Discussion Paper No. 653

Convergence of Information, Random Variables, and Noise

Kevin D. Cotter
Department of Economics
Northwestern University
Evanston, IL 60201

Revised August 1985

The function mapping any random variable to the information it generates is studied. With respect to convergence in probability of random variables, the information map is continuous with respect to pointwise convergence of information precisely at the set of completely revealing random variables. It is continuous on any subset of random variables to which the same smooth independent term has been added. Some relationships between convergence in distribution of perturbed random variables and convergence in distribution of conditional expectations are also examined.

Convergence of Information and Random Variables

1. Introduction

One characteristic of the methods used in the study of the economics of information is the use of many seemingly unrelated ways of specifying information. This has been made necessary by the wide variety of situations in which information is of interest as well as a lack of formal methods for studying information as an economic variable. Information does often needs to be given by some parametric form in order to generate meaningful results such as explicit solutions or comparative statics. However, similar restrictions on other economic parameters such as individual preferences and technology can be made by restricting attention to some subset of a familiar space of possible characteristics. One example is the choice of a subset of Cobb-Douglas utility functions from the space of consumer preference relations. By maintaining a relationship between concrete and abstract forms, the general theory of preference relations as well as the special properties of Cobb-Douglas functions can be used in a self-consistent manner. The distinction between those results which depend explicitly on the use of Cobb-Douglas functions and those which can be generalized is usually clear. The use of general consumer theory allows one to relate models using different specifications of preferences. In contrast, the specification of information is often in analogy with some more familiar set of objects such as the set of normal random variables, obscuring the dependence of the results on that specification. The results obtained by using a particular information specification cannot be easily related to other models. Furthermore, without general methods for solving problems of information, the parameterization of information must often be severely restricted in order to allow solution to be derived from more standard methods. This is seen by observing the large number of models in which information is restricted to either complete information or complete ignorance.

One purpose of studying information in an abstract context is to provide a framework for relating more concrete results using different information specifications. Many properties of the abstract space of information, such as continuity and compactness, relate to the topology of information being used. To this end two relatively tractible metric topologies of information have been proposed. One was first studied by Boylan (1971) which is analogous to the Hausdorff metric on closed sets. Allen (1983) examined some economic properties of the Boylan metric. In particular she showed that a consumer's conditional expected utility is jointly continuous with respect to the consumer's information and state-dependent utility function (Theorem 10.6). It follows easily that excess demand is continuous with respect to information (Corollary 10.7). The second information metric is the pointwise convergence metric, proposed and studied by Cotter (1985). The pointwise convergence topology is weaker than the Boylan topology (Proposition 5.1) but has the same property of joint continuity of conditional expected utility (Theorem 8.7). Both metrics are discussed in more detail in Section 2.

The next step is to relate these metrics of information to the topological properties of spaces from which commonly used specifications of information are drawn. Particular attention should be devoted to the space of random variables since a huge bulk of information specifications involves use of that space. Every random variable generates an information field. This defines an information map which is a function from the space of random variables to the abstract space of information. To relate topologies (and hence convergence concepts) of random variables to topologies of information, the continuity properties of the information map must be established. Since the information map is not one-to-one, no discussion of any map in the opposite direction can be made. The importance of such results go beyond relating the abstract properties of information to the properties of random variables. In many cases, such as in rational expectations equilibrium, a random variable plays two different roles, only one of which is to convey information. In that case the properties of the price function become directly relevant to its properties as an information field. It is immediately apparent that the information map is not continuous unless the space of random variables is given the indiscrete (trivial) topology. Two random variables on any given probability space may be arbitrarily close to each other yet one can be completely revealing while the other is completely uninformative. In Section 2 I state and prove that in a precise sense, the information map is lowersemicontinuous with respect to the pointwise convergence metric. Intuitively, this means that the sequence of information generated by a convergent sequence of random variables either converges or "collapses" in the limit. The main result of Section 2 is that the information map is continuous at a given random variable with respect to the pointwise convergence metric of information and the metric of convergence in probability of random variables if and only if that random variable completely reveals the true state of nature. For the Boylan metric, matters are even worse; then there are completely revealing random variables at which the information map is discontinuous.

