

0517

S 389

ARCHIEF

W
A

SOME REMARKS ON THE TWO-ARMED BANDIT

by

J. Fabius and W.R. van Zwet

—— Paper presented at the European Meeting of IMS, TIMS, ES and
IASPS, Amsterdam, September 2 - 7, 1968.

BIBLIOTHEEK MATHEMATISCH CENTRUM
—— AMSTERDAM ——

SOME REMARKS ON THE TWO-ARMED BANDIT 1)

by J. Fabius and W.R. van Zwet,

University of Leiden and Mathematisch Centrum, Amsterdam

1 INTRODUCTION

In this paper we consider the following situation: An experimenter has to perform a total of N trials on two Bernoulli-type experiments E_1 and E_2 with success probabilities α and β respectively, where both α and β are unknown to him. The trials are to be carried out sequentially and independently, except that for each trial the experimenter may choose between E_1 and E_2 , using the information obtained in all previous trials. The decisions on the part of the experimenter to use E_1 or E_2 in the successive trials may be randomized, i.e. for any trial he may use a chance mechanism in order to choose E_1 or E_2 with probabilities δ and $1-\delta$ respectively, where δ may depend on the decisions taken and the results obtained in the previous trials. A strategy Δ will be a set of such δ 's, completely describing the experimenters behaviour in every conceivable situation.

1) Report S 399, Mathematisch Centrum, Amsterdam.

We assume the experimenter wants to maximize the number of successes. More precisely, we assume that he incurs a loss

$$(1.1) \quad L(\alpha, \beta, s) = N \max(\alpha, \beta) - s$$

if he scores a total of s successes. If he uses a strategy Δ , his expected loss is then given by the risk function

$$(1.2) \quad R(\alpha, \beta, \Delta) = N \max(\alpha, \beta) - E(S \mid \alpha, \beta, \Delta),$$

where S denotes the random number of successes obtained. Thus the risk of a strategy Δ equals the expected amount by which the number of successes the experimenter will obtain using Δ falls short of the number of successes he would score if he were clairvoyant and would use the more favourable experiment throughout the N trials.

We say that state $(m, k; n, l)$ is reached during the series of trials if in the first $m + n$ trials E_1 is used m times, yielding k successes, and E_2 is used n times, yielding l successes. Clearly, under a strategy Δ , the probability that this will happen is of the form

$$(1.3) \quad \pi_{\alpha, \beta, \Delta}(m, k; n, l) = p_{\Delta}(m, k; n, l) \alpha^k (1 - \alpha)^{m - k} \beta^l (1 - \beta)^{n - l},$$

where $p_{\Delta}(m, k; n, l)$ depends on the state $(m, k; n, l)$ and the strategy Δ , but not on α and β . It is easy to show (e.g. by induction on N) that the class of all strategies is convex in the sense that there exists, for every pair of strategies Δ_1 and Δ_2 and for every $\lambda \in [0, 1]$, a strategy Δ such that

$$(1.4) \quad p_{\Delta}(m, k; n, l) = \lambda p_{\Delta_1}(m, k; n, l) + (1 - \lambda) p_{\Delta_2}(m, k; n, l)$$

for every state $(m, k; n, l)$.

Moreover, this strategy Δ can always be taken to be such, that according to it the experimenter should base all his decisions exclusively on the numbers of successes and failures observed with E_1 and E_2 , irrespective of the order in which these data became available. Denoting the class of all such strategies by $\mathcal{D}$ and remarking that $R(\alpha, \beta, \Delta)$ can be expressed in terms of the $\pi_{\alpha, \beta, \Delta}(m, k; n, l)$, we may conclude that $\mathcal{D}$ is an essentially complete class of strategies. We denote the probabilities δ constituting any strategy in $\mathcal{D}$ by $\delta(m, k; n, l)$: the probability with which the experimenter, having completed the first $m + n$ trials and thereby having reached state $(m, k; n, l)$, chooses E_1 for the next trial.

We note that if $p_{\Delta}(m, k; n, l) = 0$ for a state $(m, k; n, l)$, then $\delta(m, k; n, l)$ does not play any role in the description of Δ and may be assigned an arbitrary value without affecting the strategy. We shall say that any strategy Δ' such that $p_{\Delta'}(m, k; n, l) = p_{\Delta}(m, k; n, l)$ for all states $(m, k; n, l)$ constitutes a version of Δ .

