A BAYES MODEL IN SEQUENTIAL DESIGN

Samuel Karlin and S. M. Johnson

P-328 Rev.

4 January 1954

COPY	OF	
HARD COPY	$.2.00	
MICROFICHE	$.0.50	

The RAND *Corporation*

1700 MAIN ST. · SANTA MONICA · CALIFORNIA

A BAYES MODEL IN SEQUENTIAL DESIGN

Samuel Karlin and S. M. Johnson

Introduction

This paper is concerned with the Bayes problem of how to maximize the expected number of successes in N trials when at each trial we are free to choose between two machines I and II whose probabilities ρ, σ of success are unknown but have a known a priori distribution $F(\rho, \sigma)$.

We have adopted the use of the terms machine I and II to expedite the discussion. Many other interpretations and applications can be found for the theory developed below. This is a type of problem classified as sequential design. No nontrivial examples in this field have been analyzed as far as we know to the present date, and this represents an attempt to study some models, to develop some qualitative results, and to focus attention on some of the difficulties involved by suitable examples. One particular model is analyzed completely in §3. It is interesting to note that none of the intuitive simple strategies are usually optimal and, moreover, that the optimal strategies in general seem to be of a very complicated nature. However, approximate optimal strategies are discussed in several contexts.

In §2 we have analyzed the relevance of the strategy S_o which employs at each stage the machine with the maximum a priori expected value. It is shown that this strategy is rarely optimal. Other features suspected about optimal strategies are exploded. §3 deals with the case where one machine has a known probability

of success while the other has only a known a priori distribution. This case is handled completely and serves to illuminate the complex nature of optimal strategies. §4 treats of certain game extensions associated with the Bayes problem.

§1. The General Formulation

Let S denote any strategy for choosing between the two machines and let $V_S(\rho, \sigma)$ denote the expected number of successes following strategy S for given (ρ, σ). Then the expected number of successes based on policy S is

(1) $$\phi_S(F) = \int \int V_S(\rho, \sigma) dF(\rho, \sigma) .$$

The best procedure is the one maximizing $\phi_S(F)$. Since N is finite, the maximum is well defined.

In computing $\phi_S(F)$ one can extract the following formal procedure. We determine the conditional a priori distribution of (ρ, σ) on the k—th trial, given that s_1 successes and f_1 failures from I, and s_2 successes and f_2 failures from II, with $s_1 + f_1 + s_2 + f_2 = k - 1$, have preceded. In fact,

Prob. of success $(\rho, \sigma \mid s_1, f_1, s_2, f_2)$

$$= \frac{Pr(s_1, f_1, s_2, f_2 | \rho, \sigma) Pr(\rho, \sigma)}{Pr(s_1, f_1, s_2, f_2)} .$$

We thus obtain as the a posteriori distribution

$$(2) \quad dF(\rho,\sigma)|s_1 f_1 s_2 f_2) = \frac{\rho^{s_1}(1-\rho)^{f_1}\sigma^{s_2}(1-\sigma)^{f_2}dF(\rho,\sigma)}{\int_0^1 \int_0^1 \rho^{s_1}(1-\rho)^{f_1}\sigma^{s_2}(1-\sigma)^{f_2}dF(\rho,\sigma)} \quad .$$

The contribution to $\bar{\Phi}_S(F)$ then becomes the first ρ or σ moment of the distribution (2) according as to whether I or II is used at the k—th stage. This a posteriori distribution (2) is inde- pendent of the order of presentation of the information as can be easily verified.

One example of a very natural strategy S is the principle: maximize expected value at each stage. Precisely, the quantities $\int \rho \, dF(\rho,\sigma)$ and $\int \sigma dF(\rho,\sigma)$ are compared and machine I is chosen over machine II depending on whether the first integral exceeds the second. The outcome of the first trial then determines an a posteriori distribution F' for which the same criterion on the first moments of F' indicates the machine to be played for the second step, etc. This particular strategy we shall call the "stagewise maximization principle" and designate it by S_o. In the following simple example, S_o is optimal.

Example 1. If $\rho + \sigma = 1$, then $F(\rho, \sigma)$ is of the form $F(\rho, 1-\rho)$. Thus a success or failure on I is equivalent to (gives the same information as) a failure or success on II, respectively.

Let $X = \rho^{s_1}(1 - \rho)^{f_1}\sigma^{s_2}(1 - \sigma)^{f_2}$ and write dF for $dF(\rho, \sigma)$. Let $E_k(X)$ be the maximum expected number of successes when playing optimally for k more trials given the history indicated by X. Then

$$\int X\rho dF + (\int X\rho dF)E_{k-1}(X\rho) + (\int(1-\rho)XdF)E_{k-1}(X(1-\rho))$$

$$\geq \int X\sigma dF + (\int X\sigma dF)E_{k-1}(X\sigma) + (\int X(1-\sigma)dF)E_{k-1}(X(1-\sigma))$$

if and only if $\int X\rho dF \geq \int X\sigma dF$ as $\sigma = 1 - \rho$; that is, S_0 is the optimal policy.

§2. Qualitative Results about Optimal Procedures

Our first task is to obtain the complete procedure for $N = 2$ when $F(\rho, \sigma) = F(\rho)G(\sigma)$. It is important to emphasize that if the number of moves is n, then only strategies which are functions of the first n moments $\mu_1, \cdots, \mu_n$ of F and $\mu_1', \mu_2', \cdots, \mu_n'$ of G need to be considered. This is a conse-quence of the fact that the expected yield for any given strategy is an expression involving at most these moments. Thus all strategies describing a first move can be viewed as functions $S_1(\mu_1, \cdots, \mu_n, \mu_1', \cdots, \mu_n')$ such that if $S_1(\mu_1, \cdots, \mu_n, \mu_1', \cdots, \mu_n') \geq 0$, then I is chosen at the first stage and II in the contrary case. Let $\mu_1 = \int_0^1 \rho^1 dF(\rho)$ and $\mu_1' = \int_0^1 \sigma^1 dG(\sigma)$. Suppose for definiteness, $\mu_1 \geq \mu_1'$; we determine necessary and sufficient conditions that I is employed first when $N = 2$. Using the fact that at the last step one maximizes expected value, we secure in this circumstance the value

$$(1) \qquad \mu_1 + \mu_1(\frac{\mu_2}{\mu_1}) + (1 - \mu_1) \max(\mu_1', \frac{\mu_1 - \mu_2}{1 - \mu_1}) .$$

We now obtain the best possible value attainable if machine II is used first. The strategy of playing II at the first step and then changing regardless of the outcome is dominated by (1). Indeed, since $\mu_1 \geq \mu_1'$, (1) is greater than or equal to $2\mu_1 \geq \mu_1 + \mu_1'$, the value obtained according to the above strategy.