A natural question is whether there exists an economically useful subset of random variables on which the information map restricted to that subset is continuous. One answer is suggested by the use of smoothing methods in which behavior resulting from informative observations such as prices is made continuous by perturbing that observation with a smooth noise term. In Section 3 I show that on the set of random variables to which the same independent noise term has been added, the information map is continuous with respect to convergence in probability of random variables and pointwise convergence of information. The noise term is assumed to have a bounded and piecewise continuous density function. This theorem does not hold for the Boylan metric. This result applies to the case where the set of decision-relevant states is a subset of the real numbers with a probability distribution given by a well-behaved density function, and information is restricted to all possible perturbed observations of the state by independent noise terms. In this case another result more useful to the study of optimal choice of information holds. The main theorem of Section 4 is that when a sequence of extended-valued error terms converges in distribution, the corresponding sequence of the conditional expectation of any function of the decision-relevant state converges in distribution. Continuity of conditional expected utility as well as the value of information derived from any maximization problem follows. Since the set of probability distributions on a compact set is weakly compact, the value of information has a maximum whenever the set of possible errors takes values in a compact set, including the extended real line.

2. Lowersemicontinuity

Uncertainty is given as a probability space $(\Omega, \mathcal{F}, \mu)$ where Ω is the set of possible states of nature, $\mathcal{F}$ the σ-field of measurable subsets (events) of Ω, and μ a probability measure on $(\Omega, \mathcal{F})$. Information is defined as in Boylan (1971), so the space of information $\mathcal{F}^*$ is the set of all sub-σ-fields of $\mathcal{F}$ modulo null sets. Following Cotter (1985, p. 7), no ambiguity will arise by interchanging elements of $\mathcal{F}^*$ with sub-σ-fields of $\mathcal{F}$.

Let V be the set of all real-valued measurable functions on $(\Omega, \mathcal{F})$, with the metric $\theta(f,g) = \inf \{\varepsilon | \mu(|f - g| > \varepsilon) < \varepsilon\}$. Then convergence in θ is equivalent to convergence in probability. Any $f \in V$ generates a sub-σ-field of $\mathcal{F}$, which is the smallest σ containing every event of the form $\{\omega | f(\omega) < a\}$ for every $a \in \mathbb{R}$. Let $\sigma : V \to \mathcal{F}^*$ be the resulting function.

The fundamental discontinuity of σ with respect to any topology on $\mathcal{F}^*$ other than the indiscrete topology is a primary source of difficulty in the study of the economics of information. The problem is a possible "collapse" of information in the limit, which is formalized by Theorem 2.2. The imposition of stronger nontrivial conditions on the convergence of a sequence of random variables such as convergence of their density functions is of no value, as the following example illustrates.

Example 2.1: Let $T = [0,1)$, $\mathcal{T}$ be the Borel sets of T, and λ be Lebesgue measure on $(T, \mathcal{T})$. For each n define $f_n : T \to \mathbb{R}$ by

$$f_n(t) = \begin{cases} t + j2^{-n} & \text{if } j2^{-n} \leq t < (j+1)2^{-n} \leq 1/2 \\ 2 - j2^{-n} - t & \text{if } 1/2 \leq j2^{-n} \leq t < (j+1)2^{-n} \end{cases}$$

For each n, f_n is injective and therefore $\sigma(f_n) = \mathcal{T}$. In addition, f_n has a uniform (0,1) distribution. Let $f : T \to \mathbb{R}$ be defined by $f(t) = t$ for $t \in [0,1/2)$, and $f(t) = 2(1-t)$ for $t \in [1/2, 1)$. Then for each n and t, $|f_n(t) - f(t)| < 2^{-n}$, so $\{f_n\}$ converges to f in L^∞. In addition, f has a uniform (0,1) distribution, so its density is identical to the density of each f_n. However, f is exactly two-to-one, so $\sigma(f) \neq \mathcal{T}$.