Since we are considering a symmetric problem in the sense that it remains invariant when α and β are interchanged, it seems reasonable to consider strategies with a similar symmetry. Thus we are led to define the class $\mathcal{J}$ of all symmetric strategies:

$\Delta \in \mathcal{J}$ iff $\Delta \in \mathcal{D}$ and $\delta(m, k; n, l) = 1 - \delta(n, l; m, k)$ for all states $(m, k; n, l)$ with $p_{\Delta}(m, k; n, l) \neq 0$.

Clearly, for $\Delta \in \mathcal{J}$,

$$(1.5) \quad \delta(m, k; m, k) = \frac{1}{2} \quad \text{if} \quad p_{\Delta}(m, k; m, k) \neq 0, \quad \text{and}$$

$$(1.6) \quad p_{\Delta}(m, k; n, l) = p_{\Delta}(n, l; m, k) \quad \text{for all states } (m, k; n, l).$$

It follows that, for $\Delta \in \mathcal{J}$ and all (α, β) ,

$$(1.7) \quad R(\alpha, \beta, \Delta) = R(\beta, \alpha, \Delta).$$

Another argument in favour of considering $\mathcal{J}$ is the following result.

THEOREM 1

There is a strategy $\Delta \in \mathcal{J}$ with minimax risk.

PROOF

The existence of a minimax-risk strategy $\Delta_1 \in \mathcal{D}$ is well known for this type of problem. Let $\Delta_2 \in \mathcal{D}$ be defined by

$$\delta_2(m, k; n, l) = 1 - \delta_1(n, l; m, k) \quad \text{for all states } (m, k; n, l).$$

Then

$$p_{\Delta_2}(m, k; n, l) = p_{\Delta_1}(n, l; m, k) \quad \text{for all states, hence}$$

$$R(\alpha, \beta, \Delta_2) = R(\beta, \alpha, \Delta_1) \quad \text{for all } (\alpha, \beta) \quad \text{so that } \Delta_2 \text{ has}$$

minimax risk. By convexity we may construct a strategy $\Delta \in \mathcal{D}$ satisfying (1.4) with $\lambda = \frac{1}{2}$. We have

$$(1.8) \quad p_{\Delta}(m, k; n, l) = p_{\Delta}(n, l; m, k) \quad \text{for all states, and}$$

$$R(\alpha, \beta, \Delta) = \frac{1}{2}R(\alpha, \beta, \Delta_1) + \frac{1}{2}R(\alpha, \beta, \Delta_2) \quad \text{for all } (\alpha, \beta), \text{ which}$$

implies that Δ too has minimax risk. Finally, if $\Delta \notin \mathcal{J}$, we define

$\Delta^* \in \mathcal{J}$ by

$$\delta^*(m, k; n, l) = \frac{1}{2}\delta(m, k; n, l) + \frac{1}{2}[1 - \delta(n, l; m, k)] \quad \text{for all states.}$$

One easily verifies that (1.8) implies

$$p_{\Delta^*}(m, k; n, l) = p_{\Delta^*}(m, k; n, l) \quad \text{for all states, and as a result}$$

$\Delta^* \in \mathcal{J}$ has minimax risk.

In section 2 we derive a recurrence relation, which we then use in section 3 to study the structure of admissible strategies in $\mathcal{D}$. For these strategies we prove certain monotonicity properties of $\delta(m,k;n,1)$. Though these results may seem intuitively evident, one does well to remember that the two-armed bandit problem has been shown to defy intuition in many aspects (cf. [1]). Section 4 indicates how our results to some degree facilitate the search for minimax-risk strategies. Even so, the algebra involved is extremely tedious for N as small as 4. Already for slightly larger values of N it remains prohibitive.

Among the contributions to the two-armed bandit problem the work of W. Vogel, who considered the same set-up we do, deserves special mention. In [2] he discussed a certain subclass of the class $\mathcal{J}$, and in [3] he obtained asymptotic bounds for the minimax risk for $N \rightarrow \infty$. Since we shall not be concerned with asymptotics in this paper we state the following result without a formal proof: The lower bound for the asymptotic minimax risk for $N \rightarrow \infty$ that was obtained by Vogel in [3] may be raised by a factor $\sqrt{2}$. The result is proved by applying the same method that was used in [3] to the optimal symmetric strategy for $\alpha + \beta = 1$ that was discussed in [2]. Combining this lower bound with the upper bound given in [3] we find that the asymptotic minimax risk must be between $0.265 N^{\frac{1}{2}}$ and $0.376 N^{\frac{1}{2}}$.