Consequently, $\mu_1' + \mu_2' + (1 - \mu_1')\mu_1$ is the only attainable value needing to be considered if one begins with II. Choosing each term of the maximum of (1) yields the two inequalities

$$\mu_2 \geq \mu_2'$$

(2)

$$\mu_1 + \mu_1\mu_1' \geq \mu_1' + \mu_2' .$$

Combining and rewriting in a symmetric form, we have

Lemma 1. If N = 2 and the machines have independent a priori distributions of probabilities of success with moments μ_1 and μ_1', then a necessary and sufficient condition that machine II is played first is that

$$\text{Max}(\mu_2' - \mu_1\mu_1', \ \mu_1' - \mu_1) \geq \text{Max}(\mu_2 - \mu_1'\mu_1, \ \mu_1 - \mu_1') .$$

The next theorem shows that S_o is generally not optimal.

Theorem 1. If machines I and II have a priori distributions $F(\rho) = \int_o^\rho \phi(t)dt$ and $G(\sigma) = \int_o^\sigma \psi(t)dt$ respectively for the

probabilities of success where $\phi(t)$ and $\psi(t)$ are continuous and positive for $0 < t < 1$, then there exists an n so that for n trials the optimal procedure does not agree with the strategy S_o described by stagewise maximization.

Proof: (By contradiction) Suppose for definiteness that
$$1 > \int_0^1 t\phi(t)dt = b \geq a = \int_0^1 \psi(t)dt > 0,$$ then clearly at the first trial we use I. According to the strategy S_o, it is clear by the Schwartz inequality that we stay with the machine being used whenever success occurs. It is easily shown in view of the hypothesis on $\phi(t)$ that if $\frac{r}{r+s} \to t_o$, then

$$\frac{\int_0^1 t^{r+1}(1-t)^s\phi(t)dt}{\int_0^1 t^r(1-t)^s\phi(t)dt} \to t_o .$$

This can also be obtained as a consequence of the law of large numbers where the frequency of success tends to t_o.

We choose $r, s \to \infty$ sufficiently large and $\left| \frac{r}{r+s} - a \right| < \epsilon$ so that

$$a + \epsilon \geq \frac{\int t^{r+1}(1-t)^s\phi(t)dt}{\int t^r(1-t)^s\phi(t)dt} > a = \int_0^1 t\psi(t)dt, \text{ and}$$

(3)

$$a + \epsilon \geq \frac{\int t^{r+2}(1-t)^s\phi(t)dt}{\int t^{r+1}(1-t)^s\phi(t)dt} > a .$$

The approximation from above is easy to insure by approximating to a + ϵ/2. Furthermore, ϵ is chosen sufficiently small so that the above holds and also $\int t^2 \psi(t)dt > \left(\int t \psi(t)dt \right)^2 + 3\epsilon$. Let n = r + s + 2. Suppose that using I first resulted in r consecutive successes and then s failures. This agrees with the procedure prescribed by strategy S_o and this situation occurs with positive probability. At the r + s + 1 step, in view of the first equation of (3), one should continue with I according to strategy S_o (maximization stagewise). We now show that both inequalities of (2) are violated and thus machine II should be used to furnish an optimal return. Indeed, the distributions of successes at the beginning of the r + s + 1 step are, in this situation

$$dF^1(\rho) = \frac{\rho^r(1-\rho)^s \phi(\rho)d\rho}{\int_0^1 \rho^r(1-\rho)^s \phi(\rho)d\rho}$$

and

$$G(\sigma) = \int_0^\sigma \psi(\sigma)d\sigma.$$

On account of (3)

$$\int_0^1 \rho \, dF^1(\rho) > \int_0^1 \sigma \psi(\sigma)d\sigma.$$

But

$$\int \rho^2 dF^1(\rho) = \frac{\int \rho^{r+2}(1-\rho)^s\phi(\rho)d\rho}{\int \rho^r(1-\rho)^s\phi(\rho)d\rho} = \frac{\int \rho^{r+2}(1-\rho)^s\phi(\rho)d\rho}{\int \rho^{r+1}(1-\rho)^s\phi(\rho)d\rho}\frac{\int \rho^{r+1}(1-\rho)^s\phi(\rho)d\rho}{\int \rho^r(1-\rho)^s\phi(\rho)d\rho}$$

$$\leq (a + \epsilon)^2 < a^2 + 3\epsilon < \int_0^1 t^2 \psi(t)dt \quad .$$

Also,

$$\int \rho\, dF^1(\rho) + \int \rho\, dF^1(\rho) \int \sigma \psi(\sigma)d\sigma < a + \epsilon + a^2 + \epsilon$$

$$< \int t\psi(t)dt + \int t^2\psi(t)dt \quad .$$

Hence, following S_o we arrive at a nonoptimal yield and the theorem is established.