In the above example, as in every example of a sequence of random variables with discontinuous information, the limit random variable has "less" information than the limit of the information generated by each term of the sequence. To formalize this statement, the concepts of lower limit and upper limit of a sequence of information are useful. Let $L^1 = \{f \in V \mid \|f\| \equiv E[|f|]$ is finite$\}$ with norm $\|\cdot\|$. Kudō (1974, Theorem 3.1) showed that for any $\mathcal{G}$, $\mathcal{H}$ sub-σ-fields of $\mathcal{F}$, $\|E[f|\mathcal{G}]\| \leq \|E[f|\mathcal{H}]\|$ for every $f \in L^1$ if and only if $\mathcal{G} \subset \mathcal{H}$. Following Kudō, define the <u>lower limit</u> of a sequence of sub-σ-fields $\{\mathcal{G}_n\}$, denoted μ-liminf $\mathcal{G}_n$, to be the largest (modulo null sets) sub-σ-field $\mathcal{G}_o$ satisfying $\liminf_{n\to\infty} \|E[f|\mathcal{G}_n]\| \geq \|E[f|\mathcal{G}_o]\|$ for every $f \in L^1$. The <u>upper limit</u>, denoted μ-limsup $\mathcal{G}_n$, is defined to be the smallest sub-σ-field $\mathcal{G}^o$ satisfying $\limsup_{n\to\infty} \|E[f|\mathcal{G}_n]\| \leq \|E[f|\mathcal{G}^o]\|$ for every $f \in L^1$. Kudō (1974) showed that μ-liminf $\mathcal{G}_n = \{A \in \mathcal{F} \mid \lim_{n\to\infty} \inf_{B\in\mathcal{G}_n} \mu(A\Delta B) = 0\}$ where $A\Delta B = (A\cap B^c) \cup (A^c \cap B)$ (Theorem 3.2), and that μ-limsup $\mathcal{G}_n$ always exists with μ-liminf $\mathcal{G}_n \subset$ μ-limsup $\mathcal{G}_n$ (Theorem 3.4).

<u>Theorem 2.2</u>: Let $\{g_n\}$ be a sequence in V converging in probability to g. Then $\sigma(g) \subset$ μ-liminf $\sigma(g_n)$.

<u>Proof</u>: It suffices to show that for every $a \in \mathbb{R}$, $A \equiv \{\omega | g(\omega) < a\} \in$ μ-liminf $\sigma(g_n)$. Given $\varepsilon > 0$, choose $\delta > 0$ such that $\mu\{\omega | a-\delta < g(\omega) < a\} < \varepsilon/2$, and choose N such that for every $n \geq N$, $\mu\{\omega \mid |g_n(\omega) - g(\omega)| < \delta/2\} > 1 - \varepsilon/2$. Fix such an n, and let $W = \{\omega \mid |g_n(\omega) - g(\omega)| < \delta/2\}$. Let $B = \{\omega \mid g_n(\omega) \leq a - \delta/2\}$. Then for $\omega \in W \cap B^c$, $g(\omega) > a-\delta$, and for $\omega \in W \cap B$, $g(\omega) < a$. Therefore,

$$\mu(A\Delta B) \leq \mu(A\cap B^c\cap W) + \mu(A^c\cap B\cap W) + \mu(W^c) < \varepsilon/2 + 0 + \varepsilon/2 = \varepsilon$$

which completes the proof.

Consider the two topologies on $\mathcal{F}^*$ which have been proposed. The Boylan metric d, proposed by Boylan (1971) and used by Allen (1983), is defined to be

$$d(\mathcal{G},\mathcal{H}) = \sup_{G\in\mathcal{G}} \inf_{H\in\mathcal{H}} \mu(G\Delta H) + \sup_{H\in\mathcal{H}} \inf_{G\in\mathcal{G}} \mu(G\Delta H).$$

Allen (1983, Fact 9.3) showed that a sequence $\{\mathcal{G}_n\}$ converges to $\mathcal{G}$ in this metric if and only if

$$\lim_{n\to\infty} \sup_{\{f\in L^1 \mid |f(w)| \leq 1 \text{ a.e.}\}} \| E[f|\mathcal{G}_n] - E[f|\mathcal{G}]\| = 0.$$

The other principal metric on $\mathcal{F}^*$ is the pointwise convergence metric ρ, proposed by Cotter (1985). If L^1 is separable, this metric can be defined to be, where $\{f_j\}$ is a dense subset of L^1,

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

so $\{\mathcal{G}_n\}$ converges to $\mathcal{G}$ pointwise if and only if $\lim_{n\to\infty} \| E[f|\mathcal{G}_n] - E[f|\mathcal{G}]\| = 0$ for every $f \in L^1$. By Proposition 2.2 of Cotter (1985), pointwise convergence is equivalent to the strong convergence of Kudō (1974), i.e., $\{\mathcal{G}_n\}$ converges to $\mathcal{G}$ pointwise if and only if $\{P[A|\mathcal{G}_n]\}$ converges in probability to $P[A|\mathcal{G}]$ for every $A \in \mathcal{F}$. Therefore $\{\mathcal{G}_n\}$ converges pointwise if and only if $\mu\text{-}\limsup \mathcal{G}_n \subset \mu\text{-}\liminf \mathcal{G}_n$, in which case $\lim_{n\to\infty} \mathcal{G}_n = \mu\text{-}\liminf \mathcal{G}_n$. The next two results follow from Theorem 2.2.

