ON THE DESIGN AND COMPARISON OF CERTAIN DICHOTOMOUS EXPERIMENTS

By Russell N. Bradt and Samuel Karlin

University of Kansas and California Institute of Technology

1. Introduction and Summary. It may frequently happen that a researcher, wishing to decide which one of a set of alternatives to accept, finds that there are several experiments available to him which he might perform to guide him in reaching his decision. Thus he is faced with making a preliminary decision as to which experiment or experiments he is to perform. If he admits the possibility of performing more than one, then the question of how many, which ones, and in what order arises. It is such questions as these that come under the heading of comparison and design of experiments.

While a great deal of general theory of the design problem has been developed, e.g., by Wald [1] and Maguire [2], few actual solutions of particular problems, especially of the sequential type, have been investigated thus far. Robbin's paper [3] is the first published report dealing with various sequential rules for particular nontruncated design problems. The basic purpose of this paper is to investigate for certain cases the optimal design, which almost uniformly turns out to be exceedingly complicated (see Section 3), and to propose and determine some justification for certain simpler criteria.

Attention is restricted to problems in which there are but two alternatives, or hypotheses, H_1 and H_2, and it is required to decide between them with a loss of one unit if the false one is selected, while no loss results from selecting the correct one. Further, ξ will denote the a priori probability that H_1 is true, and the basic criterion for comparison will be the Bayes risks associated with the various experiments. To say that an experiment is available to a researcher is to say that there is a real random variable which he can observe whose distribution is known under each hypothesis.

As an example of a situation in which this type of question may arise, consider the problem of deciding between utilizing a use test as against a specifications test for acceptance of a lot of manufactured items. A large lot of items has been produced and a decision is to be made between, say, w_1 and w_2 as being the proportion of defectives in the lot. Let $X = 1$ or 0 according as an item selected at random is defective or not as determined by subjecting it to a use test. Let $Y = 1$ or 0 according as an item selected at random is classified as defective (because it fails to meet certain specifications) or not. If α, the probability that a nondefective item fails to meet the specifications, and β, the probability that a defective item meets the specifications, are known, then both X and Y have a binomial distribution with known parameter under each hypothesis.

Again, it might be that in the course of a series of treatments of a material there are two points at which a certain characteristic may be measured, say

Received October 27, 1954.

the breaking strength of a metal undergoing a series of heat treatments. Let X and Y, respectively, denote the value of the characteristic at the different points in the process. It is reasonable to assume that under each of two simple hypotheses concerning the process and material, X and Y have prescribed normal distributions.

In general, suppose that X and Y are two real random variables having distribution functions of F_i and G_i, respectively, under hypothesis H_i and with corresponding densities f_i and g_i with respect to a common measure, ψ, such that $f_i > 0$ if and only if $g_i > 0$. Let $R_Z(\xi)$ denote the Bayes risk against ξ when using experiment Z.

Now the computation and comparison of the risk, $R_X(\xi)$ and $R_Y(\xi)$, for all ξ, which appears to be necessary in order to obtain an optimal design, is intrinsically complicated in most cases, as will be seen in the following sections. Hence it is of some interest to investigate some more convenient criteria for choosing between experiments. Any such criterion should, of course, dictate the use of X if $R_X \leqq R_Y$ (i.e., $R_X(\xi) \leqq R_Y(\xi)$ for all ξ). To be able to check on this is but one reason for interest in conditions that $R_X \leqq R_Y$. Another is that whenever it is true that $R_X \leqq R_Y$, then regardless of the choice of actions open to the researcher or the loss function used, use of X will never yield a greater Bayes risk than Y. Also, if a total of, say, n independent experiments is to be performed, the optimal sequential design is the nonsequential rule: take all n observations of X ([4], [5], [8]).

In Section 2.1 general conditions that $R_X \leqq R_Y$ are derived, some of which are related to those obtained by Blackwell [5] via consideration of the standard experiment. In Section 2.2 the Kullbach-Leibler (abbreviated hereafter as K-L) information numbers are introduced, and it is shown, in particular, that they provide a criterion which yields an especially simple necessary condition that $R_X \leqq R_Y$. The K-L information numbers are also considered as functions of that transformation, t, such that the distribution of $t(X)$ under H_1 is the distribution of X under H_2. The case in which all the distributions involved are normal is analyzed in some detail in Section 2.3, where it is seen that the K-L numbers do not yield a sufficient condition that $R_X \leqq R_Y$, though they do yield a sufficient condition that $R_X = R_Y$. The normal case gives an example in which a second criterion, that of being "locally more informative" (Bayes) at both zero and 1, yields a condition both necessary and sufficient that $R_X \leqq R_Y$. X is termed locally more informative than Y at ξ if ξ lies in an interval $[\xi_1, \xi_2]$ such that on $[0, 1] \cap [\xi_1, \xi_2]$ $R_X(\xi) \leqq R_Y(\xi)$, with strict inequality at at least one end point. This latter criterion is discussed further in Section 2.4.

The problem of determining the optimal designs, sequential and nonsequential, for the case in which all distributions are binomial and a fixed number of experiments is to be performed is discussed in Section 3. The complete solutions are found to be exceedingly complicated; a few are given. For the sequential design, a system for obtaining the optimal design which avoids the complete calculation of the successive risk functions was found.

In the final section, a sequential rule for terminating experimentation is con-

sidered, and the problem of finding a sequential design which minimizes the expected number of experiments is posed. Two reasonable designs are proposed, are shown to be equivalent, and are shown to be better than either of the rules which require that all observations be of the same random variable.

Throughout the paper it will be convenient to consider $\xi/(1-\xi)$, and this will regularly be denoted by η.

2.1. Uniform inequality of risk functions. With respect to a measure ψ, X and Y have densities under H_i as shown by

		X	Y
(ξ)	H_1	f_1	g_1
$(1-\xi)$	H_2	f_2	g_2,

where $f_1(x) > 0$ if and only if $f_2(x) > 0$, and similarly for the g_i's. It is required to decide which of the hypotheses is true on the basis of one observation, either of X or of Y, with the usual zero-or-one loss.

Interest in the case of one observation stems not merely from general curiosity; but, as mentioned before, if for one observation $R_X \leqq R_Y$, then for any number of observations, any set of actions, and any loss function, consistent use of X yields a Bayes risk less than or equal to that of any other design. In view of this strong property it is clear that the one-observation case holds some interest, and that no criterion for choice of experiment should be seriously considered which would dictate the use of Y when $R_X \leqq R_Y$.

LEMMA 2.1. *Two conditions, each necessary and sufficient, that* $R_X(\xi) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} R_Y(\xi)$ *are*:

(i) $$\int_0^\infty \min(u-\eta, 0)\, dE(u) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \int_0^\infty \min(u-\eta, 0)\, dF(u),$$

(ii) $$\int_0^\infty \min\left(1-\frac{\eta}{u}, 0\right) dG(u) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \int_0^\infty \min\left(1-\frac{\eta}{u}, 0\right) dH(u),$$

where E and G are the c.d.f.'s of $f_2(x)/f_1(x)$ under H_1 and H_2, respectively, and F and H are the c.d.f.'s of $g_2(x)/g_1(x)$ under H_1 and H_2, respectively.

