The Annals of Mathematical Statistics
1972, Vol. 43, No. 5, 1428-1440

SOME PROPERTIES OF BAYESIAN ORDERINGS OF EXPERIMENTS[1]

By D. Feldman

Stanford University and Michigan State University[2]

Let Ω be a finite set of states and Ξ the class of prior distribution on Ω. A nonnegative, continuous, concave function on Ξ is called an uncertainty function and if $X = (\mathscr{X}, \mathscr{A}; P_\theta, \theta \in \Omega)$ and $Y = (\mathscr{Y}, \mathscr{B}; Q_\theta, \theta \in \Omega)$ are two experiments X is called at least as informative as Y with respect to U if

$$U(\xi \mid X) \leqq U(\xi \mid Y) \qquad \text{for all } \xi \in \Xi$$

where $U(\xi \mid X)$ is the expected posterior uncertainty for an observation on X when the prior is $\xi \in \Xi$. Any such U induces a partial ordering on the class of all experiments. The paper characterizes (i) the class of functions U which lead to a total ordering of the class of experiments and (ii) the class of transformations of a function U which preserve its induced ordering.

1. Introduction and summary. Consider a decision problem (Ω, A, L), where Ω is a set of states, A an action space and L a loss function on $\Omega \times A$. Let $E(\Omega)$ denote the class of all possible experiments with parameter space Ω and suppose that the decision maker is at liberty to perform one of the two experiments, $X, Y \in E(\Omega)$ before choosing an act $a \in A$. If we further assume that there is a prior distribution ξ over Ω, the decision as to which of the two experiments to perform reduces to calculating which of the experiments has the smaller expected posterior risk. In typical situations this choice will depend heavily on the nature of the decision problem and on the particular prior in question. When these are fixed $E(\Omega)$ can be totally ordered in the sense that for any pair of experiments one is preferred or indifferent to the other.

There has been considerable interest in the (partial) ordering on $E(\Omega)$ induced by the requirement that the relation described above hold whatever be the decision problem and the prior distribution. In such papers as Blackwell [1], [2] and LeCam [7], for example, it is shown that under rather general conditions this global requirement is equivalent to the relation of Blackwell sufficiency between the experiments. In this paper we shall be concerned with certain properties of orderings of $E(\Omega)$ induced by an intermediate requirement, namely that the expected posterior risk of X be smaller than that of Y for every prior ξ, but for a fixed decision problem. In particular we shall investigate (i) classes of decision problems which totally order $E(\Omega)$ and (ii) transformations of a decision problem which preserve the order relation in $E(\Omega)$.

If the sufficiency relation is at one end of the spectrum of Bayesian orderings

---

Received August 31, 1970; revised April 19, 1971.

[1] This research was in part sponsored by the Army Research Office, Office of Naval Research and Air Force Office of Scientific Research by Contract No. N00014-67-0112-0053 (NR-042-267).

[2] On sabbatical leave from Michigan State University, 1969-1970.

in $E(\Omega)$ problem (i) addresses itself to orderings at the opposite end. The solution of problem (ii) provides insight for determining those decision problems for which a specified experiment is desirable when it is known to be desirable for some given problem.

Let Ξ be the class of all priors over Ω. The concave functional

$$U(\xi) = \inf_a \int L(\theta, a)\, d\xi(\theta)\,,$$

defined for all $\xi \in \Xi$ for which the integral makes sense, is the Bayes envelope of the decision problem and we follow de Groot [5] in calling U the uncertainty function of the problem. In [5] it is shown that (when the number of states is finite) any concave functional on Ξ is the Bayes envelope of some decision problem, and, since our interest is primarily in the posterior behavior of this function we shall identify a decision problem with the uncertainty function it induces on Ξ.

Let $X = (\mathscr{X}, \mathscr{A}; P_\theta, \theta \in \Omega)$ denote an experiment with measurable sample space $(\mathscr{X}, \mathscr{A})$ and a family of probability models $P_\theta, \theta \in \Omega$, and let $Y = (\mathscr{Y}, \mathscr{B}; Q_\theta, \theta \in \Omega)$ denote an alternative experiment. We say that X is at least as (more) informative as (than) Y with respect to U if $U(\xi \mid X) \leqq U(\xi \mid Y)$ for all $\xi \in \Xi$ (and $U(\xi \mid X) < U(\xi \mid Y)$ for some $\xi \in \Xi$) where $U(\xi \mid X)$ is the expected posterior uncertainty if X is observed and the prior distribution is ξ and $U(\xi \mid Y)$ is the corresponding value for an observation on Y.

The relation "at least as informative as with respect to U" induces a partial ordering on $E(\Omega)$ and can be represented as a subset of $E(\Omega) \times E(\Omega)$ which we shall denote by $\Pi(U)$. We then express the fact that "X is at least as (more) informative as (than) Y" as $(X, Y) \in \Pi(U)((X, Y) \in \Pi(U)$ and $(Y, X) \notin \Pi(U))$.

The propositions we shall state will all be proved under the following basic assumptions concerning the decision models:

(i) Ω is a finite set with N elements. Ξ is then the N-dimensional simplex consisting of all points $\xi = (\xi(\theta_1), \cdots, \xi(\theta_N))$ with $\xi(\theta_i) \geqq 0$ and $\sum \xi(\theta_i) = 1$.

(ii) U is nonnegative, continuous and concave on Ξ and $U(\xi) = 0$ at all degenerate probability vectors $\xi = (0, 0, \cdots, 0, 1, 0, \cdots 0)$. This assumption essentially means that the payoff or loss function of the problem has been transformed to a regret function, which does not affect the class of Bayes procedures.