2 A RECURRENCE RELATION

For $\Delta \in \mathcal{D}$ we consider the expected number of successes $E(S \mid \alpha, \beta, \Delta)$ as a function of the $\delta(m,k;n,1)$. Clearly, the dependence on each $\delta(m,k;n,1)$ is linear.

We denote the coefficient of $\delta(m,k;n,1)$ in $E(S \mid \alpha, \beta, \Delta)$ (and hence also in $-R(\alpha, \beta, \Delta)$) by $p_{\Delta}(m,k;n,1)c_{\alpha, \beta, \Delta}(m,k;n,1)$. If all $\delta(m,k;n,1)$ are strictly between 0 and 1, then all $p_{\Delta}(m,k;n,1)$ are positive and as a result all $c_{\alpha, \beta, \Delta}(m,k;n,1)$ are uniquely determined. Otherwise the $c_{\alpha, \beta, \Delta}(m,k;n,1)$ are defined by continuity.

THEOREM 2

For any strategy Δ in $\mathcal{D}$ the functions $c_{\alpha, \beta, \Delta}(m,k;n,1)$ satisfy the following relations.

$$(2.1) \quad c_{\alpha, \beta, \Delta}(m,k;n,1) = (\alpha - \beta) \alpha^k (1 - \alpha)^{m-k} \beta^l (1 - \beta)^{n-l} \\ \text{if } m + n = N - 1,$$

$$(2.2) \quad c_{\alpha, \beta, \Delta}(m,k;n,1) = \delta(m+1, k+1; n, 1) c_{\alpha, \beta, \Delta}(m+1, k+1; n, 1) + \\ + \delta(m+1, k; n, 1) c_{\alpha, \beta, \Delta}(m+1, k; n, 1) + \\ + [1 - \delta(m, k; n+1, 1+1)] c_{\alpha, \beta, \Delta}(m, k; n+1, 1+1) + \\ + [1 - \delta(m, k; n+1, 1)] c_{\alpha, \beta, \Delta}(m, k; n+1, 1) \\ \text{if } m + n \leq N - 2.$$

PROOF

By continuity it is obviously sufficient to consider the case where all $\delta(m,k;n,1)$ as well as α and β are strictly between 0 and 1. This ensures that expression (1.3) is positive for all states $(m,k;n,1)$. Hence the conditional expectation $e_{\alpha, \beta, \Delta}(m,k;n,1)$ of the total number of successes S under α, β and Δ given that the state $(m,k;n,1)$ is reached, exists.

It is clearly a linear function of $\delta(m,k;n,1)$ and may thus be written in the form

$$(2.3) \quad e_{\alpha,\beta,\Delta}(m,k;n,1) = a_{\alpha,\beta,\Delta}(m,k;n,1) \delta(m,k;n,1) + b_{\alpha,\beta,\Delta}(m,k;n,1).$$

It follows that

$$(2.4) \quad c_{\alpha,\beta,\Delta}(m,k;n,1) = a_{\alpha,\beta,\Delta}(m,k;n,1) \alpha^k (1 - \alpha)^{m-k} \beta^l (1 - \beta)^{n-l}.$$

Dropping the subscripts α , β and Δ , we obtain, from the definition of $e(m,k;n,1)$,

$$\begin{aligned} e(m,k;n,1) &= \delta(m,k;n,1) [\alpha e(m+1,k+1;n,1) + (1-\alpha)e(m+1,k;n,1)] + \\ &\quad + [1 - \delta(m,k;n,1)] [\beta e(m,k;n+1,1) + (1-\beta)e(m,k;n+1,1)], \end{aligned}$$

and consequently

$$(2.5) \quad \begin{aligned} a(m,k;n,1) &= \alpha e(m+1,k+1;n,1) + (1-\alpha)e(m+1,k;n,1) + \\ &\quad - \beta e(m,k;n+1,1) - (1-\beta)e(m,k;n+1,1), \end{aligned}$$

$$(2.6) \quad b(m,k;n,1) = \beta e(m,k;n+1,1) + (1-\beta)e(m,k;n+1,1).$$

If $m+n = N-1$, then (2.5) becomes $a(m,k;n,1) = \alpha - \beta$, and hence (2.1) follows from (2.4). On the other hand, rewriting (2.5) by means of (2.3) leads to