PROOF. From the well-known theory of Bayes solutions (see [3], Chap. 6), the Bayes risk against ξ using X is given by

(1) $$R_X(\xi) = \xi \int_{f_2(x)/f_1(x) > \xi/(1-\xi)} f_1(x)\, d\psi(x) + (1-\xi) \int_{f_2(x)/f_1(x) \leqq \xi/(1-\xi)} f_2(x)\, d\psi(x).$$

With $\eta = \xi/(1 - \xi)$, this can be written as

$$(2) \qquad \frac{R_X(\xi)}{1-\xi} - \eta = \int_{f_2(x)/f_1(x) \leqq \eta} f_2(x)\, d\psi(x) - \eta \int_{f_2(x)/f_1(x) \leqq \eta} f_2(x)\, d\psi(x).$$

With

$$E(u) = \int_{f_2(x)/f_1(x) \leqq u} f_1(x)\, d\psi(x),$$

$$(3) \qquad \frac{R_X(\xi)}{1-\xi} - \eta = \int_0^\eta u\, dE(u) - \eta \int_0^\eta dE(u) = \int_0^\infty \min\,(u - \eta, 0)\, dE(u).$$

With

$$F(u) = \int_{f_2(x)/f_1(x) \leqq u} f_2(x)\, d\psi(x),$$

$$(4) \qquad \frac{R_X(\xi)}{1-\xi} - \eta = \int_0^\eta dF(u) - \eta \int_0^\eta \frac{1}{u}\, dF(u) = \int_0^\infty \min\left(1 - \frac{\eta}{u}, 0\right) df(u).$$

With the analogous expressions for the risk associated with Y, the conclusion is immediate.

LEMMA 2.2. (i) $R_X(\xi) \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} R_Y(\xi)$ *if and only if*

$$\int_0^\eta \alpha_X\left(\frac{u}{1+u}\right) du \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} \int_0^\eta \alpha_Y\left(\frac{u}{1+u}\right) du,$$

where $\alpha_X(\xi)$ is the probability, under H_1, that in following the Bayes procedure against ξ with X, H_2 will be chosen.

(ii) *If $G(u) / u \to 0$ as $u \to 0$, then* $R_X(\xi) \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} R_Y(\xi)$ *if and only if*

$$\int_0^\eta \beta_X\left(\frac{u}{1+u}\right) du \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} \int_0^\eta \beta_Y\left(\frac{u}{1+u}\right) du,$$

where $\beta_X(\xi)$ is the probability under H_2 that in following Bayes procedure against ξ with Y, H_1 will be chosen.

PROOF. From equation (3) of the preceding proof,

$$(1) \qquad \frac{R_X(\xi)}{1-\xi} - \eta = \int_0^\eta (u - \eta)\, dE(u).$$

Integrating by parts and noting that $E(u) = 1 - \alpha_X(u / 1 + u)$,

$$\frac{R(\xi)}{1 - \xi} = \int_0^\eta \alpha_X\left(\frac{u}{1 + u}\right) dE(u). \tag{2}$$

From the similar expression involving $R_Y(\xi)$, conclusion (i) follows.

A parallel argument from equation (4) in the proof of Lemma 2.1 and $G(u) = \beta_X(u / 1 + u)$ yields conclusion (ii).

THEOREM 2.1. *Two conditions, each necessary and sufficient that $R_X = R_Y$, are:*
(i) *f_2/f_1 and g_2/g_1 have the same distributions under H_1 ;*
(ii) *f_2/f_1 and g_2/g_1 have the same distributions under H_2 .*

PROOF. The sufficiency is immediate from Lemma 2.1. To show the necessity, suppose $R_X = R_Y$; then for all $\eta \geqq 0$,

$$\int_0^\infty \min(u - \eta, 0)\, dE(u) = \int_0^\infty \min(u - \eta, 0)\, dF(u). \tag{1}$$

Now, for any $a > 0$, let $\phi_n(u) = n \min(u - a, 0)$ and let $\gamma_n(u) = -n \min(u - (a + 1/n), 0)$, $n = 1, 2, 3, \cdots$. Then

$$\int_0^\infty (\phi_n(u) + \gamma_n(u))\, dE(u) = \int_0^\infty (\phi_n(u) + \gamma_n(u)\, dF(u) \tag{2}$$

for all n. Hence,

$$\begin{aligned} E(a) + \int_{a<u\leqq a+1/n} (1 - n(u - a))\, dE(u) &= F(a) \\ &\quad + \int_{a<u\leqq a+1/n} (1 - n(u - a))\, dF(u). \end{aligned} \tag{3}$$

Letting $n \to \infty$, $E(a) = F(a)$, i.e., the likelihood ratios f_2/f_1 and g_2/g_1 have the same distribution under H_1 . It follows immediately that $\alpha_X(\xi) = \alpha_Y(\xi)$, since $E(u) = 1 - \alpha_X(u / 1 + u)$. But $R_X(\xi) = \xi\alpha_X(\xi) + (1 - \xi)\beta_X(\xi)$; hence, $R_X = R_Y$ and $\alpha_X = \alpha_Y$ implies $\beta_X = \beta_Y$, which is conclusion (ii).

2.2. Relations between Bayes risk and the K-L information numbers. With these conditions that $R_X \leqq R_Y$, attention is turned to the relation between the condition $R_X \leqq R_Y$ and the K-L information numbers.

The mean information per observation of X for discriminating between H_1 and H_2 when H_i is true as defined by Kullbach and Leibler [6], [7] is

$$I_X(1{:}2) = \int_{-\infty}^\infty f_1(x) \log \frac{f_1(x)}{f_2(x)}\, d\psi(x) \qquad i = 1, \tag{2.2.1}$$

and

$$I_X(2{:}1) = \int_{-\infty}^\infty f_2(x) \log \frac{f_2(x)}{f_1(x)}\, d\psi(x) \qquad i = 2.$$

The mean divergence between H_1 and H_2 per observation of X is then defined to be

$$J_X = I_X(1:2) + I_X(2:1). \tag{2.2.2}$$

$I_X(1:2)$ and $I_X(2:1)$ will be referred to as the K-L numbers for X. The K-L numbers and the divergence for Y are similarly defined.

It is noted that if the distribution of X is of the exponential type, i.e., $f_i(x) = \beta(\omega_i)e^{\omega_i x}$, then

$$\begin{aligned} I_X(1:2) &= \log \frac{\beta(\omega_1)}{\beta(\omega_2)} + (\omega_1 - \omega_2)E_{\omega_1}[X]; \\ I_X(2:1) &= \log \frac{\beta(\omega_2)}{\beta(\omega_1)} + (\omega_2 - \omega_1)E_{\omega_2}[X]; \end{aligned} \tag{2.2.3}$$

and

$$J_X = (\omega_1 - \omega_1)(E_{\omega_1}[X] - E_{\omega_2}[X]).$$

Thus, J_X is an interesting measure of the "distance" between H_1 and H_2 relative to the random variable X, being the product of two often-considered measures.

If $I_X(1:2) \geqq I_Y(1:2)$ and $I_X(2:1) \geqq I_Y(2:1)$, then one would say that, in the K-L sense, X is the more informative. When considering an a priori probability, ξ, it seems natural to consider the numbers

$$I_X(\xi) = \xi I_X(1:2) + (1 - \xi)I_X(2:1) \tag{2.2.4}$$

and $I_Y(\xi)$ analogously defined. Then X is the more informative in the K-L sense if and only if $I_X(\xi) \geqq I_Y(\xi)$ for all ξ, i.e., $I_X \geqq I_Y$.

Comparison of $I_X(\xi)$ and $I_Y(\xi)$ provides a very simple criterion for choice between X and Y, especially in the sequential design problem, where after the jth experiment one could simply compute the a posteriori probability, ξ_j, and choose as the $(j + 1)$th experiment that corresponding to the greater of $I_X(\xi_j)$ and $I_Y(\xi_j)$. It is hardly to be expected that this rule would be optimal, but it will be shown that it will never lead to use of Y when $R_X \leqq R_Y$. This rule also possesses some other nice properties described later in the paper.