(iii) $E(\Omega)$ consists of all those experiments $X = (\mathscr{X}, \mathscr{A}; P_\theta, \theta \in \Omega)$ such that $P_\theta, \theta \in \Omega$ is a dominated family with densities $p_x(\theta)$ at x and with the property that the mapping

$$x \to p_x = (p_x(\theta_1), p_x(\theta_2), \cdots, p_x(\theta_N)) \in R_N$$

is Borel.

Under these assumptions, we show in Section 3 that the only functions U which induce a total ordering in $E(\Omega)$ are geometric means i.e., functions of the form $U(\xi) = \prod_{i=1}^N [\xi(\theta_i)]^{t_i}$ for some fixed $t = (t_1, \cdots, t_N)$ with $t_i \geqq 0$ and

$\sum_{i=1}^{N} t_i = 1$. The ordering amounts to a comparison of one value of the Laplace transforms of the experiments being compared.

In Section 4 we show that $\Pi(U_1) = \Pi(U_2)$ if and only if the loss function corresponding to U_2 differs from that of U_1 by a multiplicative function of θ which is nowhere 0.

2. Preliminaries. Let $\Omega = (\theta_1, \theta_2, \cdots, \theta_N)$ be a finite set, and Ξ the class of all prior distributions on Ω. If $\xi \in \Xi$ then $\xi = (\xi(\theta_1), \xi(\theta_2), \cdots, \xi(\theta_N))$ with $\xi(\theta_i) \geqq 0$ and $\sum_i \xi(\theta_i) = 1$. The uncertainty function U is taken to be nonnegative, continuous, and concave on Ξ with the additional property that $U(\xi) = 0$ for every degenerate probability function ξ. Without loss of generality we then assume that any experiment $X = (\mathscr{X}, \mathscr{A}; P_\theta, \theta \in \Omega)$ is dominated and we denote the density of P_θ at x, with respect to the dominating measure, by $p_x(\theta)$—the likelihood function at x is then denoted by p_x. If ξ is the prior probability function and x is an observed value of X, the posterior probability function becomes:

$$\xi_x = \frac{\xi p_x}{\sum_i \xi(\theta_i) p_x(\theta_i)} = \frac{\xi p_x}{p_x(\xi)}$$

where ξp_x is the ordinary product of the functions ξ and p_x.

Let $E(\Omega)$ denote the class of all experiments with parameter space Ω. We define the ordering induced by U on $E(\Omega)$ as the subset, $\Pi(U)$, of $E(\Omega) \times E(\Omega)$ characterized by: $(X, Y) \in \Pi(U)$ if

$$U(\xi \mid X) = \int_{\mathscr{X}} U(\xi_x) p_x(\xi)\, d\mu(x) \leqq \int_{\mathscr{Y}} U(\xi_y) q_y(\xi)\, d\eta(y) = U(\xi \mid Y)$$

for every $\xi \in \Xi$.

If $(X, Y) \in \Pi(U)$ and $(Y, X) \notin \Pi(U)$ we say that X is *more informative than Y with respect to U*, or, we may interpret $\Pi(U)$ as a preference relation.

Typically, the ordering $\Pi(U)$ will be a partial ordering of $E(\Omega)$. The ordering induced by U is *total* if for every pair of experiments X, Y either $(X, Y) \in \Pi(U)$ or $(Y, X) \in \Pi(U)$.

Instead of dealing with U and ξ_x directly we will use a convenient extension of U defined as follows: Let $F^+(\Omega) = F^+$ denote the class of all nonnegative functions, f, on Ω (F^+ is the orthant of nonnegative vectors in N-space). For every $f \in F^+$ define

$$V(f) = U\left(\frac{f}{\sum_i f(\theta_i)}\right) \Sigma f(\theta_i) \quad \text{and}$$

$$V(f \mid X) = \int_{\mathscr{X}} V(f p_x)\, d\mu(x)\,.$$

The properties of U on Ξ guarantee the following for V on F^+:

(2.1)
- (i) continuity
- (ii) superadditivity: $V(f + g) \geqq V(f) + V(g)$, $f, g \in F^+$
- (iii) nonnegative homogeneity: $V(cf) = cV(f)$, $c > 0, f \in F^+$
- (iv) $V(f) = 0$ for all f with only one nonzero component
- (v) For any experiment X, $V(f \mid X) \leqq V(f)$ for every $f \in F^+$.

A 1–1 correspondence between functions U on Ξ and V on F^+ is established by noting that V is determined by its restriction to any subset $F_0 \subset F^+$ with the property that for every $f \in F^+$ there exists $c > 0$ such that $cf \in F_0$. In particular we may take $F_0 = \{f\colon \sum_i f(\theta_i) = 1\}$, which yields U.

If U is the uncertainty function for the decision problem (Ω, A, L) then $U(\xi) = \inf_a \sum_i L(\theta_i, a)\xi(\theta_i)$. The function V merely extends the domain of U to F^+ and is related to the decision problem by:

$$V(f) = \inf_a \textstyle\sum_i L(\theta_i, a)f(\theta_i) \qquad \text{for all } f \in F^+ . \tag{2.2}$$

An experiment, X, is called *completely informative* with respect to V if $V(f \mid X) = 0$, for all $f \in F^+$. At the other extreme, any X for which $V(f \mid X) = V(f)$ for every $f \in F^+$ is called *uninformative*. An uninformative experiment is an auxiliary randomization or, alternatively, an experiment whose likelihood functions are concentrated on the ray: $\{f\colon f(\theta_i) \equiv c$, for some $c > 0\}$. Henceforth we denote the function identically equal to c by c^*.

