AN INFORMATION METRIC OF SIMILARITY OF EXPECTATIONS

by

Kevin D. Cotter

Discussion Paper No. 186, September, 1983

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

This paper along with the companion paper "Consumer Choice of Information With a 'Weak' Topology were presented under the latter title at the Summer North American Meeting of the Econometric Society, Evanston, Illinois, on June 25, 1983.

I would like to thank Professor James Jordan for generous doses of advice and encouragement as well as for pointing out a couple of counter-examples. Other comments were made by Professors Leonid Hurwicz, Marcel K. Richter, Thomas Armstrong, Beth Allen, and participants of the 1983 Summer North American Econometric Society meeting, particularly Larry Jones and Kerry Beck. They are gratefully appreciated. Any remaining errors are my own.

Financial support from the Sloan Foundation and a William F. Stout fellowship from the University of Minnesota Graduate School is acknowledged.

ABSTRACT

Although its specification in economic models with uncertainty is critical to the results obtained, agents' information has not been defined in a way allowing isolation of its effects on other variables. In addition, information needs to be allowed to be optimally chosen by agents when, for example, endogenous shifts in information are a major part of the model.

Both problems require a space of information with some appropriate mathematical structure depending only on the uncertainty in the model. Given a probability space, the space of information is the set of all sub-σ-fields of events. The structure defined here is a metric derived from the topology defined to be the weakest one such that for any integrable random variable, the function mapping information to the expected value of the random variable conditional on that information is continuous. This metric is shown to be complete and separable with two additional useful properties. First, set-theoretic convergent (monotone) sequences of information converge, so martingale convergence is modelled by this metric. Second, the set of finite partitions of the state space is dense, so any information can be finitely approximated. The topology derived from this metric is more tractible than the one used by Allen (1982). A generalization of the topology is provided to handle special cases such as information consisting of state plus noise.

1. Introduction

The specification of information in an environment with uncertainty can have a significant effect on subsequent activity. The Arrow-Debreu economy in which commodities are indexed by states of the world (cf. Debreu, 1957, Chapter 7) requires that agents have sufficient posterior information about the actual state to consummate trades in state-contingent commodities. The inability of some agents to observe states *ex post* is one reason that complete markets do not exist (Townsend, 1979), and moral hazard can result when such markets are open. Informational asymmetries about product characteristics or risk types can lead to adverse selection and subsequently to market failure or distortion.

Unfortunately, most studies of the economics of information have used a few special parameterizations, or been confined to the polar cases of "informed" and "uninformed" agents. Consequently, the findings of such models are difficult to generalize, with different results appearing to be unrelated. The reader perusing any recent survey of the literature such as Hirshleifer and Riley (1979) would learn about many phenomena involving information but find no statements of the formal relationships governing these phenomena. Given an economic environment with information, the only ways to determine the properties of equilibria have been either to compute them explicitly or to reduce the problem to a more familiar case. This limits the problems that can be studied and often suppresses the effects of information present.

The major difficulty is that economists have no formal language for making statements about information and its relationships with other variables. Assumptions about information can neither be precisely stated nor evaluated, unlike assumptions on other agent characteristics such as preferences, endowments, or technology. If, for example, the latter three were to be changed slightly, the resulting equilibria would also be changed slightly. No meaning can be given to the statement that information changes slightly. To ensure well-behaved equilibria in classical situations, assumptions on agents' characteristics are needed such as continuity and convexity of preferences. Both assumptions are based on mathematical structures on Euclidean spaces. Continuity depends on the existence of a topology (e.g., norm or metric) on the commodity space, while convexity requires that the commodity space be a vector space. To state assumptions about information necessary for the existence of equilibria, some such mathematical structure on the space of information is needed.

One set of problems receiving relatively little attention due to a lack of methods are those involving choice of information. Information is nearly always valuable to an individual, so agents have an incentive to pay for it. Firms able to produce and distribute information can therefore sell it for profit. When markets for information exist, some agents must choose what information to acquire while others must decide what information to produce and in what quantities.

Opening markets for information, or at least permitting some choice of information to use, is one way to allow agents' information to be endogenous. Shifts in agents' information due to a change in incentives are often a major aspect of economic problems. For example, the demise of the short-run Phillips curve occured during the 1970s when agents could no longer be "fooled" into increasing economic activity when faced with an increase in prices. One hypothesis is that a general and persistent price increase caused agents to devote more resources to distinguishing real from nominal price changes. This caused the Phillips curve to shift and government fiscal policy to fail. In another case, increased volatility of interest rates in recent years has caused investors to focus on money supply figures in an attempt to predict future interest rates. As a result, changes in M-1A have led to sudden and seemingly paradoxical changes in bond prices, altering the short-run effects of monetary policy. Few monetarists a decade ago would have predicted that an increase in the money supply would result in higher interest rates!

A third example of endogenous shifts in information involves the demand for physician services. With sharp increases in hospital and other medical care fees, consumers have increasingly sought second opinions regarding major expenditures such as surgery. Such second opinions, though costly, have become worthwhile given the expense of many operations. This change in consumer strategy may have an effect on the demand for medical services.

Unfortunately, most models have taken information as exogenous. Even studies of information choice such as Grossman and Stiglitz (1980), Chan and Leland (1982), and Verrecchia (1982) have involved only a few limited choices. These models are too special to extract many general principles.

To discuss optimal choices involving information, an explicit definition of the space of information is needed. Some mathematical structure such as a topology (e.g., metric) or a vector space structure must be imposed on this space. Every optimization result requires one or both of these structures on the choice set, particularly the Kuhn-Tucker theorems and the Maximum Theorem (e.g., Hildenbrand, 1974, p. 30). Since the best understood spaces are Euclidean, the chosen structure should ideally make the space of information resemble Euclidean space as well as satisfy the hypotheses of the above theorems.

The purpose of this paper is to establish such a mathematical structure, namely a metric. A formal definition of the space of information is given in Section 2, but an informal discussion is in order here. Suppose uncertainty is given by a set of states of the world with a probability distribution on those states. Call a subset of the state space an event.[1] Some "true" state is presumed to exist but is unknown. Care must be taken to distinguish between what Hirshleifer and Riley (1979) call an information service (to be called an information field or sub-σ-field in this paper) from the revelation made by it. The former reveals information depending on what the true state happens to be.

The information field is defined to be a list of events, not necessarily finite or disjoint. The information field reveals, for each event on the list, whether or not the true state of the world belongs to that event. For example, if the list of events is a disjoint partition of the state space, then the information field reveals the element of the partition in which the state lies. See Example 2.2 below for more details.

The above definition is consistent with the more common version of information as given by the observation of a random variable, where the random variable is identified with the list of events it generates. See Breiman (1968) for an explanation.

For mathematical convenience the following three conditions are imposed on any information field:

(1) the events defined by the empty set and the state space respectively are on the list,

(2) for any event on the list, its complement is also on the list,

(3) for any countably infinite or finite collection of events on the list, their union is also on the list.

The list of all possible events satisfies (1) - (3) and is called the σ-field (read sigma field) of the state space, hence other lists of events satisfying (1) - (3) are called sub-σ-fields. Thus the space of information is the set of all information (or sub-σ-) fields of the state space.

Any information field defines a conditional expectation operation. Given a random variable, which is a function mapping the state space to the real line, its conditional expectation given the information field can be defined. Since the information revealed by the information field depends on the true state of the world, so does the conditional expectation, hence the latter is a random variable. See Breiman (1968) or any measure-theoretic probability textbook for a definition. For finite partitions and information fields given by random variables, this definition of conditional expectation is the more familiar one. See Examples 2.2 and 2.3 below.

Though mathematicians have studied sub-σ-fields and conditional expectations extensively, most of the literature has been devoted to applications in statistical decision theory, communication theory, or natural stochastic phenomena. None of these applications require the type of structure on the space of information needed for economic problems.

As suggested earlier, some mathematical structure such as a metric or a vector space structure needs to be imposed on the space of information. A little thought will reveal to the reader that no vector space structure is possible in this case. The usual rules of vector addition and scalar multiplication do not apply to information. For example, two identical newspapers, added together, reveal the same information as one.[2] Though information can be combined (see Section 5), such combinations do not obey the rules of vector addition.

It is possible, however, to impose a metric on the space of information. As will be seen shortly, there are many possible choices of metric, but some are more useful than others. Loosely speaking, a metric determines how "far apart" any two points in the space are from each other. The chosen metric should embody a sound notion of similarity of information, based upon economically relevant properties. Most likely this would involve some concept of similarity of conditional expectations. The metric also defines a topology on the space. Any metric (more precisely, the topology derived from that metric) on a space determines which functions defined on that space are continuous. In the choice of information problem, value functions often arise from relationships between information and other variables such as consumer demand. Since value functions are usually required to be continuous for an optimum to exist, a useful metric on the space of information should make continuous as many of these relationships as possible. Such a metric also provides a natural meaning to the statement "a small change in information" in terms of the resulting changes in other economic variables.

Optimization problems frequently require a compact choice set, which in the choice of information problem is a subset of the space of information. Compactness is another property determined by the metric chosen. Thus a useful metric of information should make as many subsets of the space of information compact as possible. Technically this is equivalent to making the topology as weak as possible. A weaker topology, however, makes fewer functions continuous on the space.

The problem of metrizing the space of information was studied by Boylan (1971) who proposed a metric based on the properties of information fields as sets of events. This metric is analogous to the Hausdorff metric on closed sets (cf. Hildenbrand, 1974, p. 16) and is therefore called the Hausdorff metric in this paper. Allen (1982) has studied this metric further and explored its economic properties, which are mentioned in Section 4 below. It lacks some of the more important properties possessed by the Euclidean distance metric, namely separability and local compactness.

In this paper I propose a different metric of information. Here, two information fields are close if they lead to close conditional expectations for some given finite set of random variables. It is weaker (i.e., a weaker topology) than the Hausdorff metric, and possesses two important properties which fail for the latter. First, this metric is separable, and in fact, the set of finite partitions of the state space is a dense subset. In other words, any information field can be approximated in the metric to any degree of accuracy desired by some finite partition. This is useful since conditional expectations on finite partitions are easy to compute. Second, any sequence of information fields which is increasing in the sense that each information field is more informative than the preceding one converges in the metric. The latter result is important in applications to learning models.