THEOREM 2.2. *If* $R_X \leqq R_Y$, *then* $I_X \geqq I_Y$.

PROOF: Again with E and F as defined in Lemma 2.1,

$$\int_0 u\, dE(u) = \lim_{\eta \to \infty} \int_{f_2(x)/f_1(x) \leqq \eta} f_2(x)\, d\psi(x) = 1. \tag{1}$$

Similarly, $\int_0^\infty u\, dF(u) = 1$. Hence, for ϕ any linear function

$$\int_0^\infty \phi(u)\, dE(u) = \int_0^\infty \phi(u)\, dF(u). \tag{2}$$

By Lemma 2.1, $R_X \leqq R_Y$ implies

$$\int_0^\infty \min(u-\eta, 0)\, dE(u) \leqq \int_0^\infty \min(u-\eta, 0)\, dF(u). \tag{3}$$

It is easily seen by (2) and (3), then, that for any concave function, ϕ,

$$\int_0^\infty \phi(u)\, dE(u) \leqq \int_0^\infty \phi(u)\, dF(u). \tag{4}$$

In particular, for $\phi(u) = \log u$,

$$I_X(1:2) = -\int_0^\infty \log u\, dE(u) \geqq -\int_0^\infty \log u\, dF(u) = I_Y(1:2); \tag{5}$$

while for $\phi(u) = -u \log u$,

$$-I_X(2:1) = -\int_0^\infty u \log u\, dE(u) \leqq -\int_0^\infty u \log u\, dF(u) = -I_Y(2:1). \tag{6}$$

Equations (5) and (6) yield the conclusion of the theorem.

An immediate corollary is that $R_X = R_Y$ implies $I_X = I_Y$, i.e., $I_X(1:2) = I_Y(1:2)$ and $I_X(2:1) = I_Y(2:1)$.

As will be seen in Section 2.3, $I_X \geqq I_Y$ does not in general imply $R_X \leqq R_Y$. However, in all of the special cases considered by the authors, including the standard distributions—binomial, Poisson, Gamma, and normal—it was true that $I_X = I_Y$ implied $R_X = R_Y$.

An illuminating view of the K-L numbers is obtained (see Theorem 2.3, below) if it is assumed that all the densities under consideration are elements of a class, $\{f_\omega : \omega \in \Omega\}$, of densities positive on the same set, and that there is an Abelian group, T, of transformations of the domain of the f_ω's and a corresponding group of transformations, $\overline{T}$, of transformations of Ω such that if X has density f_ω, then for $t \in T$, $t(X)$ has density $f_{\bar{t}(\omega)} = \mu(t^{-1})f_\omega$, i.e., $d\psi\,(t^{-1}X) = \mu(t^{-1})\, d\psi(x)$. Finally, assume that given ω_1 and ω_2 in Ω, there is a $t \in T$ such that $\omega_2 = \bar{t}(\omega_1)$.

As an example of such a class of densities and groups of transformations, consider the Γ distributions

$$f_\omega(x) = \frac{\omega^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\omega x} (\omega > 0, x > 0), \tag{2.2.5}$$

with $T = \{t_e : t_e(x) = ex, e > 0\}$, $\mu(t_e^{-1}) = e$, and $\bar{t}_e(\omega) = e\omega$.

THEOREM 2.3. *$I_X(1:2)$ and $I_X(2:1)$ are functions only of the transformation that carries f_1 into f_2 and not of f_1 and f_2 individually.*

PROOF. Choose $t \in T$ such that $f_2(x) = f_1(t^{-1}x)\mu(t^{-1})$. Then

$$\begin{aligned} I(1:2) &= \int_{-\infty}^\infty f_1(x) \log \frac{f_1(x)}{f(\bar{t}^{-1}x)\mu(t^{-1})}\, d\psi(x) \\ &= -\log \mu(t^{-1}) + \int_{-\infty}^\infty f_1(x) \log \frac{f_1(x)}{f_1(t^{-1}x)}\, d\psi(x). \end{aligned} \tag{1}$$

To show that the value of the integral is a function of t only and does not depend on f_1, choose $f_0 \varepsilon \{f_\omega\}$ and let $f_1(x) = f_0(s^{-1}x)\mu(s^{-1})$. Then, with $y = s^{-1}x$, $t^{-1}y = t^{-1}s^{-1}x = s^{-1}t^{-1}x$ and (1) can be rewritten as

$$I(1:2) = -\log \mu(t^{-1}) + \int_{-\infty}^{\infty} f_0(y) \log \frac{f_0(y)}{f_0(t^{-1}y)}\, d\psi(y). \tag{2}$$

A similar proof holds for $I(2:1)$ and the proof is complete.

For the example of the Γ distributions, if the parameters under H_1 and H_2 are ω and $a\omega$, respectively, the K-L numbers are given explicitly as functions of a by

$$I_X(1:2) = a[-\log a - 1 + a] \tag{2.2.6}$$

and

$$I_X(2:1) = a\left[\log a - 1 + \frac{1}{a}\right].$$

It can be verified that for this class the relation between T and the K-L numbers is one to one.

Whenever the random variable Y is obtainable from X by a relabeling—more precisely, whenever there is a $t \varepsilon T$ such that Y and $t(X)$ have the same distribution under each hypothesis—we write $Y \leftrightarrow t(X)$. In such a case, clearly, $I_X = I_Y$ and $R_X = R_Y$.

If, as in the Γ distributions, the relation between T and the K-L numbers is one to one, then $I_X = I_Y$ if and only if $R_X = R_Y$, since $I_X = I_Y$ implies $Y \leftrightarrow t(X)$ implies $R_X = R_Y$ implies $I_X = I_Y$. Without the one to one condition we have

THEOREM 2.4. $Y \leftrightarrow t(X)$ implies $R_X = R_Y$, and if the likelihood ratios, f_2/f_1 and g_2/g_1, are monotone in the same direction, then $R_X = R_Y$ implies that $Y \leftrightarrow t(X)$.

PROOF. The first statement is clear. Without loss of generality, let X and Y have the common density h under H_1 and densities f and g respectively under H_2. It then suffices to show that $f = g$.

From Theorem 2.1, if $R_X = R_Y$, then

$$\int_{f(x)/h(x) \leqq \eta} h(x)\, d\psi(x) = \int_{g(x)/h(x) \leqq \eta} h(x)\, d\psi(x) \qquad \text{for all } \eta \geqq 0. \tag{1}$$

Let

$$\left\{x : \frac{f(x)}{h(x)} \leqq \eta\right\} = \{x : x \leqq \gamma_1(\eta)\},$$

and

$$\left\{x : \frac{g(x)}{h(x)} \leqq \eta\right\} = \{x : x \leqq \gamma_2(\eta)\}.$$

Since all densities are assumed positive together, it follows from (1) that $\gamma_1 = \gamma_2$. Hence,

$$\{x:f(x) \leqq \eta h(x)\} = \{x:g(x) \leqq \eta h(x)\}. \tag{2}$$

If $g(x_0) > f(x_0)$, then an η can be found such that $f(x_0) < \eta h(x_0) < g(x_0)$, contradicting (2). If $f(x_0) < g(x_0)$, then a similar contradiction arises. Therefore, $f = g$.