Note that completely informative experiments always exist: take, for example, any family of N mutually singular probability measures.

Let X_T denote the family of experiments $\{X_t, t \in T\}$. λX_T denotes the randomized experiment obtained by choosing t according to the probability measure λ over T and observing X_t. Then

$$V(f \mid \lambda X_T) = \textstyle\int_T V(f \mid X_t)\, d\lambda(t) . \tag{2.3}$$

If X is sufficient for Y then $(X, Y) \in \Pi(V)$ for every V([5]) and, in fact, the converse is also true (e.g. [2]). For $N = 2$ there is a V_0 such that $(X, Y) \in \Pi(V_0) \Leftrightarrow (X, Y) \in \Pi(V)$ for every V satisfying the conditions (2.1). Using coordinate notation, the V_0 that works in $V_0(a, b) = \min(a, b)$ as is shown in [1] and [3] (the corresponding U_0 is $U_0(\xi) = \min(\xi, 1 - \xi)$). The decision problem corresponding to this V_0 (or U_0) is a testing problem—indeed any non-degenerate testing problem (with $N = 2$) generates the same ordering. We will show in Section 4 that for higher dimensions the situation is far more complex and that no such V_0 exists.

We consider now some examples of other uncertainty functions for the case $N = 2$ ($\Xi = [0, 1]$).

EXAMPLE 1. $U(\xi) = \xi(1 - \xi)$. This uncertainty function is the one corresponding to a squared error loss problem on a two state space i.e. U is the variance of the prior distribution on $\Omega = (0, 1)$. The corresponding V is:

$$V(a, b) = \frac{ab}{a + b} .$$

$U(\xi \mid X)$ is shown in [9] to be related to the Fisher information in the family $p_x(\xi)$, $\xi \in [0, 1]$, treating ξ as the parameter of the family. For binomial experiments we get an explicit solution to the ordering problem as follows: let

X, Y be binomial experiments with parameters (p_1, p_2) and (r_1, r_2) respectively. Then $(X, Y) \in \Pi(V) \Leftrightarrow$ for every $a, b \geqq 0$:

$$\frac{p_1 p_2}{ap_1 + bp_2} + \frac{q_1 q_2}{aq_1 + bq_2} \leqq \frac{r_1 r_2}{ar_1 + br_2} + \frac{s_1 s_s}{as_1 + bs_2}$$

where $q_i = 1 - p_1$ and $s_i = 1 - r_i$. After algebraic reduction the ordering can be seen to be:

$$(X, Y) \in \Pi(U) \Leftrightarrow \frac{(p_1 - p_2)^2}{p_1(1 - p_1)} \geqq \frac{(r_1 - r_2)^2}{r_1(1 - r_1)} \quad \text{and} \quad \frac{(p_1 - p_2)^2}{p_2(1 - p_2)} \geqq \frac{(r_1 - r_2)^2}{r_2(1 - r_2)} .$$

Note that $(p_1 - p_2)^2/(p_i(1 - p_i))$ is the Fisher information of the binomial family $\xi p_1 + (1 - \xi)p_2$, $\xi \in [0, 1]$ at $\xi = 0$ for $i = 1$ and at $\xi = 1$ for $i = 2$.

EXAMPLE 2. $U(\xi) = [\xi(1 - \xi)]^{\frac{1}{2}}$. This U is the standard deviation of the prior distribution and the ordering it induces is vastly different from that of Example 1.

$$U(\xi \mid X) = [\xi(1 - \xi)]^{\frac{1}{2}} \int_{\mathscr{X}} [p_1(x)p_2(x)]^{\frac{1}{2}} \, d\mu(x) .$$

Hence, the ordering induced by U is total, depending only on the magnitude of the number $\int_{\mathscr{X}} [p_1(x)p_2(x)]^{\frac{1}{2}} \, d\mu(x)$. The corresponding V for this function is $V(a, b) = (ab)^{\frac{1}{2}}$.

EXAMPLE 3. $U(\xi) = -\xi \log \xi - (1 - \xi) \log (1 - \xi)$. This is the Shannon information suggested for comparing experiments by Lindley [8]. The V corresponding to U, here, is:

$$V(a, b) = (a + b) \log (a + b) - a \log a - b \log b .$$

Although this function has received a great deal of attention, very little is known concerning the ordering it induces.

3. Total orderings. We begin by proving a sequence of simple lemmas, each a consequence of the assumption that $\Pi(V)$ totally orders $E(\Omega)$ i.e., for any two experiments X, Y either

$$V(f \mid X) \leqq V(f \mid Y) \qquad \text{for every } f \in F^+$$

or

$$V(f \mid X) \leqq V(f \mid X) \qquad \text{for every } f \in F^+ .$$

This assumption guarantees that if for some f_0 we have $V(f_0 \mid X) < V(f_0 \mid Y)$ then $(X, Y) \in \Pi(V)$.

LEMMA 3.1. *If $\Pi(V)$ is a total ordering of $E(\Omega)$ and if the experiments X, Y are such that for some $f_0 \in F^+$*

$$V(f_0 \mid X) = V(f_0 \mid Y) \neq 0 ,$$

then $V(f \mid X) = V(f \mid Y)$ for all $f \in F^+$.

