SOME PROPERTIES OF BAYESIAN ORDERINGS OF EXPERIMENTS

BY

D. FELDMAN

TECHNICAL REPORT NO. 165
NOVEMBER 9, 1970

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
N00014-67-A-0112-0053 (NR-042-267)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SOME PROPERTIES OF BAYESIAN ORDERINGS OF EXPERIMENTS

by

D. FELDMAN

TECHNICAL REPORT NO. 165
NOVEMBER 9, 1970

PREPARED UNDER CONTRACT N00014-67-A-0112-0053
NR-042-267
OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
1. **Introduction and Summary.**

Consider a decision problem, \((\Omega, A, L)\), where \(\Omega\) 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 \(\Omega\) 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 \(\xi\) over \(\Omega\), 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 \( \xi \), 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 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 \( \Xi \) be the class of all priors over \( \Omega \). 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 any concave functional on \( \Xi \) 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 \( \Xi \).
Let $X = (\mathcal{X}, \mathcal{A}; P_\theta, \theta \in \Omega)$ denote an experiment with measurable sample space $(\mathcal{X}, \mathcal{A})$ and a family of probability models $P_\theta, \theta \in \Omega$, and let $Y = (\mathcal{Y}, \mathcal{B}; Q_\theta, \theta \in \Omega)$ denote an alternative experiment. We say that $X$ is at least as (more) informative as (than) $Y$ with respect to $U$ if $U(\xi | X) \leq U(\xi | Y)$ for all $\xi \in \Xi$ (and $U(\xi | X) < U(\xi | Y)$ for some $\xi \in \Xi$) where $U(\xi | X)$ is the expected posterior uncertainty if $X$ is observed and the prior distribution is $\xi$ and $U(\xi | Y)$ 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) $\Omega$ is a finite set with $N$ elements. $\Xi$ is then the $N$-dimensional simplex consisting of all points $\xi = (\xi(\theta_1), \ldots, \xi(\theta_N))$ with $\xi(\theta_1) \geq 0$ and $\sum \xi(\theta_1) = 1$.

(ii) $U$ is non-negative, continuous and concave on $\Xi$ and $U(\xi) = 0$ at all degenerate probability vectors $\xi = (0,0,\ldots,0,1,0,\ldots,0)$. 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 $X = (\mathcal{X}, \mathcal{A}; P_\theta, \theta \in \Omega)$ such that $P_\theta, \theta \in \Omega$ is a dominated family with densities $p_x(\theta)$ at $x$. 

3
and with the property that the mapping is

\[ x \rightarrow p_x = (p_x(\theta_1), p_x(\theta_2), \ldots, p_x(\theta_N)) \in \mathbb{R}^N \]

is Borel.

Under these assumptions, we show in section 3 that the only functions \( \mathcal{U} \) which induce a total ordering in \( E(\mathcal{O}) \) are geometric means i.e., functions of the form \( \mathcal{U}(\xi) = \prod_{i=1}^{N} [\xi(\theta_i)]^{t_i} \) for some fixed \( t = (t_1, \ldots, t_N) \) with \( t_i \geq 0 \) 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 \( \Pi(U_1) = \Pi(U_2) \iff \) the loss function corresponding to \( U_2 \) differs from that of \( U_1 \) by a multiplicative function of \( \theta \) which is nowhere 0.
2. Preliminaries.

Let $\Omega = (\theta_1, \theta_2, \ldots, \theta_N)$ be a finite set, and $\Xi$ the class of all
prior distributions on $\Omega$. If $\xi \in \Xi$ then $\xi = (\xi(\theta_1), \xi(\theta_2), \ldots, \xi(\theta_N))$
with $\xi(\theta_i) \geq 0$ and $\sum_i \xi(\theta_i) = 1$. The uncertainty function $U$ is
taken to be non-negative, continuous, and concave on $\Xi$ with the
additional property that $U(\xi) = 0$ for every degenerate probability
function $\xi$. Without loss of generality we then assume that any experi-
ment $X = (\mathcal{X}, \mathcal{A}, \mathcal{P}, \Theta, \Omega)$ is dominated and we denote the density of $\mathcal{P}_\theta$
at $x$, with respect to the dominating measure, by $p_x(\theta)$ - the likelihood
function at $x$ is then denoted by $p_x$. If $\xi$ 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 $\xi p_x$ is the ordinary product of the functions $\xi$ and $p_x$.