We further remark that Theorem 1 can be established for almost all pairs of independent distributions. Only in trivial cases where the a posteriori distribution of I for any possible outcomes will always have larger expected value than that of II, will it be true for all n that the principle of "maximization stepwise" agrees with the optimal procedure. We have chosen to illustrate this theorem by the class of distributions considered above in order to avoid some trivial technical difficulties. In the case where $F(\rho) = G(\sigma)$ and with $F(\rho)$ symmetric, i.e., $1 - F(1-\rho) = F(\rho)$, then Theorem 1 is valid in many instances with n = 4. It is clearly immaterial which is played first as the expected yield is the same and equal to 1/2. Furthermore, if a success occurs then it is optimal based on the principle

of S_0 to continue with the same choice. We suppose now that a failure occurs on the second trial; then there remain two trials with distributions $\dfrac{\rho(1-\rho)dF(\rho)}{\int\rho(1-\rho)dF(\rho)}$ and $dF(\rho)$, respectively, for I and II. Both are still symmetric and hence possess an expected value equal to $1/2$. Thus, according to the strategy which maxi-mizes stepwise, it makes no difference which machine is tried at this third step. This implies according to Lemma 1 that

$$\mu_2 = \frac{\mu_3 - \mu_4}{\mu_1 - \mu_2} \text{ or } \mu_1\mu_2 + \mu_4 = \mu_3 + \mu_2^2 .$$ This is generally impossible, particularly, e.g., when $dF(\rho) = C\rho^{\alpha}(1-\rho)^{\alpha}d\rho$.

The strategy S_0 can be described as the procedure which makes that choice at each trial which would have been optimal had there been only one trial left.

Let T_j be the strategy making the choice at each stage which would have been optimal if there were j trials left with the understanding that when fewer than j trials remain then the optimal procedure is followed thereafter.

Thus $T_1 = S_0$, and the strategy T_2 for independent distribu-tions is determined by the relations of Lemma 1.

In this way we can obtain a whole hierarchy of strategies T_j, $j = 1, 2, \cdots, N$. Intuitively one might expect that these strategies are successive improvements. Of course, when $N = 2$, then $T_2 > T_1$ (T_2 is indeed optimal), for $N = 3(T_3 > T_1, T_2)$ etc. We now produce an example for independent populations which shows that for $N = 3$, $T_1 = S_0 > T_2$. This negates and destroys the

above intended direction for improving strategies. To this end, suppose μ_1, μ_2, μ_3 and μ_1', μ_2', μ_3' represent the first three moments of two distributions $F(\rho)$ and $G(\sigma)$ respectively, and that $\mu_1 > \mu_1'$, $\mu_2 < \mu_2'$, $\mu_3 > \mu_3'$, $\mu_1 + \mu_1\mu_1' < \mu_1' + \mu_2'$, $\mu_1' > \dfrac{\mu_2 - \mu_3}{\mu_1 - \mu_2}$ and $\mu_1 > \dfrac{\mu_2' - \mu_3'}{\mu_1' - \mu_2'}$. In view of these inequalities, according to T_2 we readily obtain for $N = 3$ the expected yield $\mu_1' + \mu_2' + \mu_3' +$

$$(\mu_1' - \mu_2')\, \mu_1 + (1 - \mu_1') \left[u_1 + \mu_2 + (1 - \mu_1')\, \max\left(\frac{\mu_1' - \mu_2'}{1 - \mu_1'}, \frac{\mu_1' - \mu_2'}{1 - \mu_1'} \right) \right] .$$

In a similar manner we can get the expected yield following strategy $T_1 = S_0$. The difference becomes $T_1 - T_2 = \mu_3 - \mu_3' > 0$. To complete the example, it remains only to construct two distributions with moments satisfying the above inequalities. Let a_1, a_2, a_3 and b_1, b_2, b_3 denote any successive moments of two distributions H and K where $a_1 > b_1$, $a_2 < b_2$ and $a_3 > b_3$; e.g., $a_1 = \frac{1}{2}$, $a_2 = \frac{1}{3}$, $a_3 = \frac{1}{4}$, $b_1 = \frac{1}{2} - \eta$, $b_2 = \frac{1}{3} + \eta$, $b_3 = \frac{1}{4} - \eta$, which for η sufficiently small are the moments of a distribution since (a_1, a_2, a_3) is an interior point of the moment space of order 3. Let c_1, c_2, c_3 denote the moments of a distribution L satisfying $c_1 > \dfrac{c_2 - c_3}{c_1 - c_2}$. The Schwartz inequality implies that $c_1 > \dfrac{c_1 - c_2}{1 - c_1}$. Let $F(\rho) = \epsilon H(\rho) + (1 - \epsilon) L(\rho)$ and $G(\sigma) = \epsilon K(\sigma) + (1 - \epsilon) L(\sigma)$ with ϵ chosen sufficiently small. Then

$$\mu_1 = \epsilon a_1 + (1 - \epsilon) c_1 \quad \text{and} \quad \mu_1' = \epsilon b_1 + (1 - \epsilon) c_1 \ .$$

We immediately get that $\mu_1 > \mu_1'$, $\mu_2 < \mu_2'$, and $\mu_3 > \mu_3'$ from the corresponding properties of the a_i and b_i. Also $\mu_1 > \dfrac{\mu_1' - \mu_2'}{1 - \mu_1'}$, $\mu_1 > \dfrac{\mu_2' - \mu_3'}{\mu_1' - \mu_2'}$, $\mu_1' > \dfrac{\mu_2 - \mu_3}{\mu_1 - \mu_2}$, and $\mu_1' > \dfrac{\mu_1 - \mu_2}{1 - \mu_1}$ follow since $\mu_1 \sim \mu_1' \sim c_1$ for ϵ sufficiently small for which these inequalities hold.