$$\begin{aligned} (2.7) \quad a(m,k;n,1) &= \alpha \delta(m+1,k+1;n,1) a(m+1,k+1;n,1) + \\ &\quad + (1-\alpha) \delta(m+1,k;n,1) a(m+1,k;n,1) + \\ &\quad + \beta [1 - \delta(m,k;n+1,1)] a(m,k;n+1,1) + \\ &\quad + (1-\beta) [1 - \delta(m,k;n+1,1)] a(m,k;n+1,1) + \\ &\quad + [\alpha b(m+1,k+1;n,1) + (1-\alpha)b(m+1,k;n,1) + \\ &\quad - \beta b(m,k;n+1,1) + \\ &\quad - (1-\beta)b(m,k;n+1,1) - \beta a(m,k;n+1,1) + \\ &\quad - (1-\beta)a(m,k;n+1,1)], \end{aligned}$$

where for $m + n \leq N - 2$ the last expression between square brackets vanishes as one easily verifies using (2.5) and (2.6). This result, combined with (2.4), gives (2.2).

Let μ be a prior distribution on the closed unit square. For a strategy $\Delta \in \mathcal{D}$,

$$(2.8) \quad \rho(\mu, \Delta) = \int R(\alpha, \beta, \Delta) d\mu(\alpha, \beta) \quad \text{denotes the average risk of}$$

Δ against μ . If we define

$$(2.9) \quad \gamma_{\mu, \Delta}(m, k; n, l) = \int c_{\alpha, \beta, \Delta}(m, k; n, l) d\mu(\alpha, \beta), \quad \text{then}$$

$-p_{\Delta}(m, k; n, l) \gamma_{\mu, \Delta}(m, k; n, l)$ is the coefficient of $\delta(m, k; n, l)$ in $\rho(\mu, \Delta)$.

It follows that any strategy Δ that has $\delta(m, k; n, l) = 1$ whenever

$\gamma_{\mu, \Delta}(m, k; n, l) > 0$ and $\delta(m, k; n, l) = 0$ whenever $\gamma_{\mu, \Delta}(m, k; n, l) < 0$,

minimizes $\rho(\mu, \Delta)$ for fixed μ and is therefore a Bayes strategy against μ .

This may be seen by successively finding the optimal $\delta(m, k; n, l)$ for

$m + n = N - 1, N - 2, \dots, 0$, and noting that for $m + n = v$ these

optimal values do not depend on the values of $\delta(m, k; n, l)$ for $m + n < v$.

Conversely, every Bayes strategy against μ has a version with

$\delta(m, k; n, l) = 1$ (or 0) whenever $\gamma_{\mu, \Delta}(m, k; n, l) > 0$ (or < 0).

THEOREM 3

Let μ be a prior distribution on the closed unit square and let

$\gamma_{\mu}(m, k; n, l)$ be defined by

$$(2.10) \quad \gamma_{\mu}(m, k; n, l) = \int (\alpha - \beta) \alpha^k (1 - \alpha)^{m-k} \beta^l (1 - \beta)^{n-l} d\mu(\alpha, \beta)$$

if $m + n = N - 1$,

$$(2.11) \quad \gamma_{\mu}(m, k; n, l) = \gamma_{\mu}^{+}(m+1, k+1; n, l) + \gamma_{\mu}^{+}(m+1, k; n, l) + \\ - \gamma_{\mu}^{-}(m, k; n+1, l+1) - \gamma_{\mu}^{-}(m, k; n+1, l)$$

for $m+n \leq N-2$, where x^{+} and x^{-} denote $\max(0, x)$ and $\max(0, -x)$ respectively. Then $\Delta \in \mathcal{D}$ is a Bayes strategy against μ if and only if it has a version with $\delta(m, k; n, l) = 1$ whenever $\gamma_{\mu}(m, k; n, l) > 0$ and $\delta(m, k; n, l) = 0$ whenever $\gamma_{\mu}(m, k; n, l) < 0$.

PROOF

According to the remarks preceding the theorem, Δ is Bayes against μ iff it has a version for which $\delta(m, k; n, l) = 1$ (or 0) if $\gamma_{\mu, \Delta}(m, k; n, l) > 0$ (or < 0). Integrating (2.1) and (2.2) with respect to μ and substituting the values of the $\delta(m, k; n, l)$ we find that for this version of Δ , $\gamma_{\mu, \Delta}(m, k; n, l)$ equals $\gamma_{\mu}(m, k; n, l)$ as defined by (2.10) and (2.11) for all states.