PROOF. Let $T = (1, 2)$, $X_1 = X$ and let X_2 be completely informative. λX_T is the randomized experiment which chooses X_1 with probability λ and X_2 with probability $1 - \lambda$. Then

$$V(f \mid \lambda X_T) = \lambda V(f \mid X_1) + (1 - \lambda) V(f \mid X_2) = \lambda V(f \mid X) , \quad f \in F^+$$

since X_2 is assumed completely informative and $X_1 = X$. Under the assumption that $V(f_0 | X) = V(f_0 | Y) > 0$ we get for every λ

$$V(f_0 | \lambda X_T) = \lambda V(f_0 | X) < V(f_0 | Y) \Rightarrow (\lambda X_T, Y) \in \Pi(V)$$

since $\Pi(V)$ is a total ordering.

Suppose, now, that for some f_1, $V(f_1 | X) > V(f_1 | Y)$. Then for λ sufficiently close to 1

$$V(f_1 | \lambda X_T) > V(f_1 | Y)$$

which contradicts $(\lambda X_T, Y) \in \Pi(V)$ for every λ. Hence for every $f \in F^+$, $V(f | X) \leqq V(f | Y)$ so that $(X, Y) \in \Pi(V)$. Repeating the argument with Y instead of X we get $(Y, X) \in \Pi(V)$ which, together with the result $(X, Y) \in \Pi(V)$, implies $V(f | X) = V(f | Y)$ for all $f \in F^+$. When this relation holds between experiments, for a given V, we shall say that the experiments are equivalent with respect to V and write $X \equiv Y(V)$.

REMARK. Lemma 3.1 implies that for any pair of experiments X, Y we can have just three possibilities:

(i) $X \equiv Y(V)$
(ii) $V(f | X) < V(f | Y)$ all $f \in F^+ \ni V(f | Y) \neq 0$
(iii) $V(f | X) > V(f | Y)$ all $f \in F^+ \ni V(f | X) \neq 0$.

LEMMA 3.2. *Let $\Pi(V)$ be a total ordering of $E(\Omega)$ and let X, Y be a pair of experiments. Let $T = (1, 2)$, $X_1 = X$ and let X_2 be completely informative. If for some $f_0 \in F^+$, $V(f_0 | X) > V(f_0 | Y) > 0$ then there exists a $\lambda \in (0, 1)$ such that $\lambda X_T \equiv Y(V)$.*

PROOF. $V(f_0 | X) > V(f_0 | Y) > 0$ implies that there exists a $\lambda \in (0, 1)$ such that $\lambda V(f_0 | X) = V(f_0 | Y) > 0$ and, since $\lambda V(f_0 | X) = V(f_0 | \lambda X_T)$, we have, by Lemma 3.1, $\lambda X_T \equiv Y(V)$.

REMARK. Lemma 3.2 shows that if X, Y are any two experiments, then $V(f | X) = cV(f | Y)$:

If $X \equiv Y(V)$, take $c = 1$

If $(Y, X) \in \Pi(V)$ and if for some f_0': $V(f_0 | X) > V(f_0 | Y) > 0$ then, by Lemma 3.2, there is a $\lambda > 0$ such that $\lambda X_T \equiv Y(V)$ so that $\lambda V(f | X) = V(f | Y)$ for every $f \in F^+$ and we can take $c = 1/\lambda$.

If $(X, Y) \in \Pi(V)$ we use the same argument to produce λ such that $V(f | X) = \lambda V(f | Y)$ and take $c = \lambda$.

In particular, if Y is any uninformative experiment, we get for any experiment, X

$$V(f | X) = c_X V(f | Y) = c_X V(f)$$

where c_X depends on X but not on $f \in F^+$. If X is completely informative, $c_X = 0$. If X is uninformative, $c_X = 1$. Generally we have $0 \leqq c_X \leqq 1$. If we assume, now, that V is normalized so that $V(1^*) = 1$ where 1^* is the N-vector of all 1's, we get $c_X = V(1^* \mid X)$ and

$$V(f \mid X) = V(1^* \mid X) V(f)$$

for every $f \in F^+$.

LEMMA 3.3. *Let* $\Pi(V)$ *be a total ordering of* $E(\Omega)$. *Then, for any pair of functions* f, $g \in F^+$ *we have* $V(fg) = V(f)V(g)$.

PROOF. Let $f \leqq 1$. Consider the following experiment, X:

$$\mathscr{X} = (0, 1, 2, \cdots, N),\ p_0(\theta) = f(\theta),\ p_i(0) = 0 \quad \text{if} \quad \theta \neq \theta_i \quad \text{and}$$
$$= 1 - f(\theta_i) \quad \text{if} \quad \theta = \theta_i\,.$$

X can also be described as follows:

$$X = 0 \qquad \text{with probability} \quad f$$
$$= X^* \qquad \text{with probability} \quad 1 - f$$

where X^* is a completely informative experiment. Then, clearly,

$$V(1^* \mid X) = \textstyle\sum_i V(p_i) = V(f)$$

so that from the remark following Lemma 3.2

$$V(g \mid X) = V(f)V(g)$$

for every $g \in F^*$. On the other hand

$$V(g \mid X) = \textstyle\sum_i V(gp_i) = V(gf)$$

which proves the assertion for all $f \leqq 1$. Since V is nonnegative homogeneous the result can be extended to any $f \in F^+$.

REMARK. The result of Lemma 3.3 can obviously be extended to read: $V(\prod_{i=1}^k f_i) = \prod_{i=1}^k V(f_i)$ whence we easily get $V(f^m) = [V(f)]^m$ for any integer, m. Since:

$$[V(f^{m/n})]^n = [V(f)]^m$$

we get $V(f^r) = [V(f)]^r$ for any rational r, and finally, by continuity of V, $V(f^t) = [V(f)]^t$ for any real number, t.