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

$$
U(\xi | X) = \int_X U(\xi_x)p_x(\xi)d\mu(x) \leq \int_Y U(\xi_y)q_y(\xi)d\eta(y) = U(\xi | Y)
$$

for every $\xi \in \Xi$.

If $(X,Y) \in \Pi(U)$ and $(Y,X) \not\in \Pi(U)$ 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 $\xi_x$ directly we will use a convenient extension of $U$ defined as follows: Let $F^+(\Omega) = F^+$ denote the class of all non-negative functions, $f$, on $\Omega$ ($F^+$ is the orthant of non-negative vectors in $N$-space). For every $f \in F^+$ define

$$V(f) = U \left( \frac{f}{\sum_i f(\theta_i)} \right) \sum_i f(\theta_i)$$

and $V(f|X) = \int_{\mathbb{R}} V(fp_x) d\mu(x)$.

The properties of $U$ on $\Xi$ guarantee the following for $V$ on $F^+$:

(i) continuity

(ii) superadditivity: $V(f+g) \geq V(f)+V(g)$, $f, g \in F^+$

(iii) non-negative homogeneity: $V(cf) = cV(f)$, $c > 0$, $f \in F^+$

(iv) $V(f) = 0$ for all $f$ with only one non-zero component

(v) For any experiment $X$, $V(f|X) \leq V(f)$ for every $f \in F^+$.

A 1-1 correspondence between functions $U$ on $\Xi$ 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 $c f \in F_0$. In particular we may take

$$F_0 = \{ f : \sum_i f(\theta_i) = 1 \},$$

which yields $U$. 

6
If $U$ is the uncertainty function for the decision problem $(\Omega, A, L)$ 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:

\begin{equation}
V(f) = \inf_a \sum_i L(\theta_i, a)f(\theta_i) \text{ for all } f \in F^+.
\end{equation}

In this context it is useful to think of $f \in F^+$ as a prior likelihood function and $f_{px}$, then, as the posterior likelihood function if $x$ is the observed value of $X = (x, a, P_{\theta}, \theta, \Omega)$.

An experiment, $X$, is called **completely informative** with respect to $V$ if $V(f|X) = 0$, for all $f \in F^+$. At the other extreme, any $X$ for which $V(f|X) = V(f)$ for every $f \in F^+$ is called **uninformative**. An unininformative experiment is an auxiliary randomization or, alternatively, an experiment whose likelihood functions are concentrated on the ray: $\{f: f(\theta_i) = c, \text{ for some } c > 0\}$. 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 $(X_t, t \in T)$. $\lambda X_T$ denotes the randomized experiment obtained by choosing $t$ according to the probability measure $\lambda$ over $T$ and observing $X_t$. Then

\begin{equation}
V(f|\lambda X_T) = \int_T V(f|X_t) d\lambda(t).
\end{equation}
If \( X \) is sufficient for \( Y \) then \( (X,Y) \in \Pi(V) \) for every \( V \in [5] \) and, in fact, the converse is also true (e.g. [2]). For \( N=2 \) there is a \( V_o \) such that \( (X,Y) \in \Pi(V_o) \iff (X,Y) \in \Pi(V) \) for every \( V \) satisfying the conditions (2.1). Using coordinate notation, the \( V_o \) that works is \( V_o(a,b) = \min(a,b) \) as is shown in [1] and [3] (the corresponding \( U_o \) is \( U_o(\xi) = \min(\xi,1-\xi) \)). The decision problem corresponding to this \( V_o \) (or \( U_o \)) 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 \( V_o \) 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:

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

\( U(\xi|X) \) is shown in [9] to be related to the Fisher information in the family \( p_X(\xi), \xi \in [0,1] \), treating \( \xi \) 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 \( (X,Y) \in \Pi(V) \iff \) for every \( a,b \geq 0; \)
\[
\frac{p_1 p_2}{ap_1 + bp_2} + \frac{q_1 q_2}{aq_1 + bq_2} \leq \frac{r_1 r_2}{ar_1 + br_2} + \frac{s_1 s_2}{as_1 + bs_2}
\]

where \( q_1 = 1 - p_1 \) and \( s_1 = 1 - r_1 \). After algebraic reduction the ordering can be seen to be:

\[
(X,Y) \in \Pi(U) \iff \frac{(p_1 - p_2)^2}{p_1 (1 - p_1)} \geq \frac{(r_1 - r_2)^2}{r_1 (1 - r_1)} \quad \text{and} \quad \frac{(p_1 - p_2)^2}{p_2 (1 - p_2)} \geq \frac{(r_1 - r_2)^2}{r_2 (1 - r_2)}.
\]

Note that \( \frac{(p_1 - p_2)^2}{p_1 (1 - p_1)} \) 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.** \( U(\xi) = [\xi(1 - \xi)]^{1/2} \). This \( U \) is the standard deviation of the prior distribution and the ordering it induces is vastly different from that of example 1.

\[
U(\xi | X) = [\xi(1 - \xi)]^{1/2} \int [p_1(x)p_2(x)]^{1/2} d\mu(x).
\]

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

**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

\[
V(f|X) \leq V(f|Y) \quad \text{for every } f \in F^+, \\
\text{or } V(f|Y) \leq V(f|X) \quad \text{for every } f \in F^+.
\]

This assumption guarantees that \( V(f_o|X) < V(f_o|Y) \) for some \( f_o \in F^+ \)

\( \Rightarrow (X, Y) \in \Pi(V) \).

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

\[
V(f_o|X) = V(f_o|Y) \neq 0,
\]

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

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

\[
V(f|\lambda X_T) = \lambda V(f|X_1) + (1-\lambda)V(f|X_2) = \lambda V(f|X), \quad f \in F^+
\]
since \( X_2 \) is assumed completely informative and \( X_1 - X \). Under the assumption that \( V(f_o|X) = V(f_o|Y) > 0 \) we get for every \( \lambda \)

\[
V(f_o|\lambda X_T) = \lambda V(f_o|X) < V(f_o|Y) \Rightarrow (\lambda X_T, Y) \in \Pi(V)
\]
since \( \Pi(V) \) is a total ordering.
Suppose, now, that for some \( \tilde{f} \), \( V(\tilde{f} | X) > V(\tilde{f} | Y) \). Then for \( \lambda \) sufficiently close to 1

\[ V(f_{\lambda X_T}) > V(f_{\lambda Y}) \]

which contradicts \((\lambda X_T, Y) \in \Pi(V)\) for every \( \lambda \). Hence for every \( f \in F^+ \),

\[ V(f | X) \leq V(f | Y) \]

so that \((X, Y) \in \Pi(V)\). Repeating the argument with \( Y \) instead of \( X \) we get \((Y, X) \in \Pi(V)\) and therefore \( X = Y \) i.e.

\[ V(f | X) = V(f | Y) \quad \text{all } f \in F^+. \]

**Remark.** Lemma 3.1 implies that for any pair of experiments \( X, Y \) we can have just three possibilities:

1. \( V(f | X) = V(f | Y) \) all \( f \in F^+ \)
2. \( V(f | X) < V(f | Y) \) all \( f \in F^+ \) \( \exists V(f | Y) \neq 0 \)
3. \( V(f | X) > V(f | Y) \) all \( f \in F^+ \) \( \exists V(f | X) \neq 0 \).

**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:

\[ \exists \lambda(0,1) \ni \lambda X_T = Y. \]

**Proof.** \( V(f_0 | X) > V(f_0 | Y) > 0 \) \( \implies \exists \lambda(0,1) \ni \lambda V(f_0 | X) = V(f_0 | Y) > 0 \)

and, since \( \lambda V(f_0 | X) = V(f_0 | \lambda X_T) \), we have, by lemma 3.1, \( \lambda X_T = Y \).