2.3. The case of normal distributions. We now consider the case in which both X and Y have normal distributions under each hypothesis. Since for normal distributions, both the risk function and the K-L numbers are invariant under affine transformations, there is no loss of generality in treating the situation given by

		X	Y
(ξ)	H_1	$N(0, 1)$	$N(0, 1)$
$(1 - \xi)$	H_2	$N(\mu, \sigma^2)$	$N(m, v)$

where $\mu \geqq 0$, $m \geqq 0$, and $\sigma^2 \geqq v$.

The interest in a thoroughgoing study of this case follows from the central role of the normal distribution in statistics, especially in that determination of the cases in which $I_X > I_Y$ and in which $R_X \leqq R_Y$ in terms of the parameters μ, σ^2, m, and v would be of use in asymptotic comparisons for many types of problems.

The K-L numbers for X are:

$$I_X(1:2) = \tfrac{1}{2}\left[\log \sigma^2 - 1 + \frac{1}{\sigma^2} + \frac{\mu^2}{\sigma^2}\right], \tag{2.3.1}$$

and

$$I_X(2:1) = \tfrac{1}{2}\left[\log \frac{1}{\sigma^2} - 1 + \sigma^2 + \mu^2\right].$$

Those for Y are the same, with the obvious substitutions.

THEOREM 2.5. *The following three statements are equivalent:*

(i) $R_X = R_Y$;

(ii) $I_X = I_Y$;

(iii) $\sigma^2 = v$ and $\mu = m$.

PROOF. By Theorem 2.2, (i) implies (ii) and clearly (iii) implies (i). Assuming (ii) to be true,

$$\log\left(\frac{\sigma^2}{v}\right) + \frac{1}{\sigma^2} - \frac{1}{v} = \frac{m^2}{v} - \frac{\mu^2}{\sigma^2}, \tag{1}$$

and

$$\log\left(\frac{v}{\sigma^2}\right) + \sigma^2 - v = m^2 - \mu^2. \tag{2}$$

Then if (iii) is false, it must be that $\sigma^2 > v$, which leads to the following absurdities:

CASE I: $\sigma^2 > 1$. Upon adding equation (1) to $-v^{-1}$ times equation (2), it is seen that μ^2 is of the same sign as

$$A(\sigma^2, v) = (v + 1) \log\left(\frac{\sigma^2}{v}\right) + \frac{v}{\sigma^2} - 1 - \sigma^2 + v. \tag{3}$$

But $A(v, v) = 0$, and $\partial/(\partial\sigma^2)\, A(\sigma^2, v) = 1/\sigma^2\ (\sigma^2 - v)(1 - \sigma^2) < 0$. Hence, $\mu^2 < 0$, a clear absurdity.

CASE II: $\sigma^2 \leqq 1$. Starting with the sum of equation (1) and $-\sigma^{-2}$ times equation (2), a similar argument reaches the absurdity $m^2 < 0$.

The same line of reasoning establishes the

COROLLARY: *For* $v < \sigma^2$, $I_X(1{:}2) \geqq I_Y(1{:}2)$ *implies that* $I_X(2{:}1) > I_Y(2{:}1)$. *For* $v > \sigma^2$, $I_X(2{:}1) \geqq I_Y(2{:}1)$ *implies* $I_X(1{:}2) > I_Y(1{:}2)$.

For a further analysis of the case of normal distributions, assume μ and σ^2 fixed, $\sigma^2 > 1$, and consider the (v, m^2) plane. The results of the remainder of this section may be summarized by figure 1, where the h_i are defined as follows:

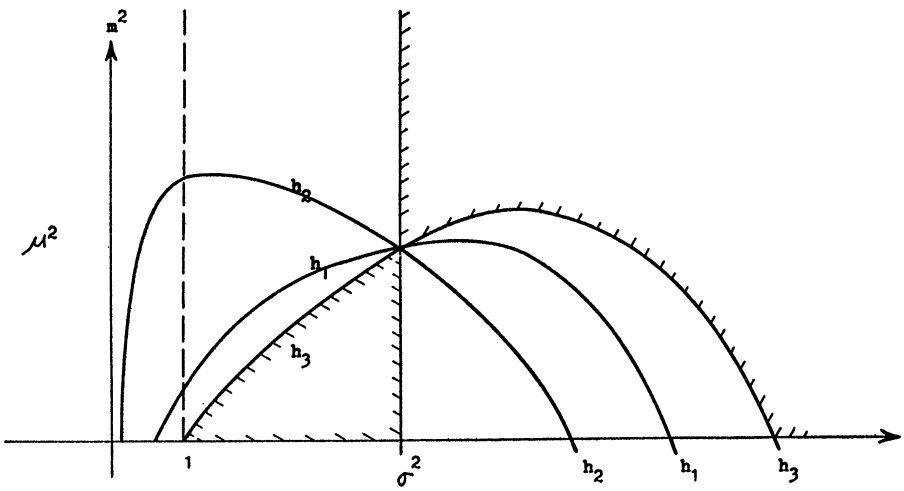

FIG. 1

h_1: $h_1(v) = v\{\log \sigma^2 + (\mu^2 + 1)/\sigma^2\} - v \log (v) - 1$, and $I_X(1{:}2) \geqq I_Y(1{:}2)$ if and only if $m^2 \leqq h_1(v)$;

h_2: $h_2(v) = \mu^2 + \sigma^2 - \log \sigma^2 + \log v - v$, and $I_X(2{:}1) \geqq I_Y(2{:}1)$ if and only if $m^2 \leqq h_2(v)$;

h_3: $h_3(v) = (v - 1)\left\{\frac{\mu^2}{\sigma^2 - 1} + \log \frac{\sigma^2}{v}\right\}$.

And h_3 has two equivalent characterizations:

(i) Let ξ_X be the maximum ξ for which $R_X(\xi) = \xi$, and let ξ_Y be the maximum ξ for which $R_Y(\xi) = \xi$. Then

$$\xi_X \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \xi_Y \quad \text{if and only if} \quad m^2 \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} h_3(v), \quad \text{respectively.}$$

(ii) $R_X \leqq R_Y$ if and only if $v \leqq \sigma^2$ and $m^2 \leqq h_3(v)$.
$R_X \geqq R_Y$ if and only if $v \geqq \sigma^2$ and $m^2 \geqq h_3(v)$.

One of the more curious aspects of this is that no matter how great may be the difference in the means, m and μ, under H_2, that random variable having the smaller variance under H_2 cannot have a risk uniformly less than or equal to that of the random variable having the larger variance.

That $h_1(v) \leqq h_2(v)$ for $v \leqq \sigma^2$ with equality only for $v = \sigma^2$, that $h_1(v) > h_2(v)$ for $v > \sigma^2$, and that $h_2(v) > 0$ follow from Theorem 2.5 and Corollary.

Thus, by Theorem 2.2, a necessary condition for $R_X \leqq R_Y$ is that $m^2 \leqq \min [h_1(v), h_2(v)]$; and for $R_X \geqq R_Y$, that $m^2 \geqq \max [h_1(v), h_2(v)]$.

We now consider the risk functions, and for simplicity it will be assumed until explicitly stated otherwise that $v \geqq 1$ and $\sigma^2 \geqq 1$. The probability of the two types of errors when using the Bayes procedure based on Y are:

$$\alpha_Y(\xi) = 1 - \Pr\left(\left| Y + \frac{m}{v-1} \right| < \frac{\sqrt{v}}{v-1} \sqrt{m^2 + (v-1)\log \eta^2 v} \,|\, H_1\right), \tag{2.3.2}$$

and

$$\beta_Y(\xi) = \Pr\left(\left| Y + \frac{m}{v-1} \right| < \frac{1}{v-1} \sqrt{m^2 + (v-1)\log \eta^2 v} \,|\, H_1\right),$$

where it is to be understood that $\alpha_Y(\xi) = 1$ and $\beta_Y(\xi) = 0$ whenever $m^2 + (v-1)\log \eta^2 v < 0$.