3 ADMISSIBLE STRATEGIES

For the type of problem considered in this paper every admissible strategy is also a Bayes strategy. In the sequel we shall, however, need a slightly stronger result. We shall say that a prior distribution is nonmarginal if, for some $\epsilon > 0$, it assigns probability 1 to the set

$$(3.1) \quad Q_{\epsilon} = \{(\alpha, \beta) \mid |\alpha - \beta| \alpha(1 - \alpha)\beta(1 - \beta) \geq \epsilon, 0 < \alpha < 1, 0 < \beta < 1\}.$$

THEOREM 4

Every admissible strategy $\Delta \in \mathcal{D}$ is Bayes against a nonmarginal prior distribution.

PROOF

Let Δ be a Bayes strategy against a prior distribution μ on the closed unit square and suppose that Δ is not Bayes against any nonmarginal prior. It is sufficient to show that Δ is not admissible.

For any sufficiently small $\epsilon_i > 0$, consider the restricted problem where the parameter space is reduced to the set $A_i = Q_{\epsilon_i}$ as defined by (3.1). Since A_i is compact, the assertion that every admissible strategy is Bayes remains true for the restricted problem. By our assumption Δ is not Bayes, and therefore not admissible in the new problem. It follows that there exists a strategy Δ_i that is Bayes against a prior distribution μ_i on A_i and for which

$$R(\alpha, \beta, \Delta_i) \leq R(\alpha, \beta, \Delta) \quad \text{for all } (\alpha, \beta) \in A_i.$$

By a standard procedure we may select a sequence $\epsilon_i \searrow 0$ and corresponding μ_i and Δ_i such that the strategies Δ_i converge to a strategy Δ_0 in the sense that $\delta_i(m, k; n, l)$ converges to $\delta_0(m, k; n, l)$ for every state $(m, k; n, l)$. Obviously

$$R(\alpha, \beta, \Delta_0) \leq R(\alpha, \beta, \Delta) \quad \text{for all } \alpha, \beta \in [0, 1],$$

since the inequality must hold on every A_i and both functions are continuous.

Since Δ_i converges to Δ_0 there exists a positive integer j for which Δ_j has the following properties:

- (a) For all states with $\delta_0(m, k; n, l) = 0$, $\delta_j(m, k; n, l) \neq 1$;
 - (b) For all states with $\delta_0(m, k; n, l) = 1$, $\delta_j(m, k; n, l) \neq 0$;
 - (c) For all states with $0 < \delta_0(m, k; n, l) < 1$, $0 < \delta_j(m, k; n, l) < 1$.
- This implies that $\delta_0(m, k; n, l) = \delta_j(m, k; n, l)$ for every state with $\delta_j(m, k; n, l) = 0$ or 1 .

Recalling that Δ_j is Bayes against μ_j and noting that this property can not be destroyed by changing only those $\delta_j(m,k;n,1)$ that are strictly between 0 and 1, we find that Δ_0 is Bayes against the prior distribution μ_j on A_j . As Δ is not Bayes against μ_j by our assumption, the inequality $R(\alpha, \beta, \Delta_0) \leq R(\alpha, \beta, \Delta)$ on the closed unit square must be strict for at least one point (α, β) and the inadmissibility of Δ follows.

We are now in a position to prove a theorem that provides some insight in the structure of admissible strategies.

THEOREM 5

If μ is a nonmarginal prior distribution and $m + n \leq N - 2$, then

$$(3.2) \quad \gamma_{\mu}(m, k; n + 1, l + 1) < \gamma_{\mu}(m + 1, k + 1; n, l)$$

$$(3.3) \quad \gamma_{\mu}(m + 1, k; n, l) < \gamma_{\mu}(m, k; n + 1, l) .$$

PROOF

For $m + n = N - 2$, (2.10) yields

$$\begin{aligned} \gamma_{\mu}(m + 1, k + 1; n, l) - \gamma_{\mu}(m, k; n + 1, l + 1) &= \\ &= \int (\alpha - \beta)^2 \alpha^k (1 - \alpha)^{m - k} \beta^l (1 - \beta)^{n - l} d\mu(\alpha, \beta) , \end{aligned}$$

which is strictly positive since μ is nonmarginal. In the same way one shows that (3.3) is satisfied for $m + n = N - 2$.

Next we suppose that the theorem is valid for $m + n = v$, where $0 < v \leq N - 2$, and we assume $m + n = v - 1$.

