(1)}) = 0$ since θ_2 is not concentrated at 0 or 1. That is,

$$\Delta(\sigma_2 G; \mathbf{A}^{(1)}) \ge 0 \quad \text{and} \quad \Delta(\phi_2 G; \mathbf{A}^{(1)}) \ge 0.$$

But for case (i) by Proposition 5.2(a),

$$\Delta(\sigma_1 G; \mathbf{A}^{(1)}) > \Delta(G; \mathbf{A}^{(1)}) > \Delta(\phi_2 G; \mathbf{A}^{(1)})$$

and for case (ii) by Proposition 5.2(b). Thus in either case (i) or (ii) we have $\Delta(\sigma_1 G; \mathbf{A}^{(1)}) > 0$, which contradicts the assumption. □

In the special case of uniform discount sequence, either (i) or (ii) of Theorem 6.1 assures the stay-with-a-winner property. Result (ii) in this special case is proven in Koloko and Benzing (1985).

The following slightly different versions of Theorem 6.1 are sometimes useful. In these, we do not impose restrictions on the support of G and also do not require the strict versions of the increment property.

COROLLARY 6.2. (a) *In Theorem* 6.1 *suppose the strict property in* (i) *is relaxed and the support of G is unrestricted. Then in either case* (i) *or* (ii),

$$\Delta(G; \mathbf{A}) > 0 \quad \textit{implies} \quad \Delta(\sigma_1 G; \mathbf{A}^{(1)}) > 0.$$

(b) *In addition to the assumptions in* (a), *suppose that* $\alpha \ne \alpha_2$ *and* $E(\theta_1|G) \ne E(\theta_2|G)$. *Then*,

$$\Delta(G; \mathbf{A}) \ge 0 \quad \textit{implies} \quad \Delta(\sigma_1 G; \mathbf{A}^{(1)}) < 0.$$

EXAMPLE 6.1. Let G be the distribution in Example 2.2 with $k = 2$. From Example 4.2 we know that if the slope of the line joining the two points of support is greater than or equal to 1 then the direction $2 \to 1$ has nonpositive association with arm 2, or equivalently the direction $1 \to 2$ has nonnegative association with arm 2. Hence by Theorem 6.1 in either of the cases $\alpha_1 = \alpha_2$ or $E(\theta_1|G) \le E(\theta_2|G)$, arm 2 has the stay-with-the-winner property. When the slope is negative, that is, when the arms are negatively correlated, by Theorem 6.1(b) the arm with smaller mean has the stay-with-the-winner property.

Bradt, Johnson and Karlin (1956, page 1067) provides a counter example to show that stay-with-a-winner rule is not true in general. The prior in their example is a special case of Example 6.1. Arm 1 is optimal initially and the optimal strategy requires a switch to arm 2 on a success. Clearly, neither of the monotone properties hold with respect to arm 1 for this prior. However, arm 2 does have the stay-with-a-winner property according to Example 6.1.

Acknowledgments. This work is part of the author's Ph.D. dissertation at University of Minnesota written under the supervision of Professor D. A. Berry, whose generous guidance and suggestions are gratefully acknowledged. I would also like to thank the referees and an Associate Editor for their helpful comments that improved the readability of the paper.

REFERENCES

BENZING, H., HINDERER K. and KOLONKO, M. (1984). On the k-armed Bernoulli bandit: Monotonicity of the total reward under an arbitrary prior distribution. *Math. Operationsforsch. Statist. Ser. Optim.* **15** 583–595.

BERRY, D. A. and FRISTEDT, B. (1985). *Bandit Problems Sequential Allocation of Experiments.* Chapman-Hall, London.

BRADT, R. N., JOHNSON, S. M. and KARLIN, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.* **27** 1060–1074.

FELDMAN, D. (1962). Contributions to the two-armed bandit problem. *Ann. Math. Statist.* **33** 847–856.

HENGARTNER, W., KARLIN, O. and THEODORESCU, R., (1981). On the Bernoulli two armed bandit problem. *Math. Operationsforsch. Statist. Ser. Optim.* **12** 307–316.

KOLONKO, M. and BENZING, H. (1985). On monotonic optimal decisions rules and the stay-on-a-winner rule for the two-armed bandit. *Metrika* **32** 395–407.

SAMARANAYAKE, K. (1988). Bernoulli k-armed bandits with dependent arms. Ph.D. dissertation, Univ. Minnesota.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF IDAHO
MOSCOW, IDAHO 83843