Thus, $R_Y(\xi) = \xi\alpha_Y(\xi) + (1-\xi)\beta_{-Y}(\xi) = \xi$ for $m^2 + (v-1)\log \eta^2 v < 0$, i.e., for small ξ. Furthermore, $R_Y(\xi) < \xi$ for ξ such that $\log \eta^2 > -m^2/(v-1) + \log v$). This latter follows from $d/d\xi\, R_Y(\xi) = \alpha_Y(\xi) - \beta_Y(\xi)$, a fact quite generally true in statistical games with two states of nature, two actions, and a zero-one loss function ([8], Sec. 6.3).

Together with the analogous results for $R_X(\xi)$, it is clear that a necessary condition that $R_X \leqq R_Y$ is that

$$\frac{\mu^2}{\sigma^2 - 1} + \log \sigma^2 \geqq \frac{m^2}{v-1} + \log v,$$

or

$$m^2 \leqq h_3(v) = (v-1)\left\{\frac{\mu^2}{\sigma^2 - 1} + \log \frac{\sigma^2}{v}\right\}. \tag{2.3.3}$$

It may be verified that

$$h_1(v) - h_3(v) \begin{cases} > 0, & 1 \leqq v < \sigma^2; \\ = 0, & v = \sigma^2; \\ < 0, & v > \sigma^2. \end{cases} \tag{2.3.4}$$

Hence, any pair (v, m^2) with $h_3(v) < m^2 < h_1(v)$ and $1 \leqq v < \sigma^2$ provides an example in which X is the more informative in the K-L sense but does not have a uniformly smaller Bayes risk.

For $1 \leqq v \leqq \sigma^2$, the most restrictive necessary condition that $R_X \leqq R_Y$ thus far obtained is that $m^2 \leqq h_3(v)$. Omitting much of the tedious algebraic details, we now show that this is also sufficient.

LEMMA 2.3. *For fixed $v > 1$, $R_Y(\xi)$ is a nonincreasing function of m for each ξ.*

PROOF: From equation (2.3.2) it follows that, with

$$\mathrm{L} = \sqrt{v} \,/\, (v-1)\sqrt{m^2 + (v-1)\log \eta^2 v},$$

$$\sqrt{2\pi}\left[\frac{R_Y(\xi)}{1-\xi} - \eta\right] = -\eta \int_{-L}^{L} \exp\left[-\tfrac{1}{2}\left(t - \frac{m}{v-1}\right)^2\right] dt + \frac{1}{\sqrt{v}} \int_{-L}^{L} \exp\left[-\frac{1}{2v}\left(t - \frac{mv}{v-1}\right)^2\right] dt. \tag{1}$$

Letting $a s b$ denote that a and b are of the same sign,

$$\frac{\partial}{\partial m} R_Y(\xi) \, s \, \frac{\partial}{\partial m}\left[\sqrt{2\pi}\,\frac{R_Y(\xi)}{1-\xi} - \eta\right] = \frac{\partial L}{\partial m} \cdot A + \int_{-L}^{L} \frac{\partial}{\partial m}\left\{\frac{1}{\sqrt{v}} \exp\left[-\frac{1}{2v}\left(t - \frac{mv}{v-1}\right)^2\right] - \eta \exp\left[-\tfrac{1}{2}\left(t - \frac{m}{v-1}\right)^2\right]\right\} dt, \tag{2}$$

where the rather complicated expression denoted here by A can be shown to be zero. Carrying out the indicated differentiation and integration in the second term and effecting a similar reduction, we have

$$\frac{\partial}{\partial m} R_Y(\xi) \, s \, \frac{v-1}{\sqrt{v}}\left\{\exp\left[-\frac{1}{2}v\left(L + \frac{mv}{v-1}\right)^2\right] - \exp\left[-\frac{1}{2v}\left(L - \frac{mv}{v-1}\right)^2\right]\right\}. \tag{3}$$

Since $L \geqq 0$, $\partial \,/\, (\partial m) R_Y(\xi) \leqq 0$, which completes the proof.

Now let ϕ be a nonnegative differentiable function of $v (v > 1)$ with $\phi(\sigma^2) = \mu^2$. Set $m^2 = \phi(v)$ and consider $R_Y(\xi)$ as a function of v. From equation (1) in the proof of Lemma 2.3,

$$\sqrt{2\pi}\left(\frac{R_Y(\xi)}{1-\xi} - \eta\right) = \int_{-L}^{L} \frac{1}{\sqrt{v}} \exp\left[-\frac{1}{2v}(t - vC)^2\right] dt - \eta \int_{-L}^{L} e^{-\frac{1}{2}(t-C)^2} \, dt, \tag{2.3.5}$$

where

$$C = \frac{1}{v-1}\sqrt{\phi(v)}$$

and

$$L = \sqrt{v}\sqrt{C^2 + \frac{\log v\eta^2}{v-1}}.$$

By a sequence of steps parallel to those of the proof of Lemma 2.3, we arrive at the conclusion:

$$\frac{\partial}{\partial v} R_Y(\xi) s \frac{dC}{dv} \frac{v-1}{\sqrt{v}} \exp\left(-\frac{L^2 + v^2C^2}{2v}\right)[e^{-LC} - e^{LC}] + \frac{1}{2v^2\sqrt{v}} \int_{-L}^{L} (t^2 - v^2C^2 - v) \exp\left[-\frac{1}{2v}(t - vC)^2\right] dt, \tag{2.3.6}$$

where

$$\frac{dC}{dv} = \frac{(v-1)\phi'(v) - 2\phi(v)}{2(v-1)^2\sqrt{\phi(v)}}.$$

Now take $\phi = h_3$, in equation (2.3.3); $dC/dv = -(1 + vC^2)/[2v(v-1)C]$. For this choice, the right member of (2.3.6) is nonpositive for all $L \geqq 0$. To show this, consider the right member of (2.3.6) as a function, ψ, of L for $L \geqq 0$. Now, $\psi(0) = \psi(+\infty) = 0$. Hence, it will suffice to show that there is an L^* such that $\psi'(L) \leqq 0$ for $L \leqq L^*$ and $\psi'(L) \geqq 0$ for all $L \geqq L^*$. Differentiating ψ and simplifying,

$$\psi'(L) s e^{2LC}\left[L - \frac{1 + vC^2}{C}\right] + L + \frac{1 + vC^2}{C}. \tag{2.3.7}$$

Let the right member of (2.3.7) be denoted by $\gamma(L)$. Then

$$\gamma'(L) = e^{2LC}\left[1 + 2C\left(L - \frac{1 + vC^2}{C}\right)\right] + 1 \tag{2.3.8}$$

and

$$\gamma''(L) = 4C^2e^{2LC}(L - vC).$$

Hence, γ is concave on the interval $(0, vC)$ and convex on $(vC, +\infty)$. But $\gamma(0) = 0$, $\gamma(+\infty) = +\infty$, and $\gamma'(0) < 0$. Therefore, there exists an L^* such that γ, and hence ψ', is negative for $L < L^*$ and positive for $L > L^*$. In this way the proof is complete for

LEMMA 2.4. *For $m^2 = h_3(v)$, $v \geqq 1$, $R_Y(\xi)$ is a nondecreasing function of v for each ξ.*

Combining the last two lemmas with the fact that for $1 \leqq v \leqq \sigma^2$, $m^2 \leqq h_3(v)$ is a necessary condition for $R_X \leqq R_Y$, while for $v \geqq \sigma^2$, $m^2 \geqq h_3(v)$ is a necessary condition for $R_Y \leqq R_Y$, gives

LEMMA 2.5. *For $1 \leqq v \leqq \sigma^2$, $R_X \leqq R_Y$ if and only if $m^2 \leqq h_3(v)$, while for $v \geqq \sigma^2$, $R_Y \leqq R_X$ if and only if $m^2 \geqq h_3(v)$.*

Since this necessary and sufficient condition that $R_X \leqq R_Y$ for $1 \leqq v \leqq \sigma^2$

was obtained from consideration of the behavior of the risk curves for small ξ, it was conjectured that their behavior for ξ near 1 would yield conditions that $R_X \geqq R_Y$.