One remaining issue regarding the two metrics is which economic variables, if any, are continuous with respect to which metrics of information. Allen (1982) showed that demand is a continuous function of information in the Hausdorff metric when a consumer faces a state-dependent utility function, initial information, and a known price and initial endowment. In a companion paper to the present one (Cotter, 1983), I demonstrate the same result using the metric of information described in this paper. This finding is used to study consumer choice of information. Other papers will study optimal supply of information and develop a theory of markets for information coincident with markets for other commodities.

In Section 2 a formal definition of the space of information and the metric are given, and some of its mathematical properties demonstrated. The denseness of the set of finite partitions of the state space is proven in Section 3. In Section 4, the metric of this paper is compared with the Hausdorff metric. Section 5 examines the continuity properties of combining information. A generalization of the metric is provided in Section 6. In Section 7 it is shown that the metric of information does not depend entirely on the probability distribution used, permitting asymmetries of beliefs in studying information across agents. Section 8 includes some concluding remarks. No economic applications are given in this paper; some may be found in Cotter (1983) and forthcoming work.

The reader lacking a firm background in general topology and functional analysis is advised to skim Section 2, concentrating on Examples 2.2 and 2.3, the statement of Proposition 2.5,, Corollaries 2.11-2.13, and the remarks after Corollary 2.13. The introductory remarks to Sections 3-7 may be read but not dwelled upon, though the results of Sections 3 and 4 have already been discussed in this section.

2. Information

Uncertainty is modeled as a probability space $(\Omega,\mathcal{F},\mu)$. Here Ω is the set of possible states of the world, $\mathcal{F}$ a σ-field of subsets of Ω, and μ a probability measure on $(\Omega,\mathcal{F})$. Information in this context has been defined by Allen (1982); we repeat the definition here for completeness. Let $\mathcal{F}^{**}$ be the set of all sub-σ-fields of $\mathcal{F}$. Define a relation $\sim$ on $\mathcal{F}^{**}$, where $\mathcal{B} \sim \mathcal{B}'$ if $\mathcal{B}$ and $\mathcal{B}'$ have the same μ-completion, that is, they differ only by null sets. This is clearly an equivalence relation. Let $\mathcal{F}^{*}$ be the set of $\sim$-equivalence classes of $\mathcal{F}^{**}$. For $\mathcal{B} \in \mathcal{F}^{**}$, and $f \in L^1(\mathcal{R}) = L^1(\Omega,\mathcal{F},\mu;\mathcal{R})$, we may define the conditional expectation $E[f|\mathcal{B}] \in L^1(\mathcal{R})$, (see Neveu (1965), p. 121, for details and properties of conditional expectation), up to sets of measure zero. Theorem 2 of Boylan (1971) motivates the equivalence relation $\sim$.

<u>Fact 2.1</u>: If $\mathcal{B}$, $\mathcal{B}' \in \mathcal{F}^{**}$, then $E[f|\mathcal{B}] = E[f|\mathcal{B}]$ a.e. for every $f \in L^1(\mathcal{R})$ iff $\mathcal{B} \sim \mathcal{B}'$.

Let $\mathcal{L}(L^1(\mathcal{R})) = \{T\colon L^1(\mathcal{R}) \to L^1(\mathcal{R}) \mid T$ is continuous and linear$\}$. Then for $\mathcal{B} \in \mathcal{F}^{**}$, and $f \in L^1(\mathcal{R})$, $\|E[f|\mathcal{B}]\| \leq \|f\|$ (with the L^1 norm), so $E[\cdot\ |\mathcal{B}] \in \mathcal{L}(L^1(\mathcal{R}))$, and the mapping $\{\mathcal{B}\} \mapsto E[\cdot\ |\mathcal{B}]$ is well-defined on equivalence classes of $\mathcal{F}^{*}$ and bijective. Henceforth, write $\mathcal{B}$ for the equivalence class $\{\mathcal{B}\} \in \mathcal{F}^{*}$. Then <u>information</u> is an element of $\mathcal{F}^{*}$.

<u>Example 2.2</u>: Let $\Omega^T = [-50^\circ F,\ 110^\circ F]$ be the outdoor temperature in Minneapolis, $\mathcal{F}^T$ the Borel sets of Ω^T, μ^T some probability measure on $(\Omega^T,\mathcal{F}^T)$. Some examples of information, to be used later, are:

(1) $\mathcal{B}^T_o = \{\Omega^T,\emptyset\}$, the <u>trivial</u> σ-field, which conveys no information about the temperature.

(2) $\mathcal{B}_1^T = \sigma\{[-50^\circ F,\ 32^\circ F),\ [32^\circ F,\ 110^\circ F]\}$, the information conveyed by a bucket of water left outside. Someone with $\mathcal{B}_1^T$ knows if it's freezing outside or not.

(3) $\mathcal{B}_2^T = \sigma\{[n - \frac{1}{2}{}^\circ F,\ n + \frac{1}{2}{}^\circ F)\}_{n=-50}^{110}$, the information conveyed by a reliable degree thermometer. Someone with $\mathcal{B}_2^T$ knows the temperature to the nearest degree.

Let $f\colon \Omega^T \to \mathbb{R}$ be integrable. For a.e. ω,

$$E[f|\mathcal{B}_o^T](\omega) = E[f] = \int_{-50}^{110} f(\nu)d\mu(\nu)$$

$$E[f|\mathcal{B}_1^T](\omega) = \begin{cases} \dfrac{1}{\mu\{[-50^\circ F,\ 32^\circ F]}\displaystyle\int_{-50}^{32} f(\nu)d\mu(\nu) & -50 \le \omega \le 32 \\[2ex] \dfrac{1}{\mu\{[n - \frac{1}{2}{}^\circ F,\ n + \frac{1}{2}{}^\circ F]\}}\displaystyle\int_{32}^{110} f(\nu)d\mu(\nu) & 32 \le \omega \le 110 \end{cases}$$

$$E[f|\mathcal{B}_2^T](\omega) = \frac{1}{\mu\{[n - \frac{1}{2}{}^\circ F,\ n + \frac{1}{2}{}^\circ F)\}}\int_{n-\frac{1}{2}}^{n+\frac{1}{2}} f(\nu)d\mu(\nu) \qquad \text{if } n - \frac{1}{2} \le \omega < n + \frac{1}{2}$$

Example 2.3: If $(\Omega,\mathcal{F},\mu)$ is any probability space and $f\colon \Omega \to \mathbb{R}$ is a random variable, the sub-σ-field $\sigma\{f^{-1}(a,b)|a < b\}$ is the information generated by f, called $\mathcal{B}(f)$. Then $E[\cdot|\mathcal{B}(f)] = E[\cdot\ |f]$, where the latter is the conditional expectation as computed by, say, Bayes' Rule.

Returning to the abstract model, we have seen how $\mathcal{F}^*$ is a subset of $(L^1(\mathbb{R}))$, which we endow with the strong operator topology as defined by Dunford and Schwartz (1957, p. 475). This is the same as the topology of pointwise convergence $\mathcal{T}_p$, the weakest topology on $\mathcal{F}^*$ such that for $f \in L^1(\mathbb{R})$, the map $\mathcal{B} \mapsto E[f|\mathcal{B}]$ is continuous.

Product topologies are not typically well-behaved, except for countable products, so one should not expect $(\mathcal{F}^*, \mathcal{T}p)$ to be a nice space. Since $\mathcal{F}^*$ is a very small subspace of $(L^1(\mathcal{R}))^{L^1(\mathcal{R})}$, however, we may be able to prove results based on the special properties of conditional expectation. In particular, elements of $\mathcal{F}^*$ are not only bounded linear operators, but the bound is unity for all elements of $\mathcal{F}^*$. This feature gives $\mathcal{F}^*$ a type of uniformity which will be used frequently in the sequel. We require one assumption which is not very restrictive. It is satisfied whenever $\mathcal{F}$ is countably generated, such as in the case where Ω is a second-countable Hausdorff space and $\mathcal{F}$ consists of all Borel subsets of Ω.

Assumption 2.4: $L^1(\mathcal{R})$ is separable, with countable dense subset $\{f_j\}_{j=1}^{\infty}$.

We now prove the central result of this section.

Proposition 2.5: If Assumption 2.4 holds, then $(\mathcal{F}^*, \mathcal{T}p)$ is a complete separable metric space with metric

$$\rho(\mathcal{B}, \mathcal{B}') = \sum_{j=1}^{\infty} 2^{-j} \min\{\|E[f_j|\mathcal{B}] - E[f_j|\mathcal{B}']\|, 1\}.$$

Proof: Throughout, let g and h be elements of $L^1(\mathcal{R})$, f_j is an element of $\{f_j\}_{j=1}^{\infty}$, we write $g = \lim_{j\to\infty} f_j$ instead of $g = \lim_{i\to\infty} f_{j_i}$ for some subsequence $\{f_{j_i}\}_{i=1}^{\infty}$, and ε, δ are positive reals.

Lemma 2.6: ρ is a metric on $\mathcal{F}^*$.

Proof: ρ is well-defined and by Fact 2.1, $\rho(\mathcal{B}, \mathcal{B}') = 0$ if $\mathcal{B} = \mathcal{B}'$. If $\rho(\mathcal{B}, \mathcal{B}') = 0$, then for each j,

$$\|E[f_j|\mathcal{B}] - E[f_j|\mathcal{B}']\| = 0$$

so given g and ε, by choosing j so that $\|g - f_j\| < \frac{\varepsilon}{2}$, we have

$$\|E[g|\mathcal{B}] - E[g|\mathcal{B}']\| \leq \|E[g|\mathcal{B}] - E[f_j|\mathcal{B}]\| + \|E[f_j|\mathcal{B}] - E[f_j|\mathcal{B}'\| + \|E[f_j|\mathcal{B}'] - E[g|\mathcal{B}']\| < \varepsilon.$$

Since ε is arbitrary, $\|E[g|\mathcal{B}] - E[g|\mathcal{B}']\| = 0$, hence by Fact 2.1, $\mathcal{B} = \mathcal{B}'$. Since $\rho(\mathcal{B},\mathcal{B}') = \rho(\mathcal{B}',\mathcal{B})$ and the triangle inequality holds, ρ is a metric.

<u>Lemma 2.7</u>: Let $\mathcal{T}$ be the metric topology on $\mathcal{F}^*$ generated by ρ. Then $\mathcal{T}p \subset \mathcal{T}$.