Corollary 2.3: Let $\{g_n\}$ be a sequence in V converging in probability to g. If $\sigma(g_n) \subset \sigma(g)$ for every n, then $\{\sigma(g_n)\}$ converges pointwise to $\sigma(g)$.

Corollary 2.4: σ is continuous with respect to ρ at every $g \in V$ which is completely revealing, i.e., $\sigma(g) = \mathcal{F}$.

The striking fact about Corollary 2.4 is that σ is continuous with respect to the pointwise convergence metric only at every completely revealing random variable. At any other random variable σ is discontinuous, as shown by the next set of results.

Lemma 2.5: For any $f \in V$ and $\varepsilon > 0$, there exists $g \in V$ such that $\theta(f,g) < \varepsilon$ and $\rho(\sigma(g), \mathcal{F}) < \varepsilon$.

Proof: There exists some $f' \in L^1$ such that $\theta(f,f') < \varepsilon/3$ (Royden, 1968, Proposition 11.26). By Theorem 3.1 of Cotter (1985), there exists a finite disjoint partition $\{B_1, B_2, \ldots, B_N\}$ of Ω such that $\rho(\mathcal{B}, \mathcal{F}) < \varepsilon$ and $\| E[f' | \mathcal{B}] - f' \| < \varepsilon^2/9$. Then by Chebyshev's inequality, $\theta(E[f' | \mathcal{B}], f') < \varepsilon/3$. Finally, let g be obtained from $E[f' | \mathcal{B}]$ by $g = \sum_{i=1}^{N} b_i I_{B_i}$ where $b_1, \ldots, b_N$ are distinct and for each i, $\left| b_i - \frac{1}{\mu(B_i)} \int_{B_i} f'(\omega) d\mu(\omega) \right| < \varepsilon/3$. Then $\sigma(g) =$ and $\theta(g, E[f' | \mathcal{B}]) < \varepsilon/3$, so $\theta(g,f) < \varepsilon$, completing the proof.

Corollary 2.6: σ is continuous with respect to ρ at $f \in V$ if and only if f is completely revealing.

Proof: Follows from Corollary 2.4 and Lemma 2.5.

As discouraging as Corollary 2.6 may be, matters are even worse when $\mathcal{F}^*$ is given the Boylan metric. In that case σ is not even continuous at some completely revealing random variables. The following example is an analogue of Example 8.1 of Allen (1983).

Example 2.7: Consider the probability space $(T, \mathcal{T}, \lambda)$ defined in Example 2.1. For each n define $g_n : T \to \mathbb{R}$ to be $g_n(t) = j2^{-n}$ for $j2^{-n} \leq t < (j+1)2^{-n}$ and $j=0, 1, \ldots, 2^n - 1$. In other words, $g_n(t)$ is the first n places of the binary expansion of t. Then $\{g_n\}$ converges in probability (in fact, in L^∞) to the identity function, so the hypotheses of Corollary 2.4 are satisfied. However, it is easy to verify that $d(\sigma(g_n), \sigma(g_{n+1})) = 1/2$ for any n, so $\{\sigma(g_n)\}$ does not converge in the Boylan metric.

3. Continuity with a Fixed Term

In light of Corollary 2.6 and Example 2.7, the best possible result regarding continuity of the information map σ would be continuity of σ on some useful subset of the space of random variables. One possibility is suggested by the use of smoothing operations on random variables when the discontinuity of σ is a problem. The best known use of smoothing is in demonstrating existence of rational expectations equilibrium. The excess demand of a trader with a state-dependent utility function is continuous with respect to the observed nonstochastic price and the trader's information with respect to the pointwise convergence metric of information (Cotter, 1985, Theorem 8.5; see also Allen, 1983, Corollary 10.3 for a proof using the Boylan metric). If the price is stochastic but the trader does not infer from the price, continuity can be obtained by use of Chebyshev's inequality. If the information map σ were to be continuous, then the map from price to the trader's combined information (initial information and price information) would be continuous with respect to the Boylan metric, due to Allen's (1983) Lemma 14.1. Combined information would also be continuous with respect to the pointwise convergence metric if the trader's initial information were a finite partition (Cotter, 1985, Theorem 6.2). In either case excess demand would be a continuous function of informative price. The only significant remaining problem in establishing existence of rational expectations would be compactness of the price space.