LEMMA 2.6. *For v and σ^2 both greater than 1, a necessary condition for $R_Y \leqq R_X$ is that $v \geqq \sigma^2$, and a necessary condition for $R_X \leqq R_Y$ is that $v \leqq \sigma^2$.*

PROOF.

$$\lim_{\xi \to 1} [R_Y(\xi)] / (1 - \xi) = -\lim_{\xi \to 1} R_Y'(\xi) = -\lim_{\xi \to 1} (\alpha_Y(\xi) - \beta_Y(\xi)) = 1,$$

and the same relation holds for $R_X(\xi)$. Then for

$$D(\eta) = \frac{R_Y(\xi) - R_X(\xi)}{1 - \xi}, \tag{1}$$

$D(\eta) \to 0$ as $\eta \to +\infty$.

The method of proof is to show that for all sufficiently large η,

$$D'(\eta) \begin{cases} < 0 & \text{if } v < \sigma^2, \\ > 0 & \text{if } v > \sigma^2, \end{cases}$$

and hence that for ξ sufficiently near 1,

$$R_X(\xi) - R_Y(\xi) \begin{cases} < 0 & \text{for } v < \sigma^2, \\ > 0 & \text{for } v > \sigma^2. \end{cases}$$

$$D'(\eta) = \frac{1}{(1 - \xi)^2} \frac{d\xi}{d\eta} \{(1 - \alpha)[R_Y'(\xi) - R_X'(\xi)] + R_Y(\xi) - R_X(\xi)\}$$

$$\cdot s\alpha_Y(\xi) - \alpha_X(\xi) = \int_{-\mu/(\sigma^2-1)-\Lambda}^{-\mu/(\sigma^2-1)+\Lambda} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt - \int_{m/(v-1)-L}^{-m/(v-1)+L} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt, \tag{2}$$

where

$$L = \sqrt{v} / (v - 1) \sqrt{m^2 + (v - 1) \log \eta^2 v} \qquad \text{and}$$

$$\Lambda = \sigma / (\sigma^2 - 1) \sqrt{\mu^2 + (\sigma^2 - 1) \log \eta^2 \sigma^2}.$$

As $\eta \to +\infty$,

$$\frac{L}{\Lambda} \to \sqrt{\frac{v(\sigma^2 - 1)}{(v - 1)\sigma^2}} \begin{cases} > 1 & \text{for } v < \sigma^2, \\ < 1 & \text{for } v > \sigma^2, \end{cases} \tag{3}$$

and both L and $\Lambda \to +\infty$. Hence, for $v < \sigma^2$, $L - \Lambda \to +\infty$ and for all sufficiently large η,

$$\left(-\frac{m}{v - 1} - L, -\frac{m}{v - 1} + L\right) \supset \left(-\frac{\mu}{\sigma^2 - 1} - \Lambda, -\frac{\mu}{\sigma^2 - 1} + \Lambda\right), \tag{4}$$

and it follows that $D'(\eta) < 0$. On the other hand, if $v > \sigma^2$, $\Lambda - L \to +\infty$; and the same reasoning shows that for all sufficiently large η, $D'(\eta) > 0$.

It has, then, been completely determined, for $v > 1$ and $\sigma^2 > 1$, precisely when $R_X \leqq R_Y$ and precisely when $R_X \geqq R_Y$, namely,

THEOREM 2.6. *For $v > 1$ and $\sigma^2 > 1$,*

(i) *$R_X \leqq R_Y$ if and only if $v \leqq \sigma^2$ and $m^2 \leqq h_3(v)$, and*

(ii) *$R_X \geqq R_Y$ if and only if $v \geqq \sigma^2$ and $m^2 \geqq h_3(v)$.*

The case $v < 1$ and $\sigma^2 < 1$ may clearly be converted, by interchange of H_1 H_2, to the above case; whereas if v and σ^2 lie on opposite sides of 1, neither $R_X \leqq R_Y$ nor $R_Y \leqq R_X$ can hold, since for $v < 1$, $R_Y(\xi) = 1 - \xi$ on an interval $(\xi_1, 1)$.

2.4. Locally more informative experiments. The notion of being "locally more informative" arises naturally from the example just analysed. We will say that X is locally more informative than Y at ξ if ξ lies in an interval $[\xi_1, \xi_2]$ such that on $[\xi_1, \xi_2] \cap [0, 1]$, $R_X(\xi) \leqq R_Y(\xi)$ with strict inequality at one end point. The example of the normal distributions provides an instance in which a necessary and sufficient condition that $R_X \leqq R_Y$ is that X be locally more informative than Y at the two points 0 and 1.

For binomial distributions, the form of the risk curves makes it immediately clear that locally more informative at 0 and 1 is equivalent to uniform inequality of risks. Thus, if the parameters are given by

	X	Y	
H_1	p_1	q_1	
H_2	p_2	q_2	$(p_2 > p_1)$,

then, in terms of the parameters, $R_X \leqq R_Y$ if and only if

$$\frac{p_2}{p_1} \geqq \frac{q_2}{q_1} \geqq 1 \geqq \frac{1 - q_2}{1 - q_1} \geqq \frac{1 - p_2}{1 - p_1} \quad \text{or} \quad \frac{p_2}{p_1} \geqq \frac{1 - q_2}{1 - q_1} \geqq 1 \geqq \frac{q_2}{q_1} \geqq \frac{1 - p_2}{1 - p_1}.$$

Another instance in which locally more informative at 0 and 1 is equivalent to uniform inequality of risks is the case of the Γ distribution (equation 2.2.5). From this it is easily found that if the parameters are given by

	X	Y
H_1	ω_1	θ_1
H_2	ω_2	θ_2,

then $R_X \leqq R_Y$ if and only if $\omega_2/\omega_1 \geqq \theta_2/\theta_1 \geqq 1$ or $\omega_2/\omega_1 \leqq \theta_2/\theta_1 \leqq 1$.

The conjecture that if all the distributions belong to the same exponential family, then being locally more informative at zero and one is equivalent to uniform inequality of risks, may be shown to be false, however, by the example in which $d\psi(x) = e^{-x^2/2}x^2\,dx$.

3. Optimal designs for a binomial testing problem. If we now consider the case in which a total of n experiments is to be carried out, the question arises

as to how difficult it would be to obtain, for a specific case, the optimal sequential designs. If they are practically obtainable, the interest in any other design criteria which have some justification though not optimal is reduced to pure curiosity. It is found, however, that this is not the situation. For the case of binomial distributions, the optimal designs for small n were found and are given below for a few of the cases in order to illustrate the intrinsically complicated structure of the optimal designs.

Suppose that X and Y have binomial distributions with parameters given by

	X	Y
H_1	p	q
H_2	q	p,

that the observations are independent, that the total number of observations to be taken, n, is fixed, and that the cost of observation is independent of the true hypothesis, of the random variable observed, and of the outcome of the observation.