<u>Proof</u>: Let $B_\rho(\mathcal{B},\delta)$ be the open δ-ball about $\mathcal{B}$ in the metric ρ. Fixing $\mathcal{B}$ and δ, let $\mathcal{B}' \in B_\rho(\mathcal{B},\delta)$. Let $\delta' = \delta - \rho(\mathcal{B}',\mathcal{B})$ and N such that $\delta'(1 + 2^N) \geq 1$. For each i let $h_i = E[f_i|\mathcal{B}']$, and let

$$W^* = \{\mathcal{B}'' \in \mathcal{F}^* |\ \|E[f_j|\mathcal{B}''] - h_i\| < \delta', \quad i=1,\dots,N\}.$$

W^* is a basis element of $\mathcal{T}p$ (recall that $\mathcal{T}p$ is just the product topology). Obviously $\mathcal{B} \in W^*$. We now show $W^* \subset B_\rho(\mathcal{B},\delta)$, so let $\mathcal{B}'' \in W^*$. Then

$$\rho(\mathcal{B}'',\mathcal{B}) \leq \rho(\mathcal{B}',\mathcal{B}) + \rho(\mathcal{B}'',\mathcal{B}') < \delta - \delta' + \sum_{j=1}^{\infty} 2^{-j}\min\{\|E[f_j|\mathcal{B}''] - E[f_j|\mathcal{B}']\|,1$$

$$< \delta - \delta' + \sum_{j=1}^{N} 2^{-j}\delta' + \sum_{j=N+1}^{\infty} 2^{-j} = \delta - \delta' + \delta'(1 - 2^{-N}) + 2^{-N}$$

$$\leq \delta - \delta' + \delta'(1 - 2^{-N}) + \delta'(1 + 2^{-N}) = \delta.$$

Hence $\mathcal{B}'' \in \mathcal{B}_\rho(\mathcal{B},\delta)$, which proves the lemma.

<u>Lemma 2.8</u>: $\mathcal{T}_p$ is second-countable.

<u>Proof</u>: Let $g_1,\dots,g_N,h_1,\dots,h_N \in L^1(R)$ and $\delta_1,\dots,\delta_N > 0$. Define

$$W = \{\mathcal{B}' \in \mathcal{F}^* |\ \|E[g_i|\mathcal{B}] - h_i\| < \delta_i, i=1,\dots,N\}, \text{ and let } \mathcal{B} \in W.$$

For each i, let n_i be an integer satisfying

$$\|E[g_i|\mathcal{B}] - h_i\| < \frac{1}{n_i} < \delta_i$$

$$\delta_i' = \min\{\frac{1}{n_1} - \|E[g_i|\mathcal{B}] - h_i\|, \quad \delta_i - \frac{1}{n_i}\}.$$

Choose f_{i_1} and f_{i_2} such that $\|g_i - f_{i_1}\| < \frac{\delta_i'}{2}$ and $\|h_i - f_{i_2}\| < \frac{\delta_i'}{2}$. Then

$$\|E[f_{i_1}|\mathcal{B}] - f_{i_2}\| \leq \|E[f_{i_1}|\mathcal{B}] - E[g_i|\mathcal{B}]\| + \|E[g_i|\mathcal{B}] - h_i\| +$$
$$\|h_i - f_{i_2}\| < \frac{1}{n_i},$$

hence, letting

$$W' = \{\mathcal{B}' \in \mathcal{F}^* |\ \|E[f_{i_1}|\mathcal{B}'] - f_{i_2}\| < \frac{1}{n_i},\ i=1,\ldots,N\}, \quad \mathcal{B} \in W'.$$

For $\mathcal{B}' \in W'$,

$$\|E[g_i|\mathcal{B}'] - h_i\| \leq \|E[g_i|\mathcal{B}'] - E[f_{i_1}|\mathcal{B}']\| + \|E[f_{i_1}|\mathcal{B}'] - f_{i_2}\|$$
$$+ \|f_{i_2} - h_i\|$$
$$< \frac{\delta_i'}{2} + \frac{1}{n_i} + \frac{\delta_i'}{2} < \delta_i$$

hence $\mathcal{B}' \in W'$. Since there are countably many sets of the form W', the lemma is proved.

<u>Lemma 2.9</u>: $\mathcal{T} \subset \mathcal{T}_\rho$, hence $\mathcal{T} = \mathcal{T}_\rho$.

<u>Proof</u>: Let W' be as before, and let $\mathcal{B} \in W'$. Let

$$\delta = \min_i\{2^{-i_1}(\frac{1}{n_i} - \|E[f_{i_1}|\mathcal{B}] - f_{i_2}\|)\}, \text{ then for } \mathcal{B}' \in B_\rho(\mathcal{B},\delta)$$

and i,

$$2^{-i_1} \min\{\|E[f_{i_1}|\mathcal{B}'] - E[f_{i_1}|\mathcal{B}]\|, 1\} < \delta \text{ so}$$

$$\|E[f_{i_1}|\mathcal{B}'] - E[f_{i_1}|\mathcal{B}]\| < \frac{1}{n_i} - \|E[f_{i_1}|\mathcal{B}] - f_{i_2}\|$$

hence $\|E[f_{i_1}|\mathcal{B}'] - f_{i_2}\| < \delta_i$, showing $B_\rho(\mathcal{B},\delta) \subset W'$ and proving the lemma.

Lemma 2.10: $\mathcal{T}$ is complete.

Proof: Let $\{\mathcal{B}_n\}_{n=1}^{\infty}$ be a Cauchy sequence. Given j and ε, there exists N such that for $m,n \geq N$, $\rho(\mathcal{B}_n,\mathcal{B}_m) < \varepsilon \cdot 2^{-j}$, hence $\|E[f_j|\mathcal{B}_n] - E[f_j|\mathcal{B}_m]\| < \varepsilon$. Then $\{E[f_j|\mathcal{B}_n]\}_{n=1}^{\infty}$ is a Cauchy sequence, so since $L^1(\mathcal{R})$ is complete, there exists $g_j \in L^1(\mathcal{R})$ with $\lim_{n\to\infty} \|E[f_j|\mathcal{B}_n] - g_j\| = 0$.

The proof now proceeds in steps.

Step 1: There exists a unique $T: L^1(\mathcal{R}) \to L^1(\mathcal{R})$ linear and continuous such that $Tf_j = g_j$ for each j.

Proof: Let $\{f_j\}_{j=1}^{\infty}$ be a subsequence with $\lim_{j,k\to\infty} \|f_j - f_k\| = 0$. Given ε, choose I such that for $j, k \geq I$, $\|f_j - f_k\| < \frac{\varepsilon}{3}$. Fixing $j,k \geq I$, choose n such that

$$\|g_j - E[f_j|\mathcal{B}_n]\| < \frac{\varepsilon}{3}$$
$$\|g_k - E[f_k|\mathcal{B}_n]\| < \frac{\varepsilon}{3}.$$

Then for $j,k \geq I$,

$$\|g_j - g_k\| \leq \|g_j - E[f_j|\mathcal{B}_n]\| + \|E[f_j|\mathcal{B}_n] - E[f_k|\mathcal{B}_n]\| + \|E[f_k|\mathcal{B}_n] - g_k\| < \varepsilon.$$

Using completeness of $L^1(\mathcal{R})$, the step is proven, where $Tf_j = g_j$ for each j, and $Tg = \lim_{j\to\infty} Tf_j$ where $\lim_{j\to\infty} f_j = g$.

Step 2: For each j, $Tg = \lim_{n\to\infty} E[g|B_n]$.

Proof: Let $\{f_j\}_{j=1}^{\infty}$ be a subsequence converging to g, and let $\varepsilon > 0$. Choose j_o such that $\|Tg - Tf_{j_o}\| < \frac{\varepsilon}{2}$ and $\|f_{j_o} - g\| < \frac{\varepsilon}{3}$, and choose N such that for $n \geq N$, $Tf_{j_o} - E[f_{j_o}|\mathcal{B}_n]\| < \frac{\varepsilon}{3}$. Then

$$\|Tg - E[g|\mathcal{B}_n]\| \leq \|Tg - Tf_{j_o}\| + \|Tf_{j_o} - E[f_{j_o}|\mathcal{B}_n]\| + \|E[f_{j_o}|\mathcal{B}_n] - E[g|\mathcal{B}_n]\| < \varepsilon.$$

This proves the step.

<u>Step 3</u>: For each g, $E[Tg] = g$.

<u>Proof</u>: $E[Tg] = E[\lim_{n\to\infty} E[g|\mathcal{B}_n]] = \lim_{n\to\infty} E[E[g|\mathcal{B}_n]] = g$.

<u>Step 4</u>: If $g_1 \in L^1(\mathcal{R})$ with ess sup $|g_1| = c < \infty$, then ess sup $|Tg_1| \le c$.

<u>Proof</u>: For each n, ess sup $|E[g_1|\mathcal{B}_n]| \le c$, hence ess sup $|Tg_1|$ = ess sup $(\lim_{n\to\infty} |E[g_1|\mathcal{B}_n]|) \le \lim_{n\to\infty}$ (ess sup $|E[g_1|\mathcal{B}_n]|) \le c$.

<u>Step 5</u>: For $g_1, g_2 \in L^1(\mathcal{R})$ with ess sup $|g_1| = c < \infty$, then $T(g_1Tg_2) = (Tg_1)(Tg_2)$ a.e.