**Remark.** Lemma 3.2 shows that if \( X, Y \) are any two experiments then

\[ V(f | X) = cV(f | Y). \]

If \( X = Y \) take \( c = 1 \)

If \( (Y, X) \in \Pi(V) \) and \( \exists f_0 \ni V(f_0 | X) > V(f_0 | Y) > 0 \),

by lemma 3.2 \( \exists \lambda > 0 \ni \lambda X_T = Y \implies \lambda V(f | X) = V(f | Y) \) for every \( f \in F^+ \) =>

we can take \( c = \frac{1}{\lambda} \).
If \((X,Y)\in \Pi(V)\) we use the same argument to produce 
\[ \lambda \in V(f|X) = \lambda V(f|Y) \] and then \( c = \lambda \).

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 \(0 \leq c_X \leq 1\). 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 \(c_X = V(1^*|X)\) and

\[ V(f|X) = V(1^*|X) V(f) \]

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 \(f \leq 1\). Consider the following experiment, \(X\):

\[ X = (0,1,2,\ldots,N), \quad p_0(\theta) = f(\theta), \quad p_1(\theta) = 0 \text{ if } \theta \neq \theta_i \text{ and } \]

\[ = 1 - f(\theta_i) \text{ if } \theta = \theta_i \]

\(X\) can also be described as follows:

\[ X = 0 \text{ with probability } f \]

\[ = X^* \text{ with probability } 1 - f \]

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

\[ V(1^*|X) = \sum_i V(p_1) = V(f) \]

12
so that from the remark following Lemma 3.2

\[ V(g|X) = V(f) V(g) \]

for every \( g \in \mathcal{F}^+ \). On the other hand

\[ V(g|X) = \sum_i V(gp_i) = V(gf) \]

which proves the assertion for all \( f \leq 1 \). Since \( V \) is non-negative homogeneous the result can be extended to any \( f \in \mathcal{F}^+ \).

**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:

\[ \left[ V(f^n) \right] = F(V^m) = [V(f)]^m \]

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:

\[ V(f) = \prod_{i=1}^N f(\theta_i)^{t_i} = e^{\sum_{i=1}^N t_i \log f(\theta_i)} \]

for some fixed \( t = (t_1, \ldots, t_N) \) where \( t_i \geq 0 \) 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 $\mu$ and $Y$ have likelihood functions $q_y$ with respect to $\eta$. Then $(X,Y) \in \Pi(V) \Rightarrow$

$$
\int_x V(p_x) d\mu(x) = \int_x \prod_i [f(\theta_i)p_x(\theta_i)]^t_i d\mu(x)
$$

$$
= \prod_i [f(\theta_i)]^t_i \int_x \prod_{i=1}^N [p_x(\theta_i)]^t_i d\mu(x)
$$

$$
\leq \prod_i [f(\theta_i)]^t_i \int_y \prod_i [q_y(\theta_i)]^t_i d\eta(y) = \int_y V(q_y) d\eta(y)
$$

for every $f \in F^+$ $\Leftrightarrow$

(3.2) \[ \int_x \prod_i [p_x(\theta_i)]^t_i d\mu(x) \leq \int_y \prod_i [q_y(\theta_i)]^t_i d\eta(y). \]

Since the ordering is determined solely by the magnitude of a single number associated with each experiment (in fact $V(1^*|X)$) 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:
\[ W(ah_1 + bh_2) = \log V(e^{ah_1} \cdot e^{bh_2}) \]
\[ = a \log V(e^{h_1}) + b \log V(e^{h_2}) \]
\[ = aW(h_1) + bW(h_2). \]

Hence \( W \) is a linear functional on \( \mathcal{H} \). Furthermore \( h_1 \geq h_2 \Rightarrow h_1' \geq h_2' \Rightarrow V(e^{h_1'}) \geq V(e^{h_2'}) \) since \( V \) is super-additive. Hence \( h_1 \geq h_2 \Rightarrow W(h_1) \geq W(h_2) \). Thus \( W \) is a positive linear functional on \( \mathcal{H} \) 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 \( t = (t_1, \ldots, t_N) \) with \( t_i \geq 0 \) and \( \sum_{i} t_i = 1 \). Since \( h(\theta_i) = \log f(\theta_i) \) for some \( f \in F^+ \), we get
\[ V(f) = e^{\sum_{i} t_i \log f(\theta_i)} = \prod_{i=1}^{N} [f(\theta_i)]^{t_i} \]

which was to be proved.