The complete form of the solution for the case $n = 3$ will now be exhibited. The solutions are expressed in terms of [0, 1] on the η axis corresponding to subintervals of [0, $\frac{1}{2}$] on the ξ axis, which by symmetry may be extended to [0, 1] on the ξ axis. (I denotes indifference.) We only present here the best optimal design of the first experiment for the case $n = 3$. The description of the optimal designs for the remaining steps will be omitted.

(3.1) For $p(1-p)^2 > q(1-q)^2$ and $p(1-p)^3 < q(1-q)^3$:

optimal choice :	I	:	X	:	I	:	X	:	I	:	Y	:	
η	0		$\left(\frac{q}{p}\right)^3$		$\left(\frac{q}{p}\right)^2$		$\frac{q^2(1-q)}{p^2(1-p)}$, $\frac{q(1-q)}{p(1-p)}$		$\frac{1-p}{1-q}$		1		

For $p(1-p)^2 > q(1-q)^2$ and $p(1-p)^3 > q(1-q)^3$:

optimal choice :	I	:	X	:	I	:	X	:	I	:	X	:	
η	0		$\left(\frac{q}{p}\right)^3$		$\left(\frac{q}{p}\right)^2$		$\frac{q^2(1-p)}{p^2(1-p)}$, $\frac{q(1-q)}{p(1-p)}$		$\frac{q(1-q)^2}{p(1-p)^2}$		1		

For $p(1-p)^2 < q(1-q)^2$ and $q^2(1-q)^3 < p^2(1-p)^3$:

optimal choice :	I	:	X	:	I	:	X	:	I	:		:	
η	0		$\left(\frac{q}{p}\right)^3$		$\left(\frac{q}{p}\right)^2$,		$\frac{q^2(1-q)}{p^2(1-p)}$	$\frac{q}{p}$	$\frac{q(1-q)}{p(1-p)}$			1	

For $p(1-p)^2 < q(1-q)^2$ and $q^2(1-q)^3 > p^2(1-p)^3$:

optimal choice :	I	:	X	:	I	:	Y	:	X	:	I	:	
η	0		$\left(\frac{q}{p}\right)^3$		$\left(\frac{q}{p}\right)^2$		$\left(\frac{1-p}{1-q}\right)^2$	A	$\frac{q}{p}\frac{q(1-q)}{p(1-p)}$		1		

where

$$A = \frac{(1-p)^2(1-p-q) - q^2(1-q)}{(1-q)^2(1-p-q) - p^2(1-p)}.$$

4. A nontruncated design problem. In the preceding sections attention has been fixed entirely on design problems in which the sample size was fixed and in which we were interested in reducing the Bayes risk. In considering nontruncated procedures, any design will, if only it calls for a sufficiently large number of observations of relevant random variables, reduce the risks from the terminal decision below any arbitrary level. What we do here is fix a risk, r, and consider certain designs with a view to minimizing the expected number of experiments required to reduce the risk from the terminal decision to at most r. If the experiments have equal cost, this is equivalent to reducing the risk to at most r in the most economical way.

One idea in considering the nontruncated problem below was to create a symmetric problem in the hope that more satisfactory results could be obtained.

Suppose, as in the previous section, that X and Y have binomial distributions with parameters given by

		X	Y	
(ξ)	H_1	p	q	
$(1-\xi)$	H_2	q	p	$(p > 0, p(1-p) > q(1-q))$.

The problem can be stated as being that of minimizing the expected number of observations required to move the a posteriori probability for H_1 to a position in either $[0, r]$ or $(1-r, 1]$.

Three reasonable rules are considered and shown to be equivalent, and are shown, by a more general theorem, to be better than always using either one of the experiments.

Let $\xi_0 = \xi$ and let ξ_j denote the a posteriori probability for H_1 after having made the first j observations. It will be convenient to consider the problem in terms of the variable $\gamma = \log \eta = \log \xi / (1-\xi)$. Then let

$$a = \log \frac{p}{q},$$

$$b = \log \frac{1-p}{1-q},$$

and

$$A = -\log \frac{r}{1-r}.$$

Let n denote the smallest value for which either $\xi_j \leqq r$ or $\xi_j \geqq 1-r$. It is seen there are two random walks, both on the γ-axis with boundaries at A and $-A$, one of which is determined by the results of observations of X and the other is determined by the results of observations of Y. After having made j observations one finds that the walk has arrived at the point γ_j. Now the choice

must be made as to whether it is better that γ_{j+1} should be determined by an observation of X or of Y, i.e., whether the next step should be taken in the X-walk or in the Y-walk. A rule is desired prescribing for every situation which walk should be taken in order to minimize the expected value of n.

If at the $(j+1)$th step X is observed, the expected movement in the walk is

$$E_X[\gamma_{j+1} - \gamma_j] = \xi_j J_X - I_X(2:1). \tag{4.1}$$

Since J_X is positive, $E_X[\gamma_{j+1} - \gamma_j]$ is an increasing function of γ_j and is positive for $\gamma_j > -\gamma^* = \log [I_X(2:1)] / J_X$.

Similarly, if Y is observed, the expected movement is

$$E_Y[\gamma_{j+1} - \gamma_j] = \xi_j J_X - I_X(1:2), \tag{4.2}$$

which is again an increasing function of γ_j and is positive for $\gamma_j > \gamma^*$.

It can be verified that $\gamma^* > 0$ and hence, for $\gamma_j > 0 (\xi_j > \frac{1}{2})$, the X-walk yields an expected step greater in magnitude than the Y-walk and in the "right" direction, i.e., toward the nearest boundary, A. For $\gamma_j < 0$, the Y-walk enjoys the same advantages, $-A$ being the nearest boundary.

These considerations lead to the conjecture that, at least for a small relative to A, the optimal design may be to take the X-walk on the $(j+1)$th step if $\gamma_j > 0$ and to take the Y-walk otherwise. (If $p(1-p) < q(1-q)$, the same results will hold with X and Y interchanged.)

Another reasonable rule would be to use X on the $(j+1)$th step whenever, starting from γ_j, observation of X consistently to termination of experimentation has a smaller expected number of observations than consistent observation of Y. Let X_∞ denote the rule requiring X at every step, let $\mathrm{E}[n \mid X_\infty, \gamma, H_i]$ denote the expected number of steps in the X-walk with ξ as the starting point when H_i is the true hypothesis, and let

$$E[n \mid X_\infty, \gamma] = \xi E[n \mid X_\infty, \gamma, H_1] + (1-\xi) E[n \mid X_\infty, \gamma, H_2].$$

Using Wald's well-known approximations, we find that

$$\begin{aligned} &E[n \mid Y_\infty, \gamma] - E[n \mid X_\infty, \gamma] \\ &\quad = \frac{I_X(1:2) - I_X(2:1)}{I_X(1:2) I_X(2:1)} \left\{ \xi \left[A - \gamma - 2A \frac{e^{-2A} - e^{-A-\gamma}}{e^{-2A} - 1} \right] \right. \\ &\qquad \left. + (1-\xi) \left[A - \gamma - 2A \frac{e^{-2A} - e^{A+\gamma}}{e^{2A} - 1} \right] \right\} \end{aligned} \tag{4.3}$$

$$\underset{\sim}{s}\psi(\gamma) = -\gamma + A(2\xi - 1) + 2A \frac{e^{A+\gamma} - 1}{e^{2A} - 1} - 2A \left\{ \frac{e^{-2A} - e^{-A-\gamma}}{e^{-2A} - 1} + \frac{e^{A+\gamma} - 1}{e^{2A} - 1} \right\}.$$