<u>Proof</u>: If $g_1 = 0$ a.e. the result is trivial so suppose $c > 0$. Given $\varepsilon > 0$, there exists $\hat{g}_2 \in L^\infty(\mathcal{R})$ with $\|\hat{g}_2 - g_2\| < \frac{\varepsilon}{c}$. Writing ess sup $|\hat{g}_2| = c'$, choose n so that

$$\|Tg_2 - E[g_2|\mathcal{B}_n]\| < \frac{\varepsilon}{c},\quad \|T(g_1Tg_2) - E[g_1Tg_2|\mathcal{B}_n]\| < \varepsilon, \text{ and}$$
$$\|E[g_1|\mathcal{B}_n] - Tg_1\| < \frac{\varepsilon}{c'}.$$

Using the triangle inequality, we have

$$\begin{aligned}
\|T(g_1Tg_2) - (Tg_1)(Tg_2)\| \le\ & \|T(g_1Tg_2) - E[g_1Tg_2|\mathcal{B}_n]\| + \|E[g_1Tg_2|\mathcal{B}_n] \\
& - E[g_1E[g_2|\mathcal{B}_n]|\mathcal{B}_n]\| + \|E[g_1|\mathcal{B}_n]E[g_2|\mathcal{B}_n] - E[g_1|\mathcal{B}_n]E[\hat{g}_2|\mathcal{B}_n]\| \\
& + \|E[g_1|\mathcal{B}_n]E[\hat{g}_2|\mathcal{B}_n] - Tg_1E[\hat{g}_2|\mathcal{B}_n]\| + \|Tg_1E[\hat{g}_2|\mathcal{B}_n] - Tg_1E[g_2|\mathcal{B}_n]\| \\
& + \|Tg_1E[g_2|\mathcal{B}_n] - Tg_1Tg_2\| \\
\le\ & \varepsilon + c\cdot\|Tg_2 - E[g_2|\mathcal{B}_n]\| + c\cdot\|g_2 - \hat{g}_2\| + \|E[g_1|\mathcal{B}_n] - Tg_1\|\cdot c' + \\
& \text{ess sup } |Tg_1|\cdot\|\hat{g}_2 - g_2\| + \text{ess sup } |Tg_1|\cdot\|E[g_2|\mathcal{B}_n] - Tg_2\| < 6\varepsilon
\end{aligned}$$

by Step 4. Since ε is arbitrary, the proof is complete.

<u>Step 6</u>: Using Neveu (1975, Proposition I-2-13) and Steps 1, 3, and 5, there exists a unique $\mathcal{B} \in \mathcal{F}^*$ with $T = E[\cdot\,|\mathcal{B}]$ a.e.. Then $\rho(\mathcal{B}_n,\mathcal{B}) \to 0$, so $\mathcal{B}_n \to \mathcal{B}$, completing the proof of the proposition. □

I now present some useful criteria for convergence in $\mathcal{T}_p$. In general, information converging in a set-theoretic sense converges in the topology.

<u>Corollary 2.11</u>: If $\{\mathcal{B}_n\}_{n=1}^{\infty}$ is a sequence in $\mathcal{F}^*$ such that for every f in a dense subset of $L^1(\mathcal{R})$, $E[f|\mathcal{B}_n]$ converges a.e. or in probability, then $\mathcal{B}_n$ converges in $\mathcal{T}p$.

<u>Proof</u>: The sequence $\{E[f|\mathcal{B}_n]\}_{n=1}^{\infty}$ is uniformly integrable, hence it converges in L^1. The rest of the proof follows the proof of Lemma 2.10. □

<u>Corollary 2.12</u>: If $\{\mathcal{B}_n\}_{n=1}^{\infty}$ is a sequence in $\mathcal{F}^*$ such that for every f in a dense subset of $L^1(\mathcal{R})$, $E[f|\mathcal{B}_n]$ converges in L^P for some $1 \leq p \leq \infty$, then $\mathcal{B}_n$ converges in $\mathcal{T}_p$.

<u>Proof</u>: Convergence in L^P for some $1 \leq p \leq \infty$ implies convergence in L^1. □

<u>Corollary 2.13</u>: If $\{\mathcal{B}_n\}_{n=1}^{\infty}$ is a sequence in $\mathcal{F}^*$ with $\bigvee_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \mathcal{B}_n = \bigcap_{k=1}^{\infty} \bigvee_{n=k}^{\infty} \mathcal{B}_n$, then $\mathcal{B}_n$ converges in $\mathcal{T}p$. In particular, if $\mathcal{B}_n$ is increasing or decreasing in the sense that $\mathcal{B}_n \subset \mathcal{B}_{n+1}$ (resp. $\mathcal{B}_n \subset \mathcal{B}_{n+1}$) for each n, then $\mathcal{B}_n$ converges in $\mathcal{T}_p$.

<u>Proof</u>: By Fetter (1975), $E[f|\mathcal{B}_n]$ converges a.e. for each $f \in L^1(\mathcal{R})$. Use Corollary 2.10. □

<u>Remark</u>: In practice, the elements of $\mathcal{F}^*$ one would wish to consider are sub-σ-fields generated by random variables. If $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables, it is not clear at this point whether convergence of $\{\mathcal{B}(X_n)\}_{n=1}^{\infty}$ (see Example 2.3) in $\mathcal{T}_p$ is related to any

notions of convergence of $\{X_n\}_{n=1}^{\infty}$. Convergence of random variables in any standard sense is neither necessary nor sufficient for convergence of their sub-σ-fields. To see why, let X be a random variable and $X_n = nX$. Then $\mathcal{B}(X_n) = \mathcal{B}(X)$ for each n but $\{X_n\}_{n=1}^{\infty}$ does not converge a.e., in probability, in L^p for any p, weakly, or in distribution. Conversely, let $(\Omega,\mathcal{F},\mu) = ([0,1],$ Borel sets, Lebesgue measure) and for each n let

$$X_n(\omega) = \begin{cases} -\frac{1}{n} & \omega < \frac{1}{2} \\ \frac{1}{n} & \omega \geq \frac{1}{2} \end{cases}$$

so $X_n \to 0$ a.e. and in L^p for every p. But $\mathcal{B}(X_n) = \{[0,1],\emptyset,[0,\frac{1}{2}),[\frac{1}{2},1]\}$ while $\mathcal{B}(0) = \{[0,1],\emptyset\}$ so $\mathcal{B}(X_n)$ does not converge to $\mathcal{B}(0)$.

This failure of convergence of sub-σ-fields to correspond to convergence of random variables is related to the problem of the nonexistence of rational expectations equilibrium (cf. Jordan and Radner (1977)). Consumer demand, as shown in Cotter (1983) is jointly continuous in state-dependent price and information. If the information available from the price were a continuous function of the price, then the existence of a rational expectations equilibrium would be immediate. As is clear from the above, the latter function is discontinuous. Therein lies the source of the difficulty.

In general, $(\mathcal{F}^*,\mathcal{T}p)$ is not compact. A probability space with a sequence of sub-σ-fields containing no convergent subsequence is given by Kudo (1974, Example 3.1). As shown in Proposition 5.1, compactness holds when the probability space is purely atomic.

Elementary conditions for compactness of subsets of $\mathcal{F}^*$ are needed.

A subset $\mathcal{G}^*$ of $\mathcal{F}^*$ is compact if and only if for every $f \in L^1(\mathcal{R})$, the set $\{E[f|\mathcal{B}]\}_{\mathcal{B}\in\mathcal{G}^*}$ is L^1-compact (cf. Dunford and Schwartz (1957, p. 511, Exercise 2). A standard diagonalization argument verifies that the latter is equivalent to compactness of $\{E[f|\mathcal{B}]\}_{\mathcal{B}\in\mathcal{G}^*}$ for every f in a dense subset of $L^1(\mathcal{R})$. This condition is not very useful since compactness in L^1-norm is difficult to verify. A useful necessary condition for relative compactness of subsets is given in the next section.

Note that $\mathcal{F}^*$ is not a convex subset of $\mathcal{L}(L^1(\mathcal{R}))^3$ so notions of convexity are difficult to apply to information. This makes single-valuedness of optimal choice functions of information harder to obtain and requires more general fixed-point theorems.

3. The Density Property

At first glance, the definition of information stated in the previous section may appear to have little practical value. Except for special cases, conditional expectations are difficult to compute, even when information is given by a random variable (see Example 2.3). The easiest situations to handle are those in which information is given by a finite partition, in which case conditional expectations can be computed as in Example 2.2. In nearly all models involving information, some restrictions on the types of information considered are necessary. Care must be taken to ensure that the assumptions made are not restrictive and do not obscure the essential features of the phenomenon under study. If the restriction of information to finite partitions of the state space is in some sense not restrictive, then the computations can be greatly simplified and the power of such models enhanced.

Since the space of information has a topology given by Proposition 2.5, the above statement can be made precise. In this section we prove that the set of all finite partitions is a dense subset of the space of information. That is, for any information field $B \in F^*$ and $\varepsilon > 0$, there exists a finite partition $\hat{B}$ such that $\rho(B,\hat{B}) < \varepsilon$. Hence if the economy depends continuously on the information it contains, restricting attention to finite partitions should cause less concern. A corollary to this result (henceforth called the density property) is that for any compact subset $C^* \subset F^*$, and any $\varepsilon > 0$, there exists a finite partition $P \in F^*$ such that for any $B \in C^*$ there exists a subpartition of P whose distance from B is less than ε. Since compactness of allowable information is often needed, we obtain at no additional cost the assumption that all information may be assumed to be bounded above by some fixed finite partition.

Another nice property of finite partitions is that if all consumer information is contained in some finite partition C, then all candidates for market clearing prices and demands may be taken to be C-measurable (provided that other other characteristics are nonstochastic). Since conditional expectation depends on the state of the world, consumer behavior is also state-dependent, hence excess demands and market clearing prices are random variables. In general, equilibrium analysis would require appeal to results about infinite-dimensional spaces which are difficult to use. By taking all such variables to be C-measurable, they become finite-dimensional, allowing results to be used applying to finite-dimensional spaces.

The following result may prove to be the most useful property of the topology of information T_p.

<u>Proposition 3.1 (The Density Property)</u>: Let P^* be the subset of F^* consisting of all finite F-measurable partitions of Ω. Then P^* is dense in F^*.

<u>Proof</u>: Let $\{f_j\}_{j=1}^N$ be a finite subset of $L^1(R)$, assuming without loss of generality that each f_j is bounded almost everywhere. Let $B \in F^*$ and $\varepsilon > 0$. For each j define

$$B_{ji} = \{\omega \in \Omega | \frac{\varepsilon \cdot i}{2} \leq E[f_j|B](\omega) \leq \frac{\varepsilon \cdot (i+1)}{2}\}$$

and $B_j = \sigma\{B_{j1}, B_{j2}, \ldots, B_{jI}, B_{j,-1}, B_{j,-2}, \ldots, B_{j0}\}$, where σ indicates closure under complementation, countable intersections, and countable unions. An easy computation then shows that for a.e. ω and for each j,

$$|E[f_j|B] - E[f_j|B_j]| < \varepsilon$$

Let $B' = \sigma\{B_1, B_2, \ldots, B_N\}$ so that for each j, $E[f_j|B_j] = E[f_j|B']$ from which the result follows.