By (2.11) we have then

$$\begin{aligned}
 \gamma_{\mu}(m+1, k+1; n, 1) - \gamma_{\mu}(m, k; n+1, 1+1) = \\
 &= [\gamma_{\mu}^{+}(m+2, k+2; n, 1) - \gamma_{\mu}^{+}(m+1, k+1; n+1, 1+1)] + \\
 &+ [\gamma_{\mu}^{+}(m+2, k+1; n, 1) - \gamma_{\mu}^{+}(m+1, k; n+1, 1+1)] + \\
 &+ [\gamma_{\mu}^{-}(m, k; n+2, 1+2) - \gamma_{\mu}^{-}(m+1, k+1; n+1, 1+1)] + \\
 &+ [\gamma_{\mu}^{-}(m, k; n+2, 1+1) - \gamma_{\mu}^{-}(m+1, k+1; n+1, 1)] \geq 0,
 \end{aligned}$$

since by hypothesis each of the four expressions is nonnegative. Equality can occur only if all four expressions vanish. However, the first and the third one can vanish only if $\gamma_{\mu}(m+1, k+1; n+1, 1+1) < 0$ and ≥ 0 respectively, and hence inequality (3.2) is strict.

Similarly (3.3) follows from

$$\begin{aligned}
 \gamma_{\mu}(m, k; n+1, 1) - \gamma_{\mu}(m+1, k; n, 1) = \\
 &= [\gamma_{\mu}^{+}(m+1, k+1; n+1, 1) - \gamma_{\mu}^{+}(m+2, k+1; n, 1)] + \\
 &+ [\gamma_{\mu}^{+}(m+1, k; n+1, 1) - \gamma_{\mu}^{+}(m+2, k; n, 1)] + \\
 &+ [\gamma_{\mu}^{-}(m+1, k; n+1, 1+1) - \gamma_{\mu}^{-}(m, k; n+2, 1+1)] + \\
 &+ [\gamma_{\mu}^{-}(m+1, k; n+1, 1) - \gamma_{\mu}^{-}(m, k; n+2, 1)] \geq 0,
 \end{aligned}$$

and the fact that the first expression in square brackets can vanish only if $\gamma_{\mu}(m+2, k+1; n, 1) < 0$ and the third one only if

$\gamma_{\mu}(m+1, k; n+1, 1+1) \geq 0$, which would imply $\gamma_{\mu}(m+2, k+1; n, 1) > 0$.

COROLLARY 1

Every admissible strategy $\Delta \in \mathcal{D}$ has a version for which

$$(3.4) \quad \delta(m, k; n + 1, l + 1) \leq \delta(m + 1, k + 1; n, l)$$

$$(3.5) \quad \delta(m + 1, k; n, l) \leq \delta(m, k; n + 1, l)$$

for all $m + n \leq N - 2$, where in each of these inequalities at least one member equals 0 or 1.

PROOF

By theorem 4, Δ is Bayes against a nonmarginal prior μ , and as a result the theorem is proved by applying theorems 5 and 3.

COROLLARY 2

Every admissible strategy $\Delta \in \mathcal{D}$ has a version for which

$$(3.6) \quad \delta(m, k; n, l) [1 - \delta(m + 1, k + 1; n, l)] [1 - \delta(m + 1, k; n, l)] = 0$$

$$(3.7) \quad [1 - \delta(m, k; n, l)] \delta(m, k; n + 1, l + 1) \delta(m, k; n + 1, l) = 0$$

for all $m + n \leq N - 2$.

PROOF

As before, we let μ denote the nonmarginal prior of theorem 4 and consider the version of Δ having $\delta(m, k; n, l) = 1$ (or 0) whenever $\gamma_\mu(m, k; n, l) > 0$ (or < 0). If (3.6) were false for this version, then $\gamma_\mu(m, k; n, l) \geq 0$, $\gamma_\mu(m + 1, k + 1; n, l) \leq 0$ and $\gamma_\mu(m + 1, k; n, l) \leq 0$. The second of these inequalities implies $\gamma_\mu(m, k; n + 1, l + 1) < 0$ by theorem 5, and hence (2.11) shows that $\gamma_\mu(m, k; n, l) < 0$, which contradicts the first inequality.

Similarly, if (3.7) were false, then $\gamma_\mu(m, k; n, l) \leq 0$, $\gamma_\mu(m, k; n + 1, l + 1) \geq 0$ and $\gamma_\mu(m, k; n + 1, l) > 0$.

The second inequality implies $\gamma_{\mu}(m+1, k+1; n, 1) > 0$ by theorem 5, and hence $\gamma_{\mu}(m, k; n, 1) > 0$ by (2.11), which contradicts the first inequality. This completes the proof.

For symmetric strategies a more explicit result may be obtained.