We are now ready to prove the main theorem of this section which will be shown to be a consequence of Lemma 3.3.

THEOREM 3.1. *The ordering induced by a normalized* V, *satisfying the conditions* (2.1) *is total if and only if* V *is of the form*:

$$V(f) = \textstyle\prod_{i=1}^N [f(\theta_i)]^{t_i} = \exp[\sum_i t_i \log f(\theta_i)] \tag{3.1}$$

for some fixed $t = (t_1, \cdots, t_N)$ *where* $t_i \geqq 0$ *and* $\sum_i t_i = 1$ (*i.e.*, V *is a geometric mean*).

PROOF. The "if" part is only a matter of checking that V's of this form do, in fact, satisfy (2.1) and generate a total ordering of the experiments. The properties (2.1) are well-known properties of the geometric mean (see e.g. [6]). Now, let X have likelihood functions p_x with respect to μ and Y have likelihood functions q_y with respect to η. Then $(X, Y) \in \Pi(V) \Rightarrow$

$$\begin{aligned}\int_{\mathscr{X}} V(fp_x)\, d\mu(x) &= \int_{\mathscr{X}} \prod_i [f(\theta_i)p_x(\theta_i)]^{t_i}\, d\mu(x) \\ &= \prod_i [f(\theta_i)]^{t_i} \int_{\mathscr{X}} \prod_{i=1}^N [p_x(\theta_i)]^{t_i}\, d\mu(x) \\ &\leqq \prod_i [f(\theta_i)]^{t_i} \int_{\mathscr{Y}} \prod_i [q_y(\theta)]^{t_i}\, d\eta(y) = \int_{\mathscr{Y}} V(fq_y)\, d\eta(y)\end{aligned}$$

for every $f \in F^+$ if and only if

$$\int_{\mathscr{X}} \prod_i [p_x(\theta_i)]^{t_i}\, d\mu(x) \leqq \int_{\mathscr{Y}} \prod_i [q_y(\theta_i)]^{t_i}\, d\eta(y)\,. \tag{3.2}$$

Since the ordering is determined solely by the magnitude of a single number associated with each experiment (in fact $V(1^* \mid X)$) the ordering is total.

To prove the converse consider the linear space $H = \{h : e^h \in F^+\}$ (H is Euclidean N-space) and let $W(h) = \log V(e^h)$.

According to Lemma 3.3 and the remark following it:

$$V(f^a g^b) = [V(f)]^a [V(g)]^b$$

so that:

$$\begin{aligned}W(ah_1 + bh_2) &= \log V(e^{ah_1} \cdot e^{bh_2}) \\ &= a \log V(e^{h_1}) + b \log V(e^{h_2}) \\ &= aW(h_1) + bW(h_2)\,.\end{aligned}$$

Hence W is a linear functional on H. Furthermore $h_1 \geqq h_2 \Rightarrow e^{h_1} \geqq e^{h_2} \Rightarrow V(e^{h_1}) \geqq V(e^{h_2})$ since V is superadditive. Hence $h_1 \geqq h_2 \Rightarrow W(h_1) \geqq W(h_2)$. Thus W is a positive linear functional on $\mathscr{H}$ and the continuity properties of V guarantee that W is continuous and bounded. V normalized $\Rightarrow W(1^*) = \log V(e^*) = \log e = 1$. Hence by the representation theorem for continuous linear functions on R_N we have $W(h) = \sum_i t_i h(\theta_i)$ for some fixed $t = (t_1, \cdots, t_N)$ with $t_i \geqq 0$ and $\sum_i = t_i = 1$. Since $h(\theta_i) = \log f(\theta_i)$ for some $f \in F^+$ we get

$$V(f) = \exp[\sum t_i \log f(\theta_i)] = \prod_{i=1}^N [f(\theta_i)]^{t_i}$$

which was to be proved.

The function $L_X(t) = \int_{\mathscr{X}} \prod_i [p_x(\theta_i)]^{t_i}\, d\mu(x)$ defined for every t in the N-dimensional simplex: $t_i \geqq 0$, $\sum t_i = 1$ is called the Laplace transform of the experiment, X. If we denote by V_t the geometric mean of order t on F^+, defined in (3.1), from (3.2) we see that $(X, Y) \in \Pi(V_t) \Leftrightarrow L_X(t) \leqq L_Y(t)$.

Since the function $L_X(t)$ uniquely represents the experiment, X, any property of experiments, in particular ordering properties, should be expressible in terms of their Laplace transforms. Efforts at establishing such expressions have thus far been unsuccessful. Note, for example that if X is sufficient for Y then $L_X(t) \leqq L_Y(t)$ for every t in the simplex. The converse, however, is not true, as shown by Torgersen in [10].

For $N = 2$, $L_X(t)$ was studied by Chernoff [4] in which he uses the value

$\rho = \inf_t \log L_X(t)$ to order experiments. The preceding theorem shows that this is not a Bayesian ordering, as we have used the term.

In Section 2, Example 2 we discussed the special case $U(\xi) = (\xi(1-\xi))^{\frac{1}{2}}$, which corresponds to $V_{\frac{1}{2}}$ and $N = 2$. Let $U_t(\xi) = \xi^t(1-\xi)^{1-t}$. Then a family of decision problems which yield this family of uncertainty functions is given by: $\Omega = (0, 1)$, $A = [0, 1]$, the unit interval, and

$$L_t(\theta, a) = t\left(\frac{1-a}{a}\right)^{1-t} \qquad \text{for } \theta = 0$$

$$= (1-t)\left(\frac{a}{1-a}\right)^{t} \qquad \text{for } \theta = 1 .$$

For $t = \frac{1}{2}$ the decision problem has the following simple structure: $\Omega = (0, 1)$, $A = \{\text{nonnegative reals}\}$ and

$$L(0, a) = a$$

$$L(1, a) - a^{-1}$$

so that if $\theta = 0$ we want to choose a small, if $0 = 1$ we want to choose a large and the Bayes envelope of this problem is $U(\xi) = 2(\xi(1-\xi))^{\frac{1}{2}}$, which yields the ordering of Example 2, Section 2.