<u>Corollary 3.2</u>: A subset G^* of F^* is relatively compact if and only if given $\varepsilon > 0$ there exists $C \in P^*$ such that for $B \in G^*$, there exists B' with $B' \subset C$ such that $\rho(B,B') < \varepsilon$.

4. Comparison With the Hausdorff Topology

An alternative topology on F^* due to Boylan (1971) uses the Hausdorff metric on the sub-σ-fields of F considered as closed subsets of F, resulting in a complete metric space. See Allen (1982) for the definition. Allen proved that the Hausdorff topology (denoted T_H) is identical to the uniform $L(L^\infty, L^1)$ operator topology. In other words, a sequence $\{B_n\}_{n=1}^{\infty}$ converges to B in T_H if and only if

$$\sup_{\{\text{ess sup } |f| \leq 1\}} \lim_{n\to\infty} \|E[f|B_n] - E[f|B]\| = 0$$

Clearly T_H is stronger than T_p. Since (F^*, T_H) is compact by Allen's Corollary 13.2, the following result is immediate.

Proposition 4.1: If (Ω, F, μ) is purely atomic, then $T_p = T_H$ and F^* is compact in either topology.

For the remainder of this section assume that the hypothesis of the above result does not hold. Allen (1982) has shown that the Hausdorff topology is not well behaved. In particular, it is not separable (Proposition 13.5) and the density property does not hold (Proposition 13.6). In addition, set-theoretic convergent sequences of information need not converge in the Hausdorff topology (Example 8.1). The latter result may prove to be troublesome in applications to learning models.

As explained earlier, finite partitions are in general the only sub-σ-fields whose conditional expectations are easily computable, so the failure of the density property for the Hausdorff metric may limit the practical usefulness of that topology. In addition, models involving infinite-dimensional commodity spaces (e.g., Bewley (1972), Mas-Colell (1975), Ostroy (1982)) often use finite-dimensional approximations of the commodity space to establish results about the general case. When information is present, subsequent demand for commodities will be state-dependent since conditional expected utility is. To obtain finite-dimensional approximations may require approximating information by finite partitions.

In the companion paper (Cotter, 1983) I show that the continuity results regarding utility and demand demonstrated by Allen (1982, Sections 10 and 11) hold in the pointwise convergence metric. At least for the results so far established, the pointwise convergence metric appears to be the more useful topology of information. More study of the continuity properties of economic variables with respect to information is needed.

5. Combining Information

When initial information is available to a decision-maker, any subsequent information received will be combined with that initial information. A natural question to ask is whether the resulting total information varies continuously with either or both of its components. Some continuity result is needed to study the case of information acquisition when initial information is present, in which case total information should be a continuous function of the information to be acquired.

The operation of combining two information sub-σ-fields, called the join operation, is not expressible in terms of the conditional expectations of the individual sub-σ-fields, as can be seen by considering $E\left[r \mid s_1, s_2\right]$ where r, s_1, and s_2 are jointly normal. In terms of set operations, however, a convenient expression for the join operation is available. If $\mathcal{A}$ and $\mathcal{B}$ are sub-σ-fields, then their join, denoted $\mathcal{A} \vee \mathcal{B}$, is the smallest sub-$\sigma$-field containing both $\mathcal{A}$ and $\mathcal{B}$. Allen (1982, Lemma 14.1) has shown that the join is jointly uniformly continuous in its arguments. That is, if d is the Hausdorff metric, then given $\varepsilon > 0$ there exists $\delta > 0$ such that if $d(\mathcal{A},\mathcal{A}') < \delta$ and $d(\mathcal{B},\mathcal{B}') < \delta$ then $d(\mathcal{A} \vee \mathcal{B},\mathcal{A}' \vee \mathcal{B}') < \varepsilon$.

Unfortunately, this continuity result does not hold in the pointwise convergence metric. In this case, the join is not even separately continuous. In other words, for fixed $\mathcal{B}$, the map $\mathcal{A} \to \mathcal{A} \vee \mathcal{B}$ is not in general continuous. The following construction is due to James Jordan.

Example 5.1: Let $(\Omega,\mathcal{F},\mu)$ be the closed unit square with Borel sets and Lebesgue measure. Let $\mathcal{B}_0$ be the sub-σ-field generated by the map $(x,y) \to y$ of Ω into $\mathcal{R}$, and $\mathcal{B}_n$ the sub-σ-field generated by the map $(x,y) \to y + x/n$. Thus $\mathcal{B}_0$ and $\mathcal{B}_n$ partition Ω into all possible strips of slope 0 and $-1/n$ respectively, as shown in Figure 1.

Let f belong to L^1. Clearly $E[f|\mathcal{B}_0](x,y) = \int_0^1 f(t,y)dt$, and on the set $A_n = \{(x,y) \mid 1/n \leq x/n + y \leq n\}$,

$$E[f|\mathcal{B}_n](x,y) = \int_0^1 f(t,y + (x - t)/n)dt .$$

If f is continuous then $f(t,y + (x - t)/n) \to f(t,y)$, and $\mu(A_n) \to 1$ as $n \to \infty$. Thus by dominated convergence (recall f is continuous on a compact set), $E[f|\mathcal{B}_n] \to E[f|\mathcal{B}_0]$ a.e., and again by dominated convergence, in L^1 norm. Since the set of continuous functions on Ω is dense in L^1, this shown that $\mathcal{B}_n \to \mathcal{B}_0$.

The reader may easily verify that $\mathcal{B}_n \vee \mathcal{B}_0 = \mathcal{F}$ for each n by demonstrating that the topology on Ω formed by the $\mathcal{B}_n$-strips and the $\mathcal{B}_0$-strips is the Euclidean topology. Therefore, $\mathcal{B}_n \vee \mathcal{B}_0 \not\to \mathcal{B}_0 \vee \mathcal{B}_0 = \mathcal{B}_0$ so the mapping $\mathcal{B} \to \mathcal{B} \vee \mathcal{B}_0$ is not continuous.

The manner in which continuity fails in the above example is of some interest. Information appears to collapse in the limit the same way that the span of a set of vectors collapses in the limit as two vectors in the set converge to each other, creating a linear dependency. In the simplest case, two vectors, however close but distinct, can span R^2 , yet a slight perturbation of one of the vectors can reduce the span by a dimension. This suggests that the map $\mathcal{A} \to \mathcal{A} \vee \mathcal{B}$ may be continuous for "most" fixed $\mathcal{B}$ and lowersemicontinuous for all fixed $\mathcal{B}$.

The reader is invited to ponder in light of the above whether or not requiring the join to be continuous in the chosen topology of information is a reasonable restriction on the topology.

When one of the arguments is a finite partition, matters work out rather nicely.

Proposition 5.2: Let $\mathcal{B}$ be a finite partition of Ω. Then the map $\mathcal{A} \to \mathcal{A} \vee \mathcal{B}$ is continuous on F^*.

Lemma 5.3: Let $\mathcal{A}_n$ and $\mathcal{B}_n$ be increasing to $\mathcal{A}$ and $\mathcal{B}$ respectively. Then $\mathcal{A}_n \vee \mathcal{B}_n$ is increasing to $\mathcal{A} \vee \mathcal{B}$.

Proof: Clearly $\mathcal{A}_n \vee \mathcal{B}_n$ is increasing to some sub-σ-field $\mathcal{C} \subset \mathcal{A} \vee \mathcal{B}$. Let $D \in \mathcal{A} \vee \mathcal{B}$, so D equals one of $A \cap B$, $A \cup B$ for some $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Given $\delta > 0$ there exists N and $A_N \in \mathcal{A}_N$, $B_N \in \mathcal{B}_N$ such that $\tau(A_N,A) < \delta$ and $\tau(B_N,B) < \delta$ (Chung, 1974, p. 251) where $\tau(E,F) = \mu(E \cup F) - \mu(E \cap F)$ is the symmetric distance function on events. Since $\cap$ and $\cup$ are uniformly continuous with respect to τ (Halmos, 1950, p. 168), given $\varepsilon > 0$ there exists N (this N may be different from the previous one) such that $\tau(A \cap B, A_N \cap B_N) < \varepsilon$ or $\tau(A \cup B, A_N \cup B_N) < \varepsilon$ as the case may be. Therefore, $\mathcal{A} \vee \mathcal{B} \subset \overline{\bigcup_{n=1}^{\infty} \mathcal{A}_n \vee \mathcal{B}_n}$. Since

$$\bigcup_{n=1}^{\infty} \mathcal{A}_n \vee \mathcal{B}_n \subset \mathcal{C} \subset \overline{\bigcup_{n=1}^{\infty} \mathcal{A}_n \vee \mathcal{B}_n}\ ,$$ this shows that $\mathcal{A} \vee \mathcal{B} = \mathcal{C}$,

completing the proof.

Lemma 5.4: Let $\mathcal{A}$ belong to F^* and let $\hat{\mathcal{A}} \subset \mathcal{A}$ be a finite partition. Then given $\varepsilon > 0$ there exists a neighborhood (in the pointwise convergence metric) such that for any $\mathcal{A}'$ in that neighborhood, there exists some $\hat{\mathcal{A}}' \subset \mathcal{A}'$ with $d(\hat{\mathcal{A}},\hat{\mathcal{A}}') < \varepsilon$ where d is the Hausdorff metric.

Proof: Write $\hat{\mathcal{A}}$ as $\{A_1, A_2, \ldots, A_M\}$ forming a disjoint partition of Ω. Let $N^* = \{\mathcal{A}' \in F^* \mid \| P[A_i|\mathcal{A}] - P[A_i|\mathcal{A}'] \|_1 < \varepsilon/2M$ for all $i\}$. Choose $\mathcal{A}' \in N^*$. Then by Chebyshev's inequality, there exists $W \in F$ with $\mu(W) > 1 - \varepsilon$ such that for every $\omega \in W$,

$$|P[A_i|A](\omega) - P[A_i|A'](\omega)| = |I_{A_i}(\omega) - P[A_i|A'](\omega)| < 1/2$$

Therefore, on $A_i \cap W$, $P[A_i|A'] > 1/2$ and on $A_i^c \cap W$, $P[A_i|A'] < 1/2$. Let $A_i' = \{\omega |\ P[A_i|A'] > 1/2\}$, so $A_i' \cap W = A_i \cap W$. Let $\hat{A}'$ be the finite partition formed by $A_1', \ldots, A_M'$, so $\hat{A}$ and $\hat{A}'$ form the same partition on W. Since $\mu(W^c) < \varepsilon$, this completes the proof.