COROLLARY 3

Every admissible strategy $\Delta \in \mathcal{J}$ has a version for which

$$(3.8) \quad \delta(m, k; n, 1) = 1, \quad \delta(n, 1; m, k) = 0$$

whenever $m + n \leq N - 1$, $k \geq 1$, $m - k \leq n - 1$ and $(m, k; n, 1) \neq (n, 1; m, k)$.

PROOF

For the version of Δ that satisfies corollary 1 we find by repeated application of (3.4) and (3.5)

$$\delta(m, k; n, 1) \geq \delta(m - k + 1, 1; n + k - 1, k) \geq \delta(n, 1; m, k)$$

where at least one of the extreme members must be 0 or 1. Since their sum equals 1 if $p_{\Delta}(m, k; n, 1) \neq 0$, (3.8) will hold in this case. If $p_{\Delta}(m, k; n, 1) = 0$, then by (1.6) we also have $p_{\Delta}(n, 1; m, k) = 0$ and choosing $\delta(m, k; n, 1) = 1$ and $\delta(n, 1; m, k) = 0$ merely leads to another version of Δ .

We conclude this section by remarking that corollaries 1, 2 and 3 obviously continue to hold if, instead of admissibility, we require that Δ be Bayes against a nonmarginal prior.

4 SYMMETRIC MINIMAX-RISK STRATEGIES

In section 1 we have shown that there exists a symmetric minimax-risk strategy. For the type of problem considered in this paper there exists a least favourable prior distribution and any minimax-risk strategy is Bayes against any least favourable prior. These assertions continue to hold if the parameter space is reduced to a compact subset of the closed unit square.

THEOREM 6

There exists a minimax-risk strategy $\Delta \in \mathcal{J}$ which obeys (3.4) through (3.8).

PROOF

By the remark at the end of section 3, it is sufficient to demonstrate the existence of a symmetric minimax-risk strategy that is Bayes against a nonmarginal prior.

For sufficiently small $\varepsilon_i > 0$ let $A_i = Q_{\varepsilon_i}$ as defined by (3.1) and let μ_i and Δ_i denote a least favourable prior and a symmetric minimax-risk strategy for the restricted problem where the parameter space is reduced to the compact set A_i . Repeating the proof of theorem 4 we may select a sequence $\varepsilon_i \searrow 0$ and corresponding μ_i and Δ_i such that the strategies Δ_i converge to a strategy Δ_0 that is Bayes against a nonmarginal prior μ_j on A_j . Since the convergence is defined as convergence of the $\delta_i(m,k;n,1)$ to the $\delta_0(m,k;n,1)$, Δ_0 is symmetric. As the maximum risk of Δ_i on A_i does not exceed the minimax risk on the entire closed unit square and $R(\alpha, \beta, \Delta_0)$ is continuous, the convergence

of Δ_1 to Δ_0 implies that Δ_0 has minimax risk.

For $N = 1$ or 2 , (1.5) and (3.8) uniquely determine a symmetric strategy. It follows from theorem 6 and corollary 3 that this strategy has minimax risk and is in fact the only admissible strategy in $\mathcal{J}$. For $N \geq 3$ the situation rapidly becomes more complicated. In order to find a symmetric minimax-risk strategy Δ_0 satisfying (3.4) through (3.8) one first has to find a general expression for the risk function $R(\alpha, \beta, \Delta)$ of an arbitrary symmetric strategy Δ satisfying (3.8). Then, with the aid of (3.4) through (3.7), one has to solve the remaining $\delta(m, k; n, 1)$ directly using the minimax property.

To accomplish the first step of computing $R(\alpha, \beta, \Delta)$ for an arbitrary symmetric strategy, one may proceed recursively. This is especially useful if one wants to find $R(\alpha, \beta, \Delta)$ for a number of values of N . If $X_v = 1 - Y_v = 1$ or 0 according to whether E_1 or E_2 is carried out on the v -th trial ($v = 1, 2, \dots, N$), then $R(\alpha, \beta, \Delta)$, being equal to $|\alpha - \beta|$ multiplied by the expected number of times the experimenter uses the less favourable experiment, is given by

$$(4.1) \quad R(\alpha, \beta, \Delta) = \frac{1}{2}N |\alpha - \beta| - \frac{1}{2}(\alpha - \beta) \sum_{v=1}^N E(X_v - Y_v | \alpha, \beta, \Delta) .$$