The next principle we examine is that of "staying on a winner." This principle involves: Does the optimal strategy have the property that whenever success occurs on a given play of a machine this same machine is tried at the next trial? This is not always optimal for the case of dependent distributions $F(\rho, \sigma)$. Consider the following example: $F(\rho, \sigma)$ concentrates at two points, $(\epsilon, 0)$, with probability λ and $(1-\epsilon, 1)$ with probability $1 - \lambda$. With $\epsilon = .1$ and $\lambda = .8$, then consider

$$(4) \qquad \epsilon^2 \lambda + (1-\epsilon)^2 (1-\lambda) < (1-\epsilon)(1-\lambda)$$

and

$$(5) \qquad \epsilon(1-\epsilon)\lambda + (1-\epsilon)\epsilon(1-\lambda) > \epsilon(1-\lambda) .$$

The inequalities yield the following: If machine I is used and success results, then (4) implies that II is to be played next, while if failure results, then (5) requires that I is again to be used. The interpretation becomes that if success results, then it is highly likely that the sample consists of machines

of high probability of success and hence II is preferred. A
similar meaning is attributed to the situation of failure.
On the other hand, if one chooses II first, then perfect infor-
mation results, and on the basis of the outcome the play is
evident for the last step.

Computing the expected value starting with I yields

$$(1-\epsilon)(1-\lambda) \left[2 + \epsilon\right] + \epsilon\lambda(2-\epsilon) = .53.$$

The expected yield beginning with II gives $2(1-\lambda) + \lambda\epsilon = .48$.
Consequently, in general, the "staying on a winner" principle
does not apply. However, it is conjectured that when the machines
come from independent populations, then this principle is valid
for the optimal strategy.

A related concept is the property of "monotonicity," defined
as follows. Let the number of trials be fixed and let the a
priori distributions be $F(\rho)$ and $G(\sigma)$. Suppose that it is
optimal to play I first. Then if F is replaced by $F^S = \dfrac{\rho dF}{\int \rho dF}$
with G unchanged, suppose it is still true that I is preferred
to II at the first step. In this case, we say that the optimal
strategy is monotone. It is trivial to show that using the same
machine is equivalent to keeping F unchanged but decreasing G
to G^f.

We assume in what follows that the machines are from inde-
pendent universes.

Lemma 2. If the optimal strategy for any number of trials is
monotone, then the principle of "staying on a winner" is valid.

Proof: Suppose that it is definitely better to play I first, but if success results, shift to II on the next trial. By the monotonocity assumption a fortiori, if failure results on the first trial, II still is played next. But playing I, then II, and optimizing from then on, is equivalent to playing II, then I, and then optimizing from then on since the order of the first two plays does not affect what follows. This contradicts the assumption that it is definitely better to play I on the first trial, and the lemma is established.

We note that to prove the proposition of "staying on a winner" for N trials it is sufficient to know the monotonicity criteria for $N-1$ trials. Using lemma 1 and lemma 2 we now verify the "staying on a winner" principle for 2 and 3 trials. It is trivial for $N = 2$. For the case $N = 3$, it is sufficient to show monotonicity for $N = 2$. We need to consider two cases where I is preferred.

Case 1.

$\mu_1 \geq \mu_1'$ and either $\mu_2 \geq \mu_2'$ or $\mu_1 + \mu_1 \mu_1' \geq \mu_1' + \mu_2'$. If μ_1 is replaced by $\dfrac{\mu_2}{\mu_1}$, then clearly any of the inequalities valid before continue to hold.

Case 2.

$\mu_1' \geq \mu_1$ but $\mu_2 > \mu_2'$ and $\mu_1 + \mu_2 > \mu_1' + \mu_1' \mu_1$. We first observe that the last inequality implies that since $\mu_1' \geq \mu_1$, $\mu_2 > \mu_1' \mu_1$ or $\dfrac{\mu_2}{\mu_1} > \mu_1'$. This combined with $\dfrac{\mu_3}{\mu_1} \geq \mu_2 > \mu_2'$ insures by lemma 1 that machine I is chosen at the first trial.

The general monotonicity property for independent machines remains an open question.

The last general property investigated is whether the a priori expected value is monotone increasing as a function of the steps when employing an optimal policy. While this is true if the strategy is S_o, it is not true in general.

First consider the following.

Lemma 3. The strategy S_o applied to any initial distribution F has the property that the a priori expected contribution at each stage is non-decreasing.

Proof: It is enough to prove the result for the first two stages. Suppose according to S_o machine I is used first; then the expected value is $\int \rho dF(\rho,\sigma)$. Thus, $\int \rho dF(\rho,\sigma) > \int \sigma dF(\rho,\sigma)$. If independent of the outcome, machine I is employed at stage 2, then the outcome is

$$\int \rho dF(\rho,\sigma) \cdot \frac{\int \rho^2 dF(\rho,\sigma)}{\int \rho dF(\rho,\sigma)} + \int (1-\rho)dF(\rho,\sigma)\left[\frac{\int \rho(1-\rho)dF(\rho,\sigma)}{\int (1-\rho)dF(\rho,\sigma)}\right] = \int \rho dF(\rho,\sigma).$$

Consequently, if the machine with maximum expected value is used, the total expected value is $\geq \int \rho dF(\rho,\sigma)$.

In contrast to this result, consider the case of N = 3 with $dF(\rho) = \frac{\rho^5(1-\rho)^3 d\rho}{\int \rho^5(1-\rho)^3 d\rho}$ and $dG(\sigma) = d\sigma$. It turns out that for optimal return, machine I is preferred first with expected value for the first move equal to .6. If success results, then I is played again, while if failure occurs then the criteria of lemma 1

require II to be chosen. The a priori expected value for the second step gives $\frac{64}{110} < .6 = \frac{66}{110}$.

Another way to express the fact is that if we let the random variable $x_r(P)$ represent the yield at the rth stage according to policy P, then one would suspect the sequence of random variables obtained by the optimal policy would form a semi-martingale. The example presented above negates this proposition. Let $x_r(P)$, as before, denote the yield at the rth stage according to the policy P. We note that always

$$\text{Exp } (\frac{x_1 + \ldots + x_n}{n}) \leq \int \int \max (\rho, \sigma) dF(\rho, \sigma) \ .$$

It can be shown using the law of large numbers that if S_o is modified so that at infinitely many trials prescribed in advance of density zero both I and II are used and otherwise the usual criteria of S_o are employed, then

$$\lim \text{Exp } (\frac{x_1 + \ldots + x_n}{n}) = \int \int \max (\rho, \sigma) dF(\rho, \sigma) \ .$$

This is a type of consistency result. Unfortunately, most procedures are consistent in the above sense and thus this concept does not help one choose among strategies.