The function \( L_X(t) = \int_{\mu} \prod_{i} [p_X(\theta_i)]^{t_i} d\mu(x) \) defined for every \( t \) in the \( N \)-dimensional simplex: \( t_i \geq 0, \sum t_i = 1 \) 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 \((X,Y) \in \Pi(V_t) \Leftrightarrow L_X(t) \leq L_Y(t) \).

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 \( L_X(t) \leq L_Y(t) \) for every \( t \) in the simplex. The converse, however is not true, as shown by Torgerssen in [10].

For \( N = 2 \) \( L_X(t) \) was studied by Chernov [4] in which he uses the value \( \rho = \inf \log L_X(t) \) 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 \( U(\xi) = \sqrt{\frac{1}{2}(1-\xi)} \), which corresponds to \( V_{1/2} \) and \( N = 2 \). Let \( U_t(\xi) = \frac{\xi}{(1-\xi)}(1-\xi)^{1-t} \). 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

\[
L_t(\theta,a) = t(1-a)^{1-t} \quad \text{for} \quad \theta = 0
\]

\[
= (1-t)(\frac{a}{1-a})^t \quad \text{for} \quad \theta = 1
\]

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

\[
L(0,a) = a
\]

\[
L(1,a) = \frac{1}{a}
\]

so that if \( \theta = 0 \) we want to choose a small, if \( \theta = 1 \) we want to choose a large and the Bayes envelope of this problem is \( U(\xi) = 2 \sqrt{\frac{1}{2}(1-\xi)} \), which yields the ordering of example 2, sec. 2.
4. **Order Preservation.**

It was shown in [5] that if \((X,Y) \in \Pi(U)\) then \((X,Y) \in \Pi(U_Z)\) where \(U_Z(\xi) = U(\xi | Z)\) 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 \(V \rightarrow V_Z\), 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., \(\forall c \in C\) 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 \(\overline{C}\) 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 \(\int W(f) \, dm(f) \geq 0\) 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 \(W \downarrow \overline{C}\), 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 \(\int W \, dm_0 < 0\) and \(\int W \, dm_0 \geq 0\) for every \(W \in \overline{C}\).

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

By suitable norming we can further assume \(\int \ell(f) \, dm_0^+(f) = \int \ell(f) \, dm_0^-(f) = \ell(1^*)\) 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, \( \mu \), on the probability simplex, \( S \), in \( N \)-space such that (a) \( \int_S \mu(s) = N \) and (b) \( \int_S s_i \mu(s) = 1, \)
i = 1, 2, \ldots, \( N \). The pair \( X, Y \) thus induces a signed measure \( m \) on \( F^+ \), with compact support and with the properties

\[
(i) \quad \int \ell(f) dm(f) = 0 \\
(ii) \quad \int \ell(f) dm^+(f) = \int \ell(f) dm^-(f) = \ell(1^x)
\]

for every linear functional, \( \ell \).

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 \( \overline{C_0}(V) \) determined by the class of functions \( \overline{V} = \{ V_\xi, \xi \in E(\xi) \} \) and denote by \( \overline{C}(V) \) the closed cone determined by \( \overline{V} \) 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., \( \Pi(V) \subset \Pi(W) \) \( \iff \) \( W \in \overline{C_0}(V) \))

**Proof.** As mentioned above the "if" part was essentially shown in [5].

The arguments preceding the statement of the theorem show that if \( W \in \overline{C}(V) \) there exists \( m \in M \) such that:

\[
(i) \quad V_\xi(f) dm_\xi(f) \geq 0 \quad \text{for every} \quad V_\xi \in \overline{C}(V) \\
\text{and} \quad (ii) \quad W(f) dm_\xi(f) < 0.
\]