4. Order preservation. It was shown in [5] that if $(X, Y) \in \Pi(U)$ then $(X, Y) \in \Pi(U_Z)$ where $U_Z(\xi) = U(\xi \mid Z)$ and Z is any experiment. Correspondingly we have $(X, Y) \in \Pi(V) \Rightarrow (X, Y) \in \Pi(V_Z)$. We show here that the transformations $V \to V_Z$, and their limits are the only transformations for which the implication holds. The argument depends on the cone structure of the orderings, a separation theorem for convex sets, and the representation theorem for linear functionals.

Let C be a cone of continuous functions on F^+ i.e., $W \in C \Rightarrow cW \in C$ for every $c > 0$ and $W_1 W_2 \in C \Rightarrow W_1 + W_2 \in C$. We suppose also that C contains the restrictions to F^+ of all linear functionals on R_N. Let $\bar{C}$ be the closure of C (in the topology of uniform convergence on compact sets) and let M be the class of regular Borel signed measures with compact support on F^+. Call $m \in M$ *C-positive* if $\int W(f)\, dm(f) \geqq 0$ for all $W \in C$. The cone C thus determines an ordering of the positive measures in M: $(\mu_1, \mu_2) \in \Pi(C)$ if $\mu_1 - \mu_2$ is C-positive.

If $W_0 \notin \bar{C}$, there is, by the separation theorem for convex sets, a linear functional (in this case an element of M) $m_0 \in M$ such that $\int W_0\, dm < 0$ and $\int W\, dm_0 \geqq 0$ for every $W \in \bar{C}$.

Note that m C-positive $\Rightarrow \int l(f)\, dm(f) = 0$ for every linear functional, l. Therefore, the $m_0 = m_0^+ - m_0^-$ determined above can be assumed to be the difference of two bounded measures such that $\int l(f)\, dm_0^+(f) = \int l(f)\, dm_0^-(f)$ for every linear functional, l.

By suitable norming we can further assume $\int l(f)\, dm_0^+(f) = \int l(f)\, dm_0^-(f) = l(1^*)$ where, as before, 1^* is the vector of all 1's.

Now consider a pair of experiments X, Y. We note first that every experiment with finite parameter space has a "standard representation" (see e.g. [2] and [10]) as a measure, μ, on the probability simplex, S, in N-space such that (a) $\int_S d\mu(s) = N$ and (b) $\int_S s_i\, d\mu(s) = 1$, $i = 1, 2, \cdots, N$. The pair X, Y thus induces a signed measure m on F^+, with compact support and with the properties

(i) $\int l(f)\, dm(f) = 0$

(ii) $\int l(f)\, dm^+(f) = \int l(f)\, dm^-(f) = l(1^*)$

for every linear functional, l.

Conversely, any measure $m \in M$ with properties (4.1) is determined by some pair of experiments X, Y.

Let V be a function with properties (2.1) and consider the closed cone $\bar{C}_0(V)$ determined by the class of functions $\tilde{V} = \{V_Z, Z \in E(\Omega)\}$ and denote by $\bar{C}(V)$ the closed cone determined by $\tilde{V}$ and all the linear functionals on F^+. We now prove the following:

THEOREM 4.1. *Let V, W be two functions with properties* (2.1). *The implication* $(X, Y) \in \Pi(V) \Rightarrow (X, Y) \in \Pi(W)$ *is valid for all* $X, Y \in E(\Omega)$ (*i.e.*, $\Pi(V) \subset \Pi(W)) \Leftrightarrow$ $W \in \bar{C}_0(V)$.

PROOF. As mentioned above the "if" part was essentially shown in [5]. The arguments preceding the statement of the theorem show that if $W \notin \bar{C}(V)$ then there exists $m_0 \in M$ such that:

(i) $V_0(f)\, dm_0(f) \geqq 0$ for every $V_0 \in \bar{C}(V)$ and

(ii) $W(f)\, dm_0(f) < 0$.

Let X_0, Y_0 be a pair of experiments which induce m_0^- and m_0^+ respectively on F^+. Then (i) $\Rightarrow (X_0, Y_0) \in \Pi(V)$ and (ii) $\Rightarrow (X_0, Y_0) \notin \Pi(W)$. It follows therefore, by contradiction that $(X, Y) \in \Pi(V) \Rightarrow (X, Y) \in \Pi(W)$ for every pair X, Y requires that $W \in \bar{C}(V)$. Since W is assumed to satisfy condition (2.1) (iv) (i.e., $W(f) = 0$ for all f with only one nonzero component) it can not have a positive linear term and therefore must be in the smaller cone $\bar{C}_0(V)$ as was to be shown.

We now consider the question: when is it true that for two functions V, W satisfying (2.1) the double implication $(X, Y) \in \Pi(V) \Leftrightarrow (X, Y) \in \Pi(W)$ is valid for every pair $X, Y \in E(\Omega)$? Equivalently—when is it true that $\Pi(V) = \Pi(W)$?