<u>Lemma 5.5</u>: Let $\mathcal{B}$ be a finite partition. Then given ε there exists δ such that if $\hat{A} \subset A$ with the former a finite partition, and if $\rho(A,\hat{A}) < \delta$, then $\rho(A \vee \mathcal{B}, \hat{A} \vee \mathcal{B}) < \varepsilon$.

<u>Proof</u>: Let $C_1, \ldots, C_N$ be events such that $\rho(A,A') < \varepsilon$ if for each i $\| P[C_i|A] - P[C_i|A']\| < \varepsilon$ and let δ be such that if $\rho(A,A') < \delta$ then $\|P[B_j|A] - P[B_j|A']\| < \mu(B_j)\varepsilon/2$ and $\|P[B_j \vee C_i|A] - P[B_j \vee C_i|A']\| < \mu(B_j)\varepsilon/2$ for each i and j, where $B_1, \ldots, B_M$ is a disjoint partition of Ω generated by $\mathcal{B}$.

Let $\hat{A} \subset A$ be finite partitions, with $\rho(A,\hat{A}) < \delta$. Then letting $A_1, \ldots, A_I$ be a disjoint partition generated by $\hat{A}$ and $A_{11}, A_{12}, \ldots, A_{IJ_I}$ be a disjoint partition generated by A with $A_{ij} \in A_i$, it follows that

$$\sum_{j,n,m} \left|\mu(A_{nm} \cap B_j) - \frac{\mu(A_n \cap B_j)\mu(A_{nm})}{\mu(A_n)}\right| < \frac{\varepsilon}{2}$$

$$\sum_{j,n,m} \left|\mu(A_{nm} \cap B_j \cap C_i) - \frac{\mu(A_n \cap B_j \cap C_i)\mu(A_{nm})}{\mu(A_n)}\right| < \frac{\varepsilon}{2}$$

A simple computation shows that $\|P[C_i|A \vee \mathcal{B}] - P[C_i|\hat{A} \vee \mathcal{B}]\| =$

$$\sum_{j,n,m} \left|\mu(A_{nm} \cap B_j \cap C_i) - \frac{\mu(A_n \cap B_j \cap C_i)\mu(A_{nm} \cap B_j)}{\mu(A_n \cap B_j)}\right|$$

where each term is taken to be 0 if $\mu(A_{nm} \cap B_j) = 0$. Thus $\|P[C_i|A \vee \mathcal{B}] - P[C_i|\hat{A} \vee \mathcal{B}]\| \leq \sum_{j,n,m} \left|\mu(A_{nm} \cap B_j \cap C_i) - \frac{\mu(A_n \cap B_j \cap C_i)\mu(A_{nm})}{\mu(A_n)}\right|$

$$+ \sum_{j,n,m} \frac{\mu(A_n \cap B_j \cap C_i)}{\mu(A_n \cap B_k)} \left| \frac{\mu(A_{nm})\mu(A_n \cap B_j)}{\mu(A_n)} - \mu(A_{nm} \cap B_j) \right| < \varepsilon.$$

Now choose δ such that if $\hat{A}' \subset A'$ are finite partitions with $\rho(\hat{A}',A') < 3\delta$ then $\rho(\hat{A}' \vee B, A' \vee B) < \frac{\varepsilon}{3}$. Let $A \in F^*$ and $\hat{A} \subset A$, $\hat{A}$ a finite partition, with $\rho(\hat{A},A) < \delta$.

By Lemma 5.4, there exists η such that if A' satisfies $\rho(A,A') < \eta$, then there exists $\hat{A}' \subset A'$ a finite partition with $d(\hat{A}',\hat{A}) < \xi$, where for $G,G' \in F^*$, $d(G,G') < \xi \Rightarrow \rho(G \vee B, G' \vee B) < \frac{\varepsilon}{3}$.

Take ζ and η to be less than δ. Choose $A' \subset A$ with $\rho(A,A') < \eta$, $\rho(A \vee B, A' \vee B) < \frac{\varepsilon}{3}$ (by Lemma 1), and $\hat{A}' \subset A'$ a finite partition with $d(\hat{A}',\hat{A}) < \xi$, so $\rho(\hat{A}',\hat{A}) < \delta$ and $\rho(\hat{A}' \vee B, \hat{A} \vee B) < \frac{\varepsilon}{3}$. Then $\rho(\hat{A}',A') \leq \rho(\hat{A}',\hat{A}) + \rho(\hat{A},A) + \rho(A,A') < 3\delta$ so $\rho(\hat{A}' \vee B, A' \vee B) < \frac{\varepsilon}{2}$. Thus $\rho(A \vee B, \hat{A} \vee B) < \varepsilon$, completing the proof. □

<u>Proof of Proposition 5.2</u>: Choose ε and $A \in F^*$. Let η be such that $\hat{G} \subset G$, $\hat{G}$ a finite partition, $\rho(\hat{G},G) < \eta$ implies $\rho(\hat{G} \vee B, G \vee B) < \frac{\varepsilon}{3}$. Choose $\hat{A} \subset A$ a finite partition with $\rho(A,\hat{A}) < \zeta$ implies $\rho(G \vee B, G' \vee B) < \frac{\varepsilon}{3}$. Choose $\delta < \frac{\eta}{3}$ as in Lemma 5.4 corresponding to $\min\{\frac{\eta}{3},\xi\}$. For A' with $\rho(A,A') < \delta$, choose $\hat{A}' \subset A'$ a finite partition with $d(\hat{A},\hat{A}') < \min\{\frac{\varepsilon}{3},\xi\}$. Then $\rho(\hat{A}',\hat{A}') < \eta$ by the triangle inequality, so $\rho(A' \vee B, A' \vee B) \leq \frac{\varepsilon}{3}$. Thus $\rho(A \vee B, A' \vee B) \leq \rho(A \vee B, \hat{A} \vee B) + \rho(\hat{A} \vee B, \hat{A}' \vee B) + \rho(A' \vee B, A' \vee B) < \varepsilon$, completing the proof. □

6. A Generalization of the Information Topology

In many situations not all elements of the state space are equally relevant in decision making. For example, information is often specified as a signal consisting of "state plus noise" which is an abuse of terminology since both are part of the description of uncertainty in the world and thus must both be considered part of the state. The reason for such a distinction in practice, however, is that only the "state" term is decision-relevant. To illustrate, suppose the state space is $(\Omega^1 \times \Omega^2, F^1 \times F^2, \mu^1 \times \mu^2)$ where (Ω^1, F^1, μ^1) and (Ω^2, F^2, μ^2) are probability spaces. Suppose for any (ω^1, ω^2) all agents care only about ω^1. Then any two sub-σ-fields which generate the same conditional expectation for all F^1-measurable random variables convey the same decision-relevant information.

Another example is given by Allen (1982, Section 15). Suppose there are finitely many decision-relevant states $a_1, a_2, \ldots, a_N$. If the state space is taken to consist of these N elements, then no "partial information" is permitted. To consider information given by, say, updated probabilities or signals correlated with the above set, one may define the state space to be the unit square with Lebesgue measure. Draw N vertical strips such that the area of the i^{th} strip is the probability of a_i. Let A be the finite partition given by these strips. To say that the state of the world is decision-relevant only through $a_1, \ldots, a_N$ is equivalent to saying that all objective functions (e.g., state-dependent utility functions) are A-measurable. Thus, two sub-σ-fields should be considered equivalent if they lead to the same conditional expectation for every A-measurable random variable.

These examples suggest a generalization of the definition of information and the pointwise convergence topology. Let (Ω, F, μ) be a probability space satisfying Assumption 2.4, and $\mathcal{G}$ a sub-σ-field of $\mathcal{F}$. Suppose all objective functions are known to be $\mathcal{G}$-measurable. In that case we wish to identify all sub-σ-fields of $\mathcal{F}$ that yield identical conditional expectations on $\mathcal{G}$-measurable functions. Accordingly, define a binary relation $\tilde{_\mathcal{G}}$ on $\mathcal{F}$ by the rule: $\mathcal{B} \,\tilde{_\mathcal{G}}\, \mathcal{B}'$ iff $E[g|\mathcal{B}'] = E[g|\mathcal{B}']$ a.e. for every $g \in L^1(\Omega,\mathcal{G},\mu;\mathcal{R}) \equiv L^1(\mathcal{G};\mathcal{R})$. This is obviously an equivalence relation, so let $\mathcal{F}^*_\mathcal{G}$ be the set of equivalence classes. We do not distinguish between a sub-σ-field and its equivalence class below. Give $\mathcal{F}^*_\mathcal{G}$ the pointwise convergence (i.e., strong operator) topology as a subset of $L(L^1(\mathcal{G};\mathcal{R}))$. Thus, the map from $\mathcal{F}^*_\mathcal{G}$ to $L^1(\mathcal{G};\mathcal{R})$ defined by $\mathcal{B} \mapsto E[g|\mathcal{B}]$ is continuous for all $g \in L^1(\mathcal{G};\mathcal{R})$, the latter with the L^1-norm topology. This is a generalization of the construction in Section 2, where $\mathcal{G}$ was the sub-σ-field generated by the null sets in $\mathcal{F}$. Hence, an analogous result holds.

Proposition 6.1: If Assumption 2.4 holds, then $\mathcal{F}^*_\mathcal{G}$ is a separable metric space with metric

$$\rho_\mathcal{G}(\mathcal{B},\mathcal{B}') = \Sigma_{j=1}^{\infty} 2^{-j} \min\{\| E[g_j|\mathcal{B}] - E[g_j|\mathcal{B}']\|,1\}$$

where $\{g_j\}_{j=1}^{\infty}$ is a dense subset of $L^1(\mathcal{G};\mathcal{R})$. If the latter is closed in $L^1(\mathcal{R})$, then $\rho_\mathcal{G}$ is complete.

Proof: Imitate the proof of Proposition 2.5, using the fact that if $L^1(\mathcal{G},\mathcal{R})$ is closed, then it is complete. □

Note that all of the subsequent results in Sections 2, 3, and 5 hold for $\mathcal{F}^*_\mathcal{G}$ where random variables and state-dependent utility functions are all taken to be $\mathcal{G}$-measurable. A slight difficulty arises in discussing finite partitions in $\mathcal{F}^*_\mathcal{G}$ since a finite partition may be equivalent to a nonfinite sub-σ-field. Take such an element to be a finite partition in reading the above results.