Remembering the definition of $\pi_{\alpha, \beta, \Delta}(m, k; n, 1)$, we have

$$(4.2) \quad E(X_v - Y_v | \alpha, \beta, \Delta) = \sum \pi_{\alpha, \beta, \Delta}(m, k; n, 1) [2\delta(m, k; n, 1) - 1] ,$$

where the summation is extended over all states $(m, k; n, 1)$ with $m + n = v - 1$, and where the $\pi_{\alpha, \beta, \Delta}(m, k; n, 1)$ can be computed recursively by means of

$$\begin{aligned}
 (4.3) \quad \pi_{\alpha, \beta, \Delta}(m, k; n, 1) = & \alpha \delta(m-1, k-1; n, 1) \pi_{\alpha, \beta, \Delta}(m-1, k-1; n, 1) + \\
 & + (1-\alpha) \delta(m-1, k; n, 1) \pi_{\alpha, \beta, \Delta}(m-1, k; n, 1) + \\
 & + \beta [1 - \delta(m, k; n-1, 1-1)] \pi_{\alpha, \beta, \Delta}(m, k; n-1, 1-1) + \\
 & + (1-\beta) [1 - \delta(m, k; n-1, 1)] \pi_{\alpha, \beta, \Delta}(m, k; n-1, 1)
 \end{aligned}$$

starting from

$$(4.4) \quad \pi_{\alpha, \beta, \Delta}(0, k; 0, 1) = \begin{cases} 1 & \text{if } k = 1 = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The work involved may be reduced somewhat by means of the relation

$$(4.5) \quad \pi_{\alpha, \beta, \Delta}(m, k; n, 1) = \pi_{\beta, \alpha, \Delta}(n, 1; m, k),$$

which is a consequence of (1.3) and (1.6).

For $N = 3$, only $\delta(2, 1; 0, 0)$ remains undetermined by the requirement that Δ be symmetric and must satisfy (3.8), and one finds

$$R(\alpha, \beta, \Delta) = \frac{3}{2} |\alpha - \beta| - \frac{1}{2} (\alpha - \beta)^2 \{1 + \delta(2, 1; 0, 0) + [1 - \delta(2, 1; 0, 0)](\alpha + \beta)\}.$$

After a little algebra one sees that Δ_0 must have $\delta(2, 1; 0, 0) = 1$ and that $R(\alpha, \beta, \Delta_0)$ attains its maximum $M(\Delta_0) = 9/16$ when $|\alpha - \beta| = 3/4$.

For $N = 4$ only $\delta(2, 1; 0, 0)$, $\delta(3, 1; 0, 0)$ and $\delta(3, 2; 0, 0)$ are to be determined and

$$\begin{aligned}
 R(\alpha, \beta, \Delta) = & 2 |\alpha - \beta| - \frac{1}{2} (\alpha - \beta)^2 \{(\alpha^2 + \beta^2 + 3\alpha\beta - \alpha - \beta + 3) - \delta(2, 1; 0, 0)\alpha\beta + \\
 & - \delta(3, 2; 0, 0)[1 + \delta(2, 1; 0, 0)](\alpha^2 + \beta^2 + \alpha\beta - \alpha - \beta) + \\
 & + \delta(3, 1; 0, 0) \delta(2, 1; 0, 0) (\alpha^2 + \beta^2 + \alpha\beta - 2\alpha - 2\beta + 1)\}.
 \end{aligned}$$

Using (3.6), one finds after lengthy calculations that Δ_0 must have $\delta(2, 1; 0, 0) = 4/5$, $\delta(3, 1; 0, 0) = 1/2$ and $\delta(3, 2; 0, 0) = 1$, so that the riskfunction of Δ_0 is given by

$$R(\alpha, \beta, \Delta_0) = 2 |\alpha - \beta| - \frac{17}{10} (\alpha - \beta)^2 + \frac{1}{5} (\alpha - \beta)^4$$

and attains its maximum $M(\Delta_0) = .617$ when $|\alpha - \beta| = .654$. For larger values of N the number of $\delta(m, k; n, 1)$ that have to be determined increases rapidly, and consequently the algebra involved becomes distressingly complicated.

REFERENCES

- [1] 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.
- [2] VOGEL, W. (1960). Ein Irrfahrten-Problem und seine Anwendung
auf die Theorie der sequentiellen Versuchs-Pläne.
Arch. Math. 11 , 310 - 320.
- [3] VOGEL, W. (1960). An asymptotic minimax theorem for the two
armed bandit problem. Ann. Math. Statist. 31 , 444 - 451.