The previous theorem makes it clear that the answer to this question amounts to finding conditions under which the cones $\bar{C}_0(V)$ and $\bar{C}_0(W)$ are identical.

Let M_0 denote the class of measures on the simplex which represent experiments and let $V_\mu(f) = \int_S V(fg)\, d\mu(g)$, $\mu \in M_0$. Then $\tilde{V} = \{V_Z, Z \in E(\Omega)\} = \{V_\mu, \mu \in M_0\}$, a convex set. $\lambda\tilde{V}$ is the class of functions obtained by multiplying each member of $\tilde{V}$ by the real number λ and the cone generated by V is $C_0(V) = \bigcup_{\lambda \geqq 0} \lambda\tilde{V}$.

Note that if $W \in \lambda\tilde{V}$ then so is αW where $0 \leqq \alpha \leqq 1$ since αW can be obtained by an experiment which randomizes between a completely informative and an informative experiment. To avoid ambiguity of representation we associate with each $\mu \in M_0$ the following number:

$$\lambda(\mu) = \inf\{\lambda:\ V_\mu \in \lambda \tilde{V}\}\,.$$

Clearly $0 \leqq \lambda(\mu) \leqq 1$ and if $\lambda(\mu) = \lambda$ then there is a $\mu^* \in M_0$ such that $\lambda(\mu^*) = 1$ and $V_\mu = \lambda V_{\mu^*}$. The measure μ^* will be called maximal wherever $\lambda(\mu^*) = 1$.

The use of maximal measures will enable us to distinguish between the elements of $\bar{C}_0(V)$ and those of $C_0(V)$.

LEMMA 4.1. *Let $W(f) = \lim_n \lambda_n V_{\mu_n}(f)$ uniformly for $f \in S$ with $\lambda_n \geqq 0$ and μ_n maximal. Then $W \in C_0(V)$ if and only if $\{\lambda_n\}$ is a bounded sequence.*

PROOF. Suppose $W \in C_0(V)$. Then for some $a > 0$ and some $\mu \in M_0$ $W = aV_\mu$ where μ is maximal. Now if λ_n is unbounded we can assume for n_0 sufficiently large and $n \geqq n_0$ $\lambda_n > a + 1$ so that $\lambda_n V_{\mu_n} \in \bigcup_{\lambda > a+1} \lambda \tilde{V}$ for all $n \geqq n_0$ and $\lambda_n V_{\mu_n} \notin \lambda \tilde{V}$ for any $n \geqq n_0$ and $\lambda \leqq a$. This contradicts $W = aV_\mu$ and therefore λ_n must be bounded. Conversely, if $\{\lambda_n\}$ is bounded we can assume $\lambda_n \to a \geqq 0$ and since M_0 is weakly compact there exists $\mu \in M_0$ such that $\int_S V(fg)\, d\mu_n(g) \to \int_S V(fg)\, d\mu$. Hence $\lim_n \lambda_n V_{\mu_n} = aV_\mu$ and $W \in C_0(V)$ as was to be shown.

LEMMA 4.2. *Let μ_1, μ_2 be maximal measures on S, μ the measure induced on S by the pair of experiments (μ_1, μ_2). Then $\lambda(\mu) = \inf\{\lambda:\ V_{\mu_1} \in \lambda \tilde{V}(\mu_2)\} = \inf\{\lambda:\ V_{\mu_2} \in \tilde{V}(\mu_1)\}$ where*

$$\tilde{V}(\mu_i) = \{V_{\mu_{i,\eta}},\ \eta \in M_0)\} = \{\textstyle\int_S V_{\mu_i}(fg)\, d\eta(g),\ \eta \in M_0\}\,.$$

PROOF. The proof follows immediately from the definition of maximality and the fact that $V_{\mu_{i,\eta}} \in \tilde{V}_{\mu_i}$ for every $\eta \in M_0$.
The following lemma will enable us to characterize those functions W which can not give rise to the same ordering as V.

LEMMA 4.3. *Let $W \in \bar{C}_0(V) - C_0(V)$. Then the only function that $\bar{C}_0(W)$ and $C_0(V)$ have in common is the zero function.*

PROOF. $W \in \bar{C}_0(V) - C_0(V)$ implies $W = \lim_n \lambda_n V_{\mu_n}$ with μ_n maximal and $\lambda_n \to \infty$. If $\eta \in M_0$ then $W_\eta(f) = \lim_n \lambda_n \int_S V_{\mu_n}(fg)\, d\eta(g)$ since the convergence is uniform on compact sets. By Lemma 4.2 the maximal representation for W_n is given by

$$W_\eta(f) = \lim_n \alpha_n \lambda_n \textstyle\int_S V_{\eta_n}(fg)\, d\mu_n(g)\,.$$

Now $\int_S V_{\eta_n}(fg)\, d\mu_n(g) \leqq \int_S V(fg)\, d\mu_n(g) \to 0$ uniformly for $f \in S$. Hence $\alpha_n \lambda_n$ is either unbounded or $W_\eta(f) \equiv 0$.

Suppose $W_0(f) = \lim_n a_n \int_S W(gf)\, d\eta_n(g)$ for a maximal sequence $\{\eta_n\}$. $W(gf) = \lim_m \lambda_m \int_S V(gf)\, d\mu_m(g)$ so that by appropriate choice of a subsequence of n's:

$$W_0(f) = \lim_n a_n \lambda_{m(n)} \textstyle\int_S \int_S V(fgh)\, d\eta_n(g)\, d\mu_{m(n)}\,.$$

Again, since the double integral converges uniformly to zero we must have $a_n \lambda_{m(n)}$ unbounded or $W_0(f) \equiv 0$. Hence either $W_0(f)$ is identically 0 or $W_0(f) \in \bar{C}_0(V) - C(V)$.