18
Let $X_o, Y_o$ be a pair of experiments which induce $m_o^-$ and $m_o^+$ respectively on $F^+$. Then (i) $\Rightarrow (X_o, Y_o) \in \Pi(V)$ and (ii) $\Rightarrow (X_o, Y_o) \notin \Pi(W)$. It follows therefore, by contradiction that $(X, Y) \in \Pi(V) \Rightarrow (X, Y) \in \Pi(W)$ for every pair $X, Y$ requires that $W \in \mathcal{O}(V)$. Since $W$ is assumed to satisfy condition (2.1) (iv) (i.e., $W(f) = 0$ for all $f$ with only one non-zero component) it cannot have a positive linear term and therefore must be in the smaller cone $\mathcal{C}_o(V)$ 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 $(X, Y) \in \Pi(V) \iff (X, Y) \in \Pi(W)$ is valid for every pair $X, Y \in \mathcal{E}(\Omega)$? 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 $\mathcal{C}_o(V)$ and $\mathcal{C}_o(W)$ are identical.

Let $M_o$ denote the class of measures on the simplex which represent experiments and let $\nu_\mu(f) = \int S V(fg) d\mu(g), \mu \in M_o$. Then $\widetilde{V} = \{V'_{\lambda}, Z \in \mathcal{E}(\Omega)\} = \{V_{\lambda', \mu} \in M_o\}$, a convex set. $\lambda \widetilde{V}$ is the class of functions obtained by multiplying each member of $\widetilde{V}$ by the real number $\lambda$ and the cone generated by $V$ is $\mathcal{C}_o(V) = \bigcup_{\lambda \geq 0} \lambda \widetilde{V}$.

Note that if $W \in \lambda \widetilde{V}$ then so is $\alpha W$ where $0 \leq \alpha \leq 1$ since $\alpha W$ 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 $\mu \in M_o$ the following number:

$$\lambda(\mu) = \inf \{\lambda : V_{\lambda} \in \lambda \widetilde{V}\}$$
Clearly $0 \leq \lambda(\mu) \leq 1$ and if $\lambda(\mu) = \lambda \exists \mu^* \in M_0 \ni \lambda(\mu^*) = 1$ and $V_\mu = \lambda V_{\mu^*}$. The measure $\mu^*$ will be called maximal wherever $\lambda(\mu^*) = 1$.

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

**Lemma 4.1.** Let $W(f) = \lim \lambda_n V_{\mu_n}(f)$ uniformly for $f \in S$ with $\lambda_n \geq 0$ and $\mu_n$ maximal. Then $W \in C_o(V)$ if and only if $\{\lambda_n\}$ is a bounded sequence.

**Proof.** Suppose $W \in C_o(V)$. Then for some $a > 0$ and some $\mu \in M_0$

$W = aV_\mu$ where $\mu$ is maximal. Now if $\lambda_n$ is unbounded we can assume

for $n_o$ sufficiently large and $n \geq n_o \lambda_n > a + 1$ so that $\lambda_n V_{\mu_n} \notin \omega \lambda \overline{V}$

for all $n \geq n_o$ and $\lambda_n V_{\mu_n} \notin \omega \lambda \overline{V}$ for any $n \geq n_o$ and $\lambda \leq a$. This contradicts $W = aV_\mu$ and therefore $\lambda_n$ must be bounded. Conversely, if $\{\lambda_n\}$ is bounded we can assume $\lambda_n \to a \geq 0$ 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 $\lim \lambda_n V_{\mu_n} = aV_\mu$ and $W \in C_o(V)$ as was to be shown.

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

$\overline{V}(\mu) = \{V_{\mu_1, \eta_1} \in M_0 \mid (\int_S V_{\mu_1}(fg) d\eta(g), \eta \in M_0)\}$.