§3. __The Case of One Known and One Unknown Probability of Success__

In this section we examine in detail the situation where $F(\rho, \sigma) = F(\rho)G(\sigma)$ with $G(\sigma) = I_\sigma$. In other words, the distributions are independent with the probability of success of machine II known to be σ. Let n trials be allowed and F be the initial

a priori distribution of success for Machine I. Define $K_n(F)$ by the condition that if $\sigma > K_n(F)$, the optimal procedure is to use the known machine for the first step, and if $\sigma < K_n(F)$, then machine I is the optimal choice while if $\sigma = K_n(F)$ either choice is optimal. We adopt the convention that if at any trial it is optimal to use either machine, then in that case we choose I. We seek to determine the form of $K_n(F)$ which represents the decision function.

The optimal procedure then is given as follows:

If $\sigma \leq K_n(F)$, then at the first step one uses the machine of unknown probability of success. On the other hand, if $\sigma > K_n(F)$, one uses the known machine. After the first step, depending on what happened, we compute the new a posteriori distributions I_σ and $F'(\rho)$ and compare σ and $K_{n-1}(F')$ following the above rules as to what to do at the second stage, etc.

We now establish a series of lemmas describing the form of the optimal strategy.

Lemma 4. If the known machine II is employed at any trial according to an optimal strategy, then it is used thereafter.

Proof: If the optimal procedure uses II r times ($r < n$) and then I, the expected value is

$$(6) \qquad r\sigma + E(F) + E(F)Y(F^s, n-r-1) + \left[1 - E(F)\right] Y(F^f, n-r-1)$$

where $E(F)$ is the expected value of the distribution F; F^s is an a posteriori distribution given success has occurred on I; F^f

corresponds to the case where failure happened on I and $Y(F, n)$ is the optimal expected yield when the a priori distribution is F and n trials remain. The strategy using I first followed by r tries on II and then optimal continuation gives the same yield as in (6). Thus by our convention the optimal procedure calls for use of I first.

Lemma 5. For any distribution F and $n \geq 2$,

$$K_{n-1}(F) \leq K_n(F) .$$

Proof: In fact, suppose the contrary and that σ is such that $K_{n-1}(F) > \sigma > K_n(F)$. Consequently, the optimal procedure begins with machine II and then must play I at the second trial. This contradicts lemma 4.

Lemma 6. For any distribution F and any n

$$Y(F^s, n) \geq Y(F^f, n)$$

where $Y(F, n)$ represents the expected yield following an optimal policy for n moves when F is the given a priori distribution of ρ.

Proof: The proof is by induction on n. If $K_n(F^s) > \sigma > K_n(F^f)$ then $Y(F^s, n) \geq n\sigma \geq Y(F^f, n)$ from which we conclude that the lemma is valid. If $\sigma \geq K_n(F^s)$ and $\sigma \geq K_n(F^f)$, then $Y(F^s, n) = Y(F^f, n) = n\sigma$, and again lemma 6 is true. Thus suppose both $K_n(F^s), K_n(F^f) > \sigma$, then

$$Y(F^s, n) = E(F^s) + E(F^s)Y(F^{ss}, n-1)$$

$$+ \left[1 - E(F^s)\right]Y(F^{sf}, n-1) = E(F^s) + A_n$$

while

$$Y(F^f, n) = E(F^f) + E(F^f)Y(F^{fs}, n-1)$$

$$+ \left[1 - E(F^f)\right]Y(F^{ff}, n-1) = E(F^f) + B_n .$$

The induction hypothesis shows that

$$Y(F^{ss}, n-1) \geq Y(F^{sf}, n-1) = Y(F^{fs}, n-1) \geq Y(F^{ff}, n-1)$$

and thus any convex combination of the first two terms is larger than or equal to any convex combination of the last two terms. This yields $A_n \geq B_n$, but evidently $E(F^s) \geq E(F) \geq E(F^f)$ and hence

$$Y(F^s, n) \geq Y(F^f, n)$$

Lemma 7. For any distribution F, we have

$$K_n(F^s) \geq K_n(F^f) .$$

Proof: (By contradiction) Suppose σ is such that

$$K_n(F^s) < \sigma < K_n(F^f) .$$

We secure

$$n\sigma < Y(F^f, n) \le Y(F^s, n)$$

by lemma 6. This contradicts the fact that $n\sigma$ is the optimal yield when the a priori distribution is F^s.

Lemma 8. If success occurs on either machine while following an optimal procedure, then the same machine is employed at the next trial.

Proof: It has been shown by lemma 4 that if the unknown machine is ever used, then one never departs from it according to an optimal procedure. To complete the proof, it remains to show that if success occurs on I, then one chooses this same machine the next time. It is clearly sufficient to show this for the first two trials. Suppose the lemma is false, that I is used, a success occurs and one switches to II. Thus $\sigma > K_{n-1}(F^s) \ge E(F^s) \ge E(F)$ by lemma 5. By lemma 7, also $K_{n-1}(F^f) < \sigma$. Consequently,

$$n\sigma < Y(F, n) = E(F) + (n-1)\sigma$$

and thus $E(F) > \sigma$ which contradicts the above inequality.

Another property valid for this model is contained in lemma 9.

Lemma 9. The a priori expected value for each stage is non-decreasing when pursuing an optimal strategy.

<u>Proof</u>: It is sufficient to show this for the first two steps. When II is used at the first step, the result is trivial in view of lemma 4. On the other hand, if I is used, then the expected value for the first step is $E(F)$. If one continues with I regardless of the outcome of the first trial, then the a priori expected value is again $E(F)$ for the second stage, which substantiates the conclusion of the lemma.