We shall now drop the standard measure notation and revert to the notation developed earlier. We may now prove:

THEOREM 4.2. *Let V, W be two functions with properties* (2.1). *The double implication* $(X, Y) \in \Pi(V) \Leftrightarrow (X, Y) \in \Pi(W)$ *is valid for all* $X, Y \in E(\Omega) \Leftrightarrow W(f) = V(gf)$ *for some g such that* $g(\theta_i) > 0$ *for all* $i = 1, 2, \cdots, N$.

PROOF. First note that if W is of the stated form W and V do generate identical orderings since the cones $\bar{C}_0(V)$ and $\bar{C}_0(W)$ are identical.

By Lemma 4.3 the double implication means that $W = aV_Z$ and $V = bW_Y$ for some $a, b > 0$ and some $Y, Z \in E(\Omega)$. Hence we must have $W = abW_{(Y,Z)}$ where (Y, Z) is the joint experiment consisting of an observation on Y and an independent observation on Z.

Now, W is an extreme point of the convex set $\tilde{W} = \{W_Z, Z \in E(\Omega)\}$ as can be seen from:

$$W = \alpha W_{Z_1} + (1 - \alpha)W_{Z_2} = W_{\alpha Z_1 + (1-\alpha)Z_2}$$

implies $\alpha Z_1 + (1 - \alpha)Z_2$ is uninformative $\Rightarrow Z_i$ is uninformative for $i = 1, 2$ and hence $W_{Z_i} = W$.

Since $W_{(Y,Z)} = 1/(ab)W$, $W_{(Y,Z)}$ must be a function in the extreme ray cW, $c \geqq 0$ and can not be a convex combination of functions in other rays of the cone generated by W. That is to say if

$$W_{(Y,Z)}(f) = \int W(gq_y r_z)\, d\eta(y)\, d\nu(z)$$

then $W(fq_y r_z) = c(y, z)W(f)$ for every y and z and every $f \in F^+$. Hence $q_y r_z = c^*(y, z)$ for every pair y, z where $c^*(y, z)$ is a vector each of whose coordinates is the constant $c(y, z)$. On the set $c(y, z) > 0$ we get, then

$$q_y = c(y, z)r_z^{-1}$$

and by varying z for fixed y and then y for fixed z we see that the likelihood vectors r_z and q_y must be concentrated on rays cg, cg^{-1}, $g > 0$ respectively, except for those y's and z's for which q_y or r_z are roots of W. In particular, Z is equivalent to an experiment of the type used in Lemma 3.3, namely:

$$\begin{aligned} Z &= 0 \quad \text{with probability} \quad g > 0 \\ &= X^* \quad \text{with probability} \quad 1 - g \end{aligned}$$

where X^* is a completely informative experiment. Since $W = V_Z$ we get $W(f) = V(gf) = V_g(v)$ for some $g > 0$ and every $F \in F^+$.

REMARK. We noted in Section 2 that for $n = 2$ $(X, Y) \in \Pi(V)$ for every V is equivalent to $(X, Y) \in \Pi(V_0)$ where $V_0(a, b) = \min(a, b)$. This is due to the fact that the extreme rays of the convex cone consisting of all continuous nonnegative superadditive functions on the positive quadrant which vanish on the axes are related to one another by transformation of the type described in the above theorem. More precisely if V_1, V_2 are any two functions in different extreme rays of the cone, there exist positive numbers c_1, c_2 such that

$$V(a, b) = V_1(c_1 a, c_2 b) \qquad \text{for all } a, b .$$

This phenomenon does not occur in higher dimensions. For example if $N = 3$ the functions $V_1(a, b, c) = \min(a, b) + \min(b, c) + \min(a, c)$ and $V_2(a, b, c) = \min(a + b, b + c, a + c)$ are in different extreme rays and can not be related as above.

The following corollary expresses the result in terms of the loss function of the decision problem:

COROLLARY 4.1. *The decision problems* (Ω, A, L_1) *and* (Ω, A, L_2) *induce the same ordering on the class* $E(\Omega)$ *if and only if there exists* g *such that* $g(\theta_i) > 0$ *for* $i = 1, 2, \cdots, N$ *such that*

$$L_2(\theta, a) = g(\theta)L_1(\theta, a) .$$

The proof follows from Theorem 4.1 and the representation (2.2).

Note also that the fact that the action spaces are then to be the same for both decision problems is not a restriction. First, because any two action spaces with a 1-1 correspondence between them are essentially identical and secondly because it is always possible to amplify an action space by adding inadmissible or redundant acts.

REFERENCES

[1] BLACKWELL, D. (1951). Comparison of experiments, *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[2] BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

[3] BRADT, R. N. and KARLIN, S. (1956). On the design and comparison of certain dichotomous experiments. *Ann. Math. Statist.* **27** 390–409.

[4] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations. *Ann. Math. Statist.* **23** 493–507.

[5] DE GROOT, M. H. (1962). Uncertainty, information and sequential experiments. *Ann. Math. Statist.* **33** 404–419.

[6] HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press.

[7] LECAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.

[8] LINDLEY, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Statist.* **27** 986–1005.

[9] SAKAGUCHI, M. (1964–1965). Topics in information and decision processes. Technical Report, The George Washington Univ.

[10] TORGERSEN, E. N. (1968). Comparison of experiments when the parameter space is finite. Ph. D. dissertation. Univ. of California, Berkeley.

DEPT. OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48823