The advantage of the new topology is that more sets are compact. In some cases, F_G^* may be compact whereas F^* is not, though the reverse situation cannot hold if $L^1(\Omega,G,\mu;R)$ is closed. In some applications, this new definition of information may correspond to some more familiar ones (e.g., probability distributions).

7. Dependence of the Topology on the Probability Space

The underlying probability space chosen determines the space of information and its topology. In economic models, however, one might expect different agents to possess different beliefs about the uncertainty they face. This presents some problems in making statements across agents or across models such as continuity of aggregate demand with respect to information. Such statements would depend on which assessement of uncertainty is chosen. In addition, the probability distribution is particularly subjective and prone to error. One would therefore hope that results about information do not depend too critically on how different events are weighted.

Clearly then it is useful to know which characteristics of the probability space determine which properties of information. The definition of a sub-σ-field depends on the state space and the σ-field of events on that space. Hence the measurable space (Ω,F) is relevant, which is misleading since two measurable spaces may be isomorphic (cf. Royden, 1968, Chapter 15). In any case, the specification of the states and events that can occur is not likely to be controversial. For the remainder of this section the space (Ω,F) is assumed to be fixed.

Most differences of opinion about uncertainty occur when discussing probabilities. One immediate way in which the probability measure affects the space of information is in the process of identifying sub-σ-fields generating the same conditional expectation (p. 9). Recall that two sub-σ-fields are considered the same if they differ by sets of measure zero. The events which have measure zero (also called null sets) are determined by the probability distribution, so one cannot expect two probability measures to give the same space of information unless they generate the same null sets. Two such probability measures are said to be <u>mutually absolutely continuous</u>. Since this is necessary for generating the same space of information, one would hope that this is sufficient for yielding the same topology on that space. Allen (1982, pp. 20-22) proved that two probability measures satisfying a stronger condition called uniform mutual absolute continuity generate the same Hausdorff topology on F^*. This condition cannot be stated only in terms of null sets but depends on how events are weighted. Mutual absolute continuity is sufficient, however, for two probability measures to generate the same pointwise convergence topology. As a result, the topological properties of information discussed in this paper and companion papers (e.g., Cotter, 1983) depend only on the state space, the σ-field of events, and the events assigned measure zero. In fact, this result extends to uniform properties of information as will be shown later.

Let μ and ν be mutually absolutely continuous probability measures on (Ω,F). Since they generate the same null sets, there is no ambiguity in the definitions of "almost everywhere", the space of information F^*, and the space of essentially bounded random variables L^∞. Unfortunately, $L^1(\Omega,F,\mu;R) \neq L^1(\Omega,F,\nu;R)$ but L^∞ is dense in both spaces. In addition, the conditional expectation operators given $B \subset F^*$ under μ and ν, denoted $E_\mu[\cdot|B]$ and $E_\nu[\cdot|B]$ respectively, are different.

Let $d\mu/d\nu$ and $d\nu/d\mu$ be the Radon-Nikodym derivatives (cf. Royden, 1968, pp. 238-240), so both are strictly positive a.e.. The following results relate the two conditional expectation operators.

<u>Lemma 7.1</u>: For a.e. bounded f and all B, $E_\mu[f|B] = E_\mu[d\nu/d\mu|B] \cdot E_\nu[f(d\mu/d\nu)|B]$ a.e.

Proof: For $B \in \mathcal{B}$, $\int_B E_\mu[d\upsilon/d\mu|\mathcal{B}] \cdot E_\upsilon[f(d\mu/d\upsilon)|\mathcal{B}]\, d\mu$

$= \int_B E_\mu[(d\upsilon/d\mu)E_\upsilon[f(d\mu/d\upsilon)|\mathcal{B}]|\mathcal{B}] \quad d\mu \quad = \quad \int_B E_\upsilon[f(d\mu/d\upsilon)|\mathcal{B}] \, d\nu \quad = \quad \int_B f d\mu$.

Since the former is $\mathcal{B}$-measurable, this completes the proof.

Corollary 7.2: $E_\upsilon[d\mu/d\upsilon|\mathcal{B}] = (E_\mu[d\upsilon/d\mu|\mathcal{B}])^{-1}$

Proof: Let $f = 1$ in Lemma 7.1.

The main result of this section can now be stated.

Proposition 7.3: Let $\mathcal{T}_p^\mu$ and $\mathcal{T}_p^\nu$ be the pointwise convergence topologies on F^* generated by μ and ν respectively. Then $\mathcal{T}_p^\mu = \mathcal{T}_p^\nu$.

Proof: Let f be an a.e. bounded random variable and $\mathcal{B}$, $\mathcal{B}' \in F^*$. Choose $\varepsilon > 0$ and let $\delta > 0$ be such that for $A \in F$ with $\nu(A) < \delta$, it follows that $\mu(A) < \varepsilon$. See Neveu (1965, p. 110) for details. Since

$\int_\Omega E_\mu[d\upsilon/d\mu|\mathcal{B}](d\mu/d\upsilon)d\nu \quad = \quad \int_\Omega E_\mu[d\upsilon/d\mu|\mathcal{B}] \, d\mu \quad = \quad 1$, the set

$A = \{\omega | E_\mu[d\upsilon/d\mu|\mathcal{B}](d\mu/d\upsilon) \leqq 1/\delta\}$ satisfies $\mu(A) \geqq 1 - \varepsilon$ by Chebyshev's inequality.

Let $\|\cdot\|_\mu$ and $\|\cdot\|_\upsilon$ denote the L^1-norms generated by μ and ν respectively. Then

$$\|E_\mu[f|\mathcal{B}] - E_\mu[f|\mathcal{B}']\|_\mu = \int_{N^c} |E_\mu[f|\mathcal{B}] - E_\mu[f|\mathcal{B}']|d\mu$$

$$+ \int_N |E_\mu[d\upsilon/d\mu \; |\mathcal{B}] \cdot E_\upsilon[f(d\mu/d\upsilon)|\mathcal{B}] - E_\mu[f|\mathcal{B}'] \cdot E_\nu[f(d\mu/d\nu)|\mathcal{B}'] \cdot (d\mu/d\nu)d\nu$$

$$\leqq 2\varepsilon\alpha + \int_N E_\mu[d\nu/d\mu|\mathcal{B}] \, |E_\nu[f(d\mu/d\nu)|\mathcal{B}] - E_\nu[f(d\mu/d\nu)|\mathcal{B}']|(d\mu/d\nu)d\nu$$

$$+ \int_N E_\nu[f(d\mu/d\nu)|\mathcal{B}] \, E_\mu[d\nu/d\mu|\mathcal{B}] - E_\mu[d\nu/d\mu|\mathcal{B}']|(d\mu/d\nu)d\nu$$

where $\alpha = \text{ess sup} |f|$. Repeated use of Lemma 7.1 and Corollary 7.2

reduces the above expression to being no greater than

$$2\varepsilon\alpha + (1/\delta)\|E_\nu[f(d\mu/d\nu)|\mathcal{B}] - E_\nu[f(d\mu/d\nu)|\mathcal{B}']\|_\nu + (\alpha/\delta)\|E_\nu[d\mu/d\nu|\mathcal{B}] - E_\nu[d\mu/d\nu|\mathcal{B}']\|_\nu$$

Now given $\eta > 0$ and $\mathcal{B} \in F^*$, choose the following $\mathcal{T}_p^\nu$-neighborhood in F^* where $\varepsilon = \eta/4\alpha$ and δ is chosen as before:

$$V^* = \{\mathcal{B}' \mid \|E_\nu[f(d\mu/d\nu)|\mathcal{B}] - E_\nu[f(d\mu/d\nu)|\mathcal{B}']\|_\nu \leq \frac{\eta\delta^*}{4} \text{ and}$$
$$\|E_\nu[d\mu/d\nu|\mathcal{B}] - E_\nu[d\nu/d\mu|\mathcal{B}']\|_\nu < \eta\delta^*/4\alpha\}$$

Then $\mathcal{B}' \in V^*$ implies that $\|E_\mu[f|\mathcal{B}] - E_\mu[f|\mathcal{B}']\|_\mu < \eta$. Hence $\mathcal{T}_p^\nu \subset \mathcal{T}_p^\mu$. Reversing the above argument completes the proof.

A question that naturally arises at this point is whether the metrics generated by μ and ν, denoted ρ_μ and ρ_ν, are uniformly equivalent. That is, given $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho_\mu(\mathcal{B},\mathcal{B}') < \delta$ implies $\rho_\nu(\mathcal{B},\mathcal{B}') < \varepsilon$ and $\rho_\nu(\mathcal{B},\mathcal{B}') < \delta$ implies $\rho_\mu(\mathcal{B},\mathcal{B}') < \varepsilon$. If the two metrics are uniformly equivalent, then their uniform structures are equivalent, meaning that properties such as uniform continuity are preserved. Though uniform equivalence is not nearly as important as equivalence, the former is still occasionally useful.

Allen (1982, pp. 20-21) showed that if μ and ν are uniformly mutually absolutely continuous then their Hausdorff metrics are uniformly equivalent.

Of course, such a question must be independent of the pointwise convergence metric chosen (see Proposition 2.5). Thus any two metrics defined under the same probability measure must be uniformly equivalent. This is proven next.

Proposition 7.4: Let μ be fixed and $\{f_j\}_{j=1}^{\infty}$ and $\{g_k\}_{k=1}^{\infty}$ be dense subsets of L^1 . Then their respective metrics as defined in Proposition 2.5 are uniformly equivalent.

Proof: Let $\varepsilon > 0$ and choose $N > 0$ such that $2^{-N+1} < \varepsilon$. Let $M > 0$ be such that for all $j \leq N$ there exists $k \leq M$ satisfying $\|f_j - g_k\| < \varepsilon/8$. Let $\delta = \varepsilon 2^{-M-1}$. Then for $\mathcal{B}$, $\mathcal{B}'$ satisfying $\rho_g(\mathcal{B},\mathcal{B}') < \delta$, for $k \leq M$, $\|E[g_k|\mathcal{B}] - E[g_k|\mathcal{B}']\| < 2^{k-M-1}\varepsilon < \varepsilon/4$

For $j \leq N$, then, $\|E[f_j|\mathcal{B}] - E[f_j|\mathcal{B}']\| + \|E[f_j - g_j|\mathcal{B}]\| + \|E[g_j|\mathcal{B}] - E[g_j|\mathcal{B}']\|$

$+ \|E[g_k - f_j|\mathcal{B}']\| < \varepsilon/2$.