It remains only to consider the case where the second trial depends on the result of the first trial. On account of lemma 8, if success occurred first, then I is again chosen. Suppose a failure occurs and the optimal strategy calls for a switch, then $\sigma > K_{n-1}(F^f) \geq E(F^f)$. Consequently, the expected value at the second stage is

$$\mu_1 \left[\frac{\mu_2}{\mu_1} \right] + (1-\mu_1)\sigma \geq \mu_2 + (1-\mu_1)\frac{(\mu_1-\mu_2)}{1-\mu_1} = \mu_1$$

where μ_1 are the moments about zero of F.

As we have seen in §2, lemma 9 is not always true.

The above lemmas enable us to describe completely the optimal strategy. To determine the explicit value of $K_n(F)$, we assume that $\sigma = K_n(F)$. It is clear in view of lemma 8 that the optimal strategy has the following form for appropriate k_1 (defined below).

(A) At the first step choose I and stay with it until a failure occurs.

(B) There exists an integer $k_1 \geq 0$ such that if at least k_1 successes have occurred before the one failure, then proceed

with I. Otherwise, if less than k_1 successes occur before the failure, then change to II from there on.

(C) A corresponding integer k_2 is attached to two failures, i.e., if two failures have resulted and less than $k_1 + k_2$ successes, then switch to II; otherwise continue with I.

(D) Generally, if r failures and s successes have occurred where $k_1 + \cdots + k_{r-1} < s < k_1 + k_2 + \cdots k_r$, then change to II; otherwise continue with I.

The yield due to the strategy prescribed above can be collected in the following way: All the terms with no failure have the form

$$I_0 = \int (\rho + \rho^2 + \cdots \rho^n) dF .$$

All the terms with one failure for machine I combine to yield according to the choice of k_1 the value

$$I_1 + I_1(\sigma) = \int \left[\binom{1}{1}\rho + \binom{2}{1}\rho^2 + \cdots + \binom{n-k_1-1}{1}\rho^{n-k_1-1} \right] \rho^{k_1}(1-\rho)dF(\rho)$$

$$+ \sigma \left[\int \left\{ (n-1) + (n-2)\rho + \cdots + (n-k_1)\rho^{k_1-1} \right\}(1-\rho)dF(\rho) \right].$$

Analogously, the contributions corresponding to two failures for machine I give the quantity

$$I_2 + I_2(\sigma) = \int k_2 \left[\binom{1}{1}\rho + \binom{2}{1}\rho^2 + \cdots \binom{n-k_1-k_2-2}{1}\rho^{n-k_1-k_2-2} \right] \rho^{k_1+k_2}(1-\rho)^2 dF(\rho)$$

$$+ \int \left[\binom{2}{2}\rho + \binom{3}{2}\rho^2 + \cdots \binom{n-k_1-k_2-2}{2}\rho^{n-k_1-k_2-2} \right] \rho^{k_1+k_2}(1-\rho)^2 dF(\rho)$$

$$+ \sigma \int \left[(n-k_1-2)\binom{1}{1} + (n-k_1-3)\binom{2}{1}\rho + \cdots (n-k_1-k_2-1)\binom{k_2}{1}\rho^{k_2-1} \right] \rho^{k_1}(1-\rho)^2 dF(\rho) .$$

The terms involving exactly r failures for machine I yield

$$(7) \quad I_r + I_r(\sigma) = \int \sum_{a_1=0}^{1} \binom{a_1-1}{a_1-1} + \cdots + \binom{k_2+a_1-1}{a_1-1} \sum_{a_2=0}^{2-a_1} \left\{ \binom{a_2-1}{a_2-1} + \cdots + \binom{k_3+a_2-1}{a_2-1} \right\} \cdots$$

$$\sum_{a_r=0}^{r-a_1-a_2\cdots a_{r-1}} \left\{ \binom{a_r-1}{a_r-1} + \cdots \binom{k_{r+1}+a_r-1}{a_r-1} \right\} \left[\binom{b_r}{b_r}\rho + \binom{b_r+1}{b_r}\rho^r + \cdots \binom{b_r+n-\sum_{i=1}^{r}k_i-r-1}{b_r}\rho^{n-\sum_{i=1}^{r}k_i-r} \right].$$

$$\rho^{k_1+k_2+\cdots k_r}(1-\rho)^r dF(\rho)$$

$$+ \sigma \int \sum_{a_1=0}^{1} \left\{ \binom{a_1-1}{a_1-1} + \cdots \binom{k_2+a_1-1}{a_1-1} \right\} \sum \cdots \sum_{a_{r-1}=0}^{r-1-\sum_{i=1}^{r-2}a_i} \left\{ \binom{a_{r-1}-1}{a_{r-1}-1} + \cdots + \binom{k_r+a_{r-1}-1}{a_{r-1}-1} \right\} .$$

$$\left[(n-\sum_{i=1}^{r-1}k_i-r)\binom{c_r}{c_r} + (n-\sum_{i=1}^{r-1}k_i-r-1)\binom{c_r+1}{c_r}\rho + \cdots (n-\sum_{i=1}^{r}k_i-r)\binom{c_r+k_r}{c_r}\rho^{k_r-1} \right] .$$

$$\rho^{\sum_{i=1}^{r-1}k_i}(1-\rho)^r$$

where $b_r = r - a_1 - a_2 - \cdots - a_r$, $c_r = r - 1 - a_1 \cdots - a_{r-2}$, and we interpret $\binom{c}{-1} = 0$ for $c \neq -1$ with $\binom{-1}{-1} = 1$. Our objective is fulfilled in the following theorem.

Theorem 2.

$$K_n(F) = \sup_{k_1, k_2, \cdots} \frac{I_0 + I_1 + I_2 + \cdots + I_r + \cdots}{I_0' + I_1' + I_2' + \cdots + I_r' + \cdots} = \sup_{k_1, k_2, \cdots} \frac{J}{J'}$$

where $J = I_0 + I_1 + I_2 + \cdots$ and where I_r' is obtained from I_r by replacing $dF(\rho)$ by $\frac{dF(\rho)}{\rho}$ and the k_1 are subject to the restrictions $0 < k_1 < n - 1$, $0 < k_2 < n - 2 - k_1$, $\cdots$, $0 < k_r < n - r - \sum_{i=1}^{r-1} k_1$, $\cdots$, with the understanding, for instance, that if $n - \ell - \sum_{i=1}^{\ell-1} k_1 = 0$, then $\frac{J}{J'} = \frac{I_0 + I_1 + \cdots I_{\ell-1}}{I_0' + I_1' + \cdots I_{\ell-1}'}$.