**Proof.** The proof follows immediately from the definition of maximality and the fact that $V_{\mu_1, \eta_1} \in \overline{V}$ 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 $W \in \overline{C}_o(V) - C_o(V)$. Then the only function that $\overline{C}_o(W)$ and $C_o(V)$ have in common is the zero function.
Proof. \( W \in \overline{C_o(V)} - C_o(V) \) implies \( W = \lim \lambda_n V \eta_n \) with \( \lambda_n \) maximal and \( \lambda_n \to \infty \). If \( \eta \in M_o \) then \( W_\eta(f) = \lim \lambda_n \int_S V\eta_n (fg) d\eta(g) \) since the convergence is uniform on compact sets. By Lemma 4.2 the maximal representation for \( W_\eta \) is given by

\[
W_\eta(f) = \lim \lambda_n \int_S V\eta_n (fg) d\mu_n(g).
\]

Now \( \int_S V\eta_n (fg) d\mu_n(g) \leq \int_S V(fg) d\mu_n(g) \to 0 \) uniformly for \( f \in S \). Hence \( \lambda_n \) is either unbounded or \( W_\eta(f) = 0 \).

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

\[
W_\eta(f) = \lim a_n \lambda_m(\eta) \int_S \int_S V(fgh) d\eta_n(g) d\mu_m(h).
\]

Again, since the double integral converges uniformly to zero we must have \( a_n \lambda_m(\eta) \) unbounded or \( W_\eta(f) = 0 \). Hence either \( W_\eta(f) \) is identically 0 or \( W_\eta(f) \in \overline{C_o(V)} - C(V) \).

We shall now drop the standard measure notation and resort 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 \( (X, Y) \in \Pi(V) \iff (X, Y) \in \Pi(W) \) is valid for all \( X, Y \in \mathcal{E}(\eta) \iff W(f) = V(gf) \) for some \( g \) such that \( g(\theta_i) > 0 \) for all \( i = 1, 2, \ldots, N \).
Proof. First note that if \( W \) is of the stated form \( W \) and \( V \) do generate identical orderings since the cones \( \overline{C}_o(V) \) and \( \overline{C}_o(W) \) 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 \( Y, Z \in \mathcal{E}(\Omega) \). 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 \( \tilde{W} = \{W_z, Z \in \mathcal{E}(\Omega)\} \) as can be seen from:

\[
W = cW_{Z_1} + (1-c)W_{Z_2} = \alpha Z_1 + (1-\alpha)Z_2
\]

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

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

\[
W_{(Y,Z)}(f) = \int W(g_{\frac{f}{g} r_z})d\eta(y)d\nu(z)
\]

then \( W(f_{\frac{f}{g} r_z}) = c(y,z)W(f) \) for every \( y \) and \( z \) and every \( f \in \mathbb{P}^+ \).

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 \( cg \),

22
\( c_{g-1}, g > 0 \) 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:

\[
Z = 0 \quad \text{with probability } \quad g > 0 \\
= X^* \quad \text{with probability } \quad 1-g
\]

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 \( FcF^+ \).

**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 non-negative 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 exists positive numbers \( c_1, c_2 \) such that

\[
V(a,b) = V_1(c_1a, c_2b) \quad \text{for all } a,b.
\]

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 cannot be related as above.

The following corollary expresses the result in terms of the loss function of the decision problem:
Corollary 4.2. The decision problems $(\omega, A, L_1)$ and $(\omega, A, L_2)$ induce the same ordering on the class $E(\omega)$ if and only if there exists $g$ such that $g(\omega_i) > 0$ for $i=1,2,\ldots,N.$ such that

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

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


Let $\mathcal{N}$ be a finite set of states and $\Xi$ the class of prior distributions on $\mathcal{N}$. Following DeGroot [5], a non-negative, continuous, concave function on $\Xi$ is called an uncertainty function and if $X = (\xi, \alpha; P_{\xi}, \theta_{\xi})$ and $Y = (\eta, \beta; Q_{\eta}, \theta_{\eta})$ are two experiments $X$ is called at least as informative as $Y$ with respect to $U$ if

$$U(\xi | X) \leq U(\eta | Y)$$

for all $\xi \in \Xi$.

where $U(\xi | X)$ 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 $Y$ which preserve its induced ordering.
Ordering of Experiments

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR’S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."
(2) "Foreign announcement and dissemination of this report by DDC is not authorized."
(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through "
(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through"
(5) "All distribution of this report is controlled. Qualified DDC users shall request through"

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (T), (R), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.