Then $\rho_f(\mathcal{B},\mathcal{B}') < \sum_{j=1}^{N} 2^{-j}\varepsilon/2 + \sum_{j=N+1}^{\infty} 2^{-j} < \varepsilon/2 + 2^{-N+1} < \varepsilon$.

Reversing the argument completes the proof.

Fortunately, if μ and ν are mutually absolutely continuous, then their pointwise convergence metrics are uniformly equivalent. Hence all uniform properties of information in this topology are independent of the probability measure chosen, once the sets of measure zero are defined.

Lemma 7.5: Let $\{f_j\}_{j=1}^{\infty}$ L^{∞} be dense in $L^1(\Omega,F,\mu;R)$. Then $\{f_j(d\mu/d\nu)\}_{j=1}^{\infty}$ is dense in $L^1(\Omega,F,\nu;R)$.

Proof: Let $g \in L^1(\Omega,F,\nu;R)$. Then $\|g - f_j(d\mu/d\nu)\|_{\nu}$

$= \int_{\Omega} |g - f_j(d\mu/d\nu)|d\nu = \int_{\Omega} |g(d\nu/d\mu) - f_j|(d\mu/d\nu)d\nu = \|g(d\nu/d\mu) - f_j\|_{\mu}$

By hypothesis this completes the proof.

Proposition 7.6: If μ and ν are mutually absolutely continuous then their metrics are uniformly equivalent.

Proof: Let $\{f_j\}_{j=1}^{\infty} \subset L^{\infty}$ be dense in $L^1(\Omega,F,\mu;R)$ with $f_1 = 1$.

Then $\rho_{\nu}(\mathcal{B},\mathcal{B}') = \sum_{j=1}^{\infty} 2^{-j}\min\{\|E_{\nu}[f_j(d\mu/d\nu)|\mathcal{B}] - E_{\nu}[f_j(d\mu/d\nu)|\mathcal{B}']\|_{\nu},1\}$

is a metric generating $\mathcal{T}_p^{\nu}$ by Lemma 7.5. The result follows from the last part of the proof of Proposition 7.3.

8. Conclusion

In this paper, a new tool has been proposed for studying information as an explicit parameter of economic models and an instrument of choice. As in Allen (1982), the set of information has been defined to be a topological space depending only on the underlying probability space. In fact, all relevant properties of this space depend only on the set of states of the world, the set of events, and the set of events defined to have probability zero. The topology used here is the weakest one making the function mapping information to conditional expectation continuous for any integrable random variable. In virtually all conceivable cases this topology is a complete separable metric space.

Sequences of information converging in a set-theoretic sense (e.g., increasing sequences) converge in the metric. In addition, the set of all finite partitions of the state space is dense. The space of information is not in general compact unless the underlying probability space is purely atomic. The join operation of combining two information fields is not separately continuous. For any finite partition, however, the function mapping any information field into its join with the finite partition is continuous.

This topology is strictly weaker than the Hausdorff topology (cf. Allen, 1982) unless the probability space is purely atomic, in which case they are identical. Since the latter lacks some of the above properties and appears to be very fine, the pointwise convergence topology may prove to be more useful in economic applications. The major remaining issue to be addressed is the continuity of economic variables with respect to the information they depend on. If variables such as consumer demand are continuous with respect to the pointwise convergence metric, then the notion of closeness defined by the metric has a useful economic interpretation. In addition, consumers with a choice of information would be expected to face a continuous value function of information, which would in turn make the consumer choice problem of information well-behaved.

Another problem needing further work is the fact that this metric is not easy to compute explicitly. It would be very useful to derive simpler forms for the metric in special cases such as when the probability space is given by the unit interval with its Borel sets and Lebesgue measure.

The continuity of consumer demand and properties of the resulting choice problem of information are considered in Cotter (1983).

FOOTNOTES

1. For ease of exposition, issues of measurability of sets and functions, as well as null sets, are ignored in this section.

2. One possible source of confusion is that two independent identically distributed signals do <u>not</u> convey the same information. Hence, combining them does increase total information.

3. To see that $\mathcal{F}^*$ is not convex as a subset of $\mathcal{L}(L^1(\mathcal{R}))$, consider the following example due to James Jordan. Let the state space be the unit interval with Borel sets and Lebesgue measure. Let $\mathcal{B}_0 = \{\emptyset, \Omega\}$ and $\mathcal{B}_1 = \{[0,\frac{1}{2}), [\frac{1}{2},1], \emptyset, \Omega\}$, and $\lambda \in (0,1)$. For $f \in L^1(\mathcal{R})$ with $\int_0^{\frac{1}{2}} f(x)dx \neq \int_{\frac{1}{2}}^1 f(x)dx$, the expression $\lambda E[f|\mathcal{B}_0] + (1-\lambda)E[f|\mathcal{B}_1]$ is $\mathcal{B}_1$-measurable. Hence, this is the conditional expectation of f with respect to some sub-σ-field $\mathcal{B}_2$, $E[f|\mathcal{B}_2] = E[f|\mathcal{B}_1]$, therefore $E[f|\mathcal{B}_0] = E[f|\mathcal{B}_1]$, a contradiction.

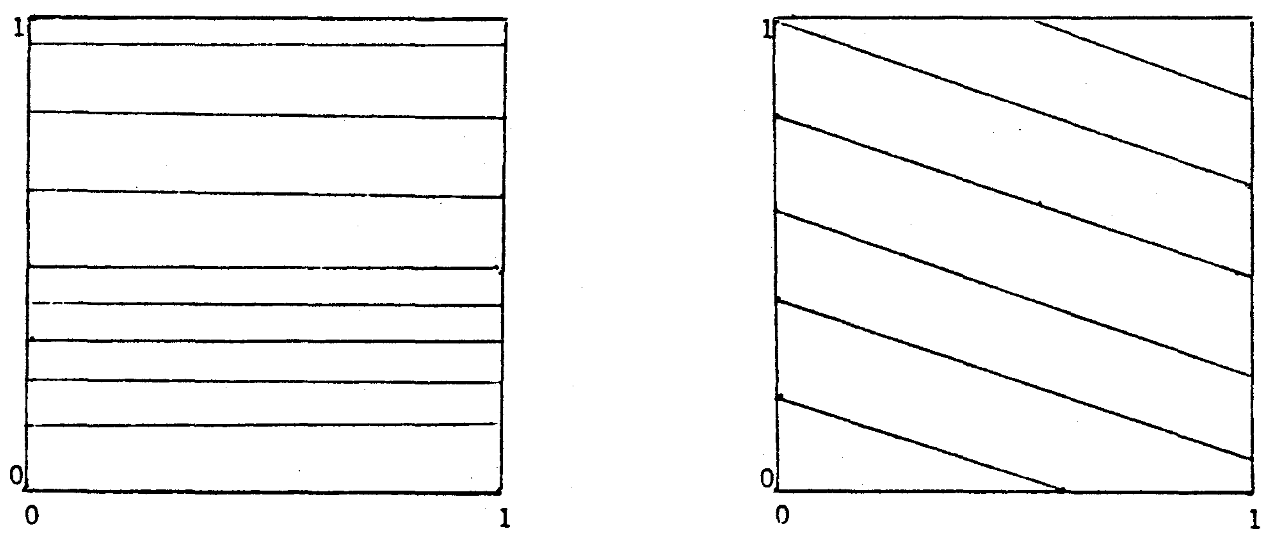
FIGURE 1

For any (x,y) in the unit square,

B_0 reveals y

B_n reveals $y + x/n$

Each partitions the unit square into strips, samples of which are shown below:



B_0 B_n

REFERENCES

Allen, B. (1982), "Neighboring Information and Distributions of Agents' Characteristics Under Uncertainty", CARESS Working Paper #82-02, University of Pennsylvania.

Bewley, T. (1972), "Existence of Equilibria in Economies with Infinitely Many Commodities", Journal of Economic Theory, 4, 514-540.

Boylan, E. (1971), "Equiconvergence of Martingales", Annals of Mathematical Statistics, 42, 552-559.

Breiman, L. (1968), Probability, Reading, Mass.: Addison-Wesley.

Chan, Y., and H. Leland (1982), "Prices and Qualities in Markets with Costly Information", Review of Economic Studies, 69, 499-516.

Chung, K. (1974), A Course in Probability Theory, 2nd. ed. New York: Academic Press.

Cotter, K. (1983), "Consumer Choice of Information With a 'Weak' Topology, University of Minnesota.

Debreu, G. (1957), Theory of Value. New York: Wiley.

Dunford, N., and J. Schwartz (1957), Linear Operators, Part I. New York: Interscience.

Fetter, H. (1977), "On the Continuity of Conditional Expectations", Journal of Mathematical Analysis and Applications, 61, 227-231.

Grossman, S., and J. Stiglitz (1980), "The Impossibility of Informationally Efficient Markets", American Economic Review, 70, 393-408.

Halmos, P. (1950), Measure Theory. Priceton: Van Nostrand.

Hildenbrand, W. (1974), Core and Equilibria of a Large Economy. Princeton: Princeton University Press.

Hirshleifer, J., and J. Riley (1979), "The Analytics of Uncertainty and Information - An Expository Survey", Journal of Economic Literature, 62, 1375-1421.

Jordan, J., and R. Radner (1982), "Rational Expectations in Microeconomic Theory: An Overview", Journal of Economic Theory, 26, 201-223.

Kudo, H. (1974), "A Note on the Strong Convergence of σ-Algebras", Annals of Probability, 2, 76-83.

Mas-Colell, A. (1975), "A Model of Equilibrium with Differentiated Commodities", Journal of Mathematical Economics, 2, 263-295.

Neveu, J. (1965), Mathematical Foundations of the Calculus of Probability (A. Feinstein, tr.). San Francisco: Holden-Day.

Neveu, J. (1975), Discrete Parameter Martingales (T. Speed, tr.). New York: American Elsevier.

Ostroy, J. (1982), "On the Existence of Walrasian Equilibrium in Large Square Economies", unpublished manuscript.

Townsend, R. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification", Journal of Economic Theory, 21, 265-293.

Verrecchia, R. (1982), "Information Acquisition in a Noisy Rational Expectations Economy", Econometrica, 50, 1415-1430.