The proof consists in showing that

$$n\sigma - I_1(\sigma) - I_2(\sigma) - I_3(\sigma) \cdots = \sigma \left[I_0' + I_1' + I_2' + \cdots \right] .$$

The general formula is established by a long induction argument and we shall illustrate the method of proof by considering only the first few sums. The general proof can be established by an extension of the argument used. The basic identity used extensively in the proof is

(8) $n - (n-1) \int (1-\rho)dF(\rho) - (n-2) \int \rho(1-\rho)dF - \cdots - \int \rho^{n-2}(1-\rho)dF(\rho)$

$$= \int (1+\rho+\rho^2+\cdots+\rho^{n-1})dF \ .$$

This can be verified directly by a simple induction. As an immediate consequence of (8), we obtain

(9) $n - (n-1) \int (1-\rho)dF - (n-2) \int \rho(1-\rho)dF - \cdots - (n-k_1) \int \rho^{k_1-1}(1-\rho)dF$

$$= (n-k_1) \int \rho^{k_1}dF + \int (1+\rho+\cdots+\rho^{k_1-1})dF \ .$$

Using (9), we secure

(10) $n\sigma - I_1(\sigma) = \sigma \left\{ \int (1+\rho+\cdots+\rho^{k_1-1})dF + (n-k_1) \int \rho^{k_1}dF \right\}.$

Repeated application of (9) gives

(11) $\sigma(n-k_1) \int \rho^{k_1}dF - I_2(\sigma) = \sigma \left\{ k_2(n-k_1-k_2-1) \int \rho^{k_1+k_2}(1-\rho)dF(\rho) \right.$

$$+ (n-k_1-k_2) \int \rho^{k_1+k_2}dF(\rho) + \int (\rho^{k_1}+\rho^{k_1+1}+\cdots+\rho^{k_1+k_2-1})dF$$

$$\left. + \int \rho^{k_1}(1-\rho) \left[\binom{1}{1} + \binom{2}{1}\rho + \cdots + \binom{k_2}{1}\rho^{k_2-1} \right]dF(\rho) \right\} \ .$$

To describe one more step in the process we find again by using (9) several times that

$$(12) \quad \sigma \left\{ k_2(n-k_1-k_2-1) \int \rho^{k_1+k_2}(1-\rho)dF + (n-k_1-k_2) \int \rho^{k_1+k_2}dF \right\} - I_3(\sigma)$$

$$= \sigma \left\{ (n-k_1-k_2-k_3) \int \rho^{k_1+k_2+k_3}dF + (k_2+k_3)(n-k_1-k_2-k_3-1) \int \rho^{k_1+k_2+k_3}(1-\rho)dF \right.$$

$$+ \left[k_2 k_3 + \binom{1}{1} + \binom{2}{1} + \cdots \binom{k_3}{1} \right] (n-k_1-k_2-k_3-2) \int \rho^{k_1+k_2+k_3}(1-\rho)^2 dF$$

$$+ \int (\rho^{k_1+k_2} + \cdots + \rho^{k_1+k_2+k_3-1})dF + \int (1-\rho)\rho^{k_1+k_2} \left[\binom{k_2+1}{1} + \binom{k_2+2}{1}\rho \right.$$

$$+ \cdots + \binom{k_2+k_3}{1}\rho^{k_3-1} \right]dF + \int k_2 \left[\binom{1}{1} + \binom{2}{1}\rho + \cdots + \binom{k_3}{1}\rho^{k_3-1} \right]\rho^{k_1+k_2}(1-\rho)^2 dF$$

$$+ \int \left[\binom{2}{2} + \binom{3}{2}\rho + \cdots + \binom{k_3+1}{2}\rho^{k_3-1} \right]\rho^{k_1+k_2}(1-\rho)^2 dF \quad .$$

The pattern is now clear that on combining (10), (11), (12) and continuing in the same manner we find that $\sigma J' = J$ or $\sigma = \dfrac{J}{J'}$. Hence $K_n(F) = \sup\limits_{k_1} \dfrac{J}{J'}$.

Some special cases are worth noting:

$$K_2(F) = \frac{\int (\rho + \rho^2)dF(\rho)}{\int (1+\rho)dF(\rho)} = \frac{I_0}{I_0'}$$

$$K_3(F) = \max \left(\frac{I_0}{I_0'} , \frac{I_0+I_1}{I_0'+I_1'} \right)$$

$$= \max \left(\frac{\int (\rho + \rho^2 + \rho^3)dF}{\int (1+\rho+\rho^2)dF} , \frac{\int (\rho + 2\rho^2)dF(\rho)}{\int (1+2\rho)dF(\rho)} \right) .$$

Unfortunately, both terms in $K_3(F)$ can occur; e.g., if $F(\rho) = \rho$,

then $K_3(\rho) = \dfrac{I_0}{I_0'} = \dfrac{\frac{1}{2} + \frac{1}{3} + \frac{1}{4}}{1 + \frac{1}{2} + \frac{1}{3}} = \dfrac{13}{22}$ while if $F(\rho) = \rho^{1/5}$, then

$$K_3(\rho) = \frac{\int (\rho + 2\rho^2)\rho^{-4/5}d\rho}{\int (1 + 2\rho)\rho^{-4/5}d\rho} = \frac{23}{88} \ .$$

In general, the expression for $K_n(F)$ in Theorem 2 can not be simplified in any way and represents the simplest form for the decision function available which again testifies to the complex nature of optimal strategies in such sequential design problems.