$\psi(0) = \psi(A) = 0$ and, at least for $A \geqq \log \frac{3}{2}$, $\psi''(\gamma)$ is positive and then negative as γ increases from 0 to A. Hence, $\psi(\gamma) \geqq 0$ on $[0, A]$ if $\psi'(0)$ can be shown to be positive. Now,

$$\psi'(0) s 4 + e^{2A}(A-2) + 2Ae^{A} - 2Ae^{-A} - e^{-2A}(A+2). \tag{4.4}$$

At $A = 0$, the right member of equation (4.4), together with its first three derivatives with respect to A, is zero, while the fourth derivative is positive for all $A \geqq 0$. Therefore, $\psi'(0) > 0$ for $A > 0$, and it follows that

$$E[n \mid Y_\infty, \gamma] - E[n \mid X_\infty, \gamma] > 0 \quad \text{for} \quad \gamma > 0; \tag{4.5}$$

and by the evident symmetry,

$$E[n \mid Y_\infty, \gamma] - E[n \mid X_\infty, \gamma] < 0 \quad \text{for} \quad \gamma < 0. \tag{4.6}$$

Thus, the design which requires the use of X on the $(j+1)$th step if $\gamma_j > 0$ and of Y if $\gamma_j < 0$ coincides with

(i) Choosing the random variable which has the largest expected movement and that toward the nearest boundary, and

(ii) Choosing the random variable corresponding to the smaller of $E[n \mid X_\infty, \gamma_j]$ and $E[n \mid Y_\infty, \gamma_j]$.

It is easily seen that it also coincides with

(iii) Choosing the random variable corresponding to the larger of $I_X(\xi_j)$ and $I_Y(\xi_j)$.

Of course, in view of the fact that we have used the Wald approximations, the description in (ii) should be understood as being only approximately valid.

Let the design thus characterized in four ways be denoted by M. This section concludes with a general result which shows that M is better than either X_∞ or Y_∞.

By a *stationary* design we shall mean a design in which the choice at the $(j+1)$th step is a function only of the a posteriori probability after the jth step, ξ_j.

THEOREM 4.1. *Let X and Y have densities f_i and g_i, respectively, under hypothesis H_i such that both* $\log f_2/f_1$ *and* $\log g_2/g_1$ *assume positive and negative values with positive probability. Let D_1 and D_2 be two stationary designs, and let D be that design which requires, at the $j+1$th step, the random variable corresponding to the smaller of $E[n \mid D_1, \gamma_j]$ and $E[n \mid D_2, \gamma_j]$. Then $E[n \mid D, \gamma] \leqq \min\{E[n \mid D_1, \gamma], E[n \mid D_2, \gamma]\}$.*

PROOF. Let $\Gamma_i = \{\gamma$: for $\gamma_j = \gamma$, D_i requires X at the $(j+1)$th step$\}$. With

$$T_X(\xi) = \frac{\xi f_1(X)}{\xi f_1(X) + (1-\xi) f_2(X)} \quad \text{and} \quad T_X(\gamma) = \log \frac{T_X(\xi)}{1 - T_X(\xi)},$$

$$E[n \mid D_i, \gamma] = \begin{cases} 1 + E_\gamma[n \mid D_i, T_X(\gamma)], & \text{if } \gamma \in \Gamma_i, \\ 1 + E_\gamma[n \mid D_i, T_Y(\gamma)], & \text{if } \gamma \notin \Gamma_i, \\ 0, & \text{if } |\gamma| > A. \end{cases} \tag{1}$$

By some simple probability considerations it is easily seen that for any stationary design D', $E[n \mid D', \gamma]$ is uniformly bounded in γ.

Define the function H and the set Θ by

$$H(\gamma) = \min_i E[n \mid D_i, \gamma] = \begin{cases} E[n \mid D_1, \gamma] & \text{for } \gamma \in \Theta, \\ E[n \mid D_2, \gamma] & \text{for } \gamma \notin \Theta, \\ 0 & \text{for } |\gamma| > A. \end{cases} \tag{2}$$

Then with $\Gamma = (\Gamma_1 \cap \Theta) \cup (\Gamma_2 - \Theta)$ (which represents the region according to the design D where D requires X), we have

$$G(\gamma) = H(\gamma) - E(n \mid D, \gamma)$$

$$= \begin{cases} E_\gamma(n \mid D_i' T_X(\gamma)) - E_\gamma[n \mid D, T_X(\gamma)], & \gamma \in \Gamma, \\ E_\gamma(n \mid D_j' T_Y(\gamma)) - E_\gamma[n \mid D, T_Y(\gamma)], & \gamma \notin \Gamma, \\ 0, & |\gamma| > A, \end{cases} \tag{3}$$

where D_i' and D_j' are the appropriate designs. It is also clear that D_i' calls for a continuation of $T_X(\gamma)$ when γ is in Γ, and a similar result is valid for D_j'. From (2) and (3), we obtain that

$$G(\gamma) \geqq \begin{cases} E_\gamma[G(T_X(\gamma)], & \gamma \in \Gamma, \\ E_\gamma[G(T_Y(\gamma)], & \gamma \notin \Gamma, \\ 0, & |\gamma| > A. \end{cases} \tag{4}$$

It suffices to show $G(\gamma) \geqq 0$. Let γ_n denote a sequence of points where $|\gamma_n| < A$ and such that $\inf_\gamma G(\gamma) = \lim_{n \to \infty} G(\gamma_n)$. It follows from (4) that either a subsequence $\gamma_{n_i} + \log f_2(x) / f_1(x)$ exists such that

$$\inf_\gamma G(\gamma) = \lim_{n_i} G(\gamma_{n_i} + \log \frac{f_2(x)}{f_1(x)}$$

with probability 1, or a subsequence γ_m of γ_n exists such that

$$\inf_\gamma G(\gamma) = \lim_{m_i} G\left(\gamma_{m_i} + \log \frac{g_2(x)}{g_1(x)}\right)$$

with probability 1. As both likelihood ratios are less than one with positive probability, there exists a choice $\lambda (\lambda > 0)$ such that either $\lim_i G(\gamma_{n_i} - \lambda) = \inf G(\gamma)$ or $\lim_i G(\gamma_{m_i} - \lambda) = \inf G(\gamma)$. Let us again denote the sequence which fulfills this last requirement by $\gamma_n^{(1)}$ so that $\lim_n G(\gamma_n^{(1)}) = \inf G(\gamma)$. Repeating the argument a finite number of times, we arrive at a sequence $\gamma_n^k \leqq -A$ and $\lim_n G(\gamma_n^{(k)}) = 0 = \inf_\gamma G(\gamma)$, or $G(\gamma) \geqq 0$.

REFERENCES

[1] A. WALD, *Statistical Decision Functions*, John Wiley & Sons, New York (1950).

[2] C. A. MAGWIRE, *Sequential Decisions Involving the Choice of Experiments*, Technical Report No. 19, Department of Statistics, Stanford University, Palo Alto, Calif. (1953).

[3] H. ROBBINS, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, Vol. 58, No. 5 (1952), pp. 527–535.

[4] D. BLACKWELL, "Equivalent comparisons of experiments," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 265–272.

[5] D. BLACKWELL, "Comparison of experiments," *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951.

[6] S. KULLBACH AND R. A. LEIBLER, "On information and sufficiency," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 79–86.

[7] S. KULLBACH, "An application of information theory to multivariate analysis," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 88–102.

[8] BLACKWELL D. AND GIRSHICK, M. A., *Theory of Games and Statistical Decisions*, John Wiley & Sons, New York (1954).