For practical purposes a reasonable approximation to $K_n(F)$ can be obtained by choosing $k_1 = n - 1$. In that case, one compares σ with

$$L_n(F) = \frac{\int (\rho + \rho^2 + \cdots + \rho^n)dF}{\int (1 + \rho + \cdots + \rho^{n-1})dF} \ .$$

This applies well for n small ($n \leq 10$). In the case of $F(\rho) = \rho$, for example, then $K_n(\rho) = L_n(\rho)$ when $n \leq 4$, but they cease to be equal for $n = 5$ and on. It is worth noting that $L_n(F)$ shares many of the properties of $K_n(F)$.

Lemma 10. $L_n(F)$ is monotonic increasing in n for any F and $L_n(F^s) > L_n(F) \geq L_n(F^f)$.

Proof: The proof of lemma 10 is based on the well-known result that if $\dfrac{a_r}{b_r}$ is increasing with $a_r, b_r > 0$, then $\dfrac{a_1 + \cdots + a_n}{b_1 + \cdots + b_n}$ is also

increasing. Since $\dfrac{\int \rho^r dF}{\int \rho^{r-1} dF} \le \dfrac{\int \rho^{r+1} dF}{\int \rho^r dF}$ by virtue of Holder's inequality we obtain that $L_n(F)$ is monotone increasing as a consequence of the above cited result. The latter part of the lemma can be proved readily using this same result.

<u>Theorem 3</u>. If F has the property that $\int \dfrac{1}{1-\rho} dF(\rho) = \infty$, <u>then</u> $K_n(F) \to 1$. Moreover, the <u>known</u> machine II <u>is never used</u> if

$$\sigma \le \frac{\int \rho(1-\rho)^{n-1} dF(\rho)}{\int (1-\rho)^{n-1} dF(\rho)} \ .$$

<u>Proof</u>:

$$1 \ge K_n(F) \ge \frac{\int(\rho + \cdots + \rho^n)dF}{\int(1+\rho + \cdots + \rho^{n-1})dF} = 1 - \frac{\int(1-\rho^n)dF}{\int \frac{1-\rho^n}{1-\rho} dF} \ge 1 - \epsilon$$

for n sufficiently large. The right-hand side of the second part of Theorem 3 is the expected value at trial n if $n - 1$ tries on machine I produced all failures. The last assertion follows since $K_n(F) \ge E(F)$ for any F and any n.

The interpretation of Theorem 3 is the intuitive fact that if there exists substantial positive probability of success and if the number of trials is sufficiently large, the unknown machine should be played first unless $\sigma = 1$.

§4. Certain Game Aspects of the General Problem.

The first type of game problem we consider in this section is as follows: Let $\Phi_N(F(\rho,\sigma),S)$ denote the expected value obtained when

$F(\rho, \sigma)$ is the a priori distribution for the probabilities of successes ρ and σ of machines I and II respectively, and S defines a strategy. The number N denotes the fixed number of trials to be used throughout these game considerations. We therefore drop the subscript N.

The function $\Phi(F, S)$ is evaluated as follows: The a priori distribution is given first and the policy S is a procedure in terms of inequalities involving the complete first n moments of the distribution and in $F(\rho, \sigma)$. (See the beginning of §2.)

Theorem 4. If $\Phi(F, S)$ <u>is</u> <u>evaluated</u> <u>as</u> <u>indicated</u> <u>above</u>, <u>then</u>

$$\min_{F} \max_{S} \Phi(F, S) = \max_{S} \min_{F} \Phi(F, S) = \max(N\alpha, N\beta)$$ <u>where</u> <u>the</u> <u>class</u>

<u>of</u> <u>distributions</u> <u>is</u> <u>restricted</u> <u>by</u> <u>the</u> <u>condition</u> $\int \rho \, dF(\rho, \sigma) = \alpha$

<u>and</u> $\int \sigma dF(\rho, \sigma) = \beta$. <u>An</u> <u>optimal</u> <u>minimax</u> <u>distribution</u> <u>is</u> $F = I_{\alpha, \beta}$
(<u>the</u> <u>distribution</u> <u>concentrating</u> <u>fully</u> <u>at</u> (α, β)) <u>while</u> S_o <u>is</u> <u>an</u>
<u>optimal</u> <u>minimax</u> <u>policy</u>.

Proof: If one considers the distribution $I_{\alpha, \beta}$, then regardless of the strategy S employed, an upper bound for the yield is max $(N\alpha, N\beta)$. This is evident since after every performance, the a posteriori distribution is unchanged and equal to $I_{\alpha, \beta}$. It represents the only distribution where the information is complete and no experimentation contributes any value. On the other hand, if the statistician employs policy S_o, then by virtue of the conditions on the moments of F the yield at the first step is $\max(\alpha, \beta)$. Lemma 3 implies that $\Phi(F, S_o) \geq n \max(\alpha, \beta)$ for any F of the type examined. The proof of Theorem 4 is hereby complete.

Thus, the intuitive strategy S_o does assume a certain general significance on account of Theorem 4. We remark that there exist many other optimal minimax policies aside from S_o.

Another type of game can be introduced where decisions S are not functions of an a priori distribution but depend only on the observed number of successes and failures to that point. The expected value $\Phi(F, S)$ is evaluated in terms of F and the game is considered where F is restricted by $\int \rho dF = \alpha$ and $\int \sigma dF = \beta$. The analysis of this game remains an open question.

<div align="right">
The RAND Corporation and
The California Institute of Technology
</div>

BIBLIOGRAPHY

1. Herbert Robbins, <u>Bulletin of American Mathematical Society</u>, Vol. 58, No. 5, September 1952, Some aspects of the sequential design of experiments.

2. W. R. Thompson, <u>Biometrika</u>, Vol. 25, 1933, pp. 285–294, On the likelihood that one unknown probability exceeds another in view of the evidence of two samples.

3. W. R. Thompson, <u>American Journal of Mathematics</u>, Vol. 57, April 1935, On the theory of apportionment.