A stylized fact is that when the observation of the price function is perturbed by a smooth independent noise term, continuity of excess demand with respect to price can be restored. This fact has been used by Allen (1984a, 1984b) to establish existence of approximate rational expectations, with the approximation due to the fact that traders are acting on perturbed rather than market-clearing prices. A related method was used by Anderson and Sonnenschein (1982) where traders' models of the price-state relationship were subjected to a convolution with a smooth density function. These results suggest that by restricting the space of random variables to those which have been perturbed by the same independent smooth noise term, the information map σ would become continuous on that subset.

Interestingly enough, there is a dual case to the one discussed above which is perhaps of greater practical importance. Suppose the set of decision relevant states is a subset of the real line with a density function. A decision-relevant subset of information is the set of observations of the true state which have been perturbed by all possible independent noise terms. Note that the error terms are also determined by the actual state of the world, and the space of informatin must be defined accordingly. Hence the definition of information allows for information about the error term (or the states which determined that error). This case is therefore identical to the one discussed above with the noise and underlying state terms interchanged. The continuity result hypothesized in the previous paragraph would then imply that the map from noise terms to information about both the error and the decision-relevant state is continuous.

Both of the above cases can be modelled by the probability space $(\Omega^1 \times \Omega^2, \mathcal{F}^1 \times \mathcal{F}^2, \mu^1 \times \mu^2)$ where $\Omega^2 = \mathbb{R}$ and $\mathcal{F}^2$ is the σ-field of Borel sets on $\mathbb{R}$. Let $V^* = \{f \in V \mid$ for some Borel-measurable $f^1 : \Omega^1 \to \mathbb{R}$, $f(\omega^1, \omega^2) = f^1(\omega^1) + \omega^2$ for a.e. $(\omega^1, \omega^2)\}$. In the case of rational expectations with price observations perturbed by a fixed noise term, utility would be a function of Ω^1, f^1 would be the market price function, and ω^2 the perturbation. For the case of variable noise terms, ω^2 would be the decision-relevant state and f^1 would be the noise term.

As in Section 2, V^* is given the metric $\theta(f,f') = \inf\{\varepsilon \geq 0 \mid \mu[\omega \mid |f(\omega) - f'(\omega)| > \varepsilon] < \varepsilon\}$. Since clearly $\theta(f,f') = \inf\{\varepsilon \geq 0 \mid \mu^1[\omega^1 \mid |f^1(\omega^1) - f^{1'}(\omega^1)| > \varepsilon] < \varepsilon\}$ V^* is isomorphic to the space of random variables on $(\Omega^1, \mathcal{F}^1, \mu^1)$, denoted V^1. Let $\sigma^* : V^* \to \mathcal{F}^*$ be the information map restricted to V^*. The following assumption provides sufficient smoothness of μ^2 to obtain continuity of σ^* with respect to the pointwise convergence metric on $\mathcal{F}^*$.

<u>Assumption 3.1</u>: The measure μ^2 is absolutely continuous with respect to Lebesgue measure, and its Radon-Nikodym derivative (density) ϕ can be chosen to be bounded and continuous on $\mathbb{R}$ except at possibly finitely many points.

Then any $f \in V^*$ is absolutely continuous (with respect to Lebesgue measure) with density function $\mu*\phi(t) = \int_{\Omega^1} \phi(t - f^1(\nu))d\mu^1(\nu)$. For any $A^1 \in \mathcal{F}^1$ any interval $A^2 = (-\infty, a)$ for some a, define $A = A^1 \times A^2$ and $\mu*\phi_A : \mathbb{R} \to \mathbb{R}$ to be $\mu*\phi_A(t) = \int_{A^1} I_{A^2}(t - f^1(\nu))\phi(t - f^1(\nu))d\mu^1(\nu)$. The following lemma is of interest in its own right.

<u>Lemma 3.2</u>: $P[A|\sigma(f)] = \mu*\phi_A(f)/\mu*\phi(f)$ a.e..

<u>Proof</u>: By the definition of conditional expectation it suffices to show that for any $t_o \in \mathbb{R}$, $\mu(A\cap G) = \int_G [\mu*\phi_A(f)/\mu*\phi(f)]d\mu$ where $G = f^{-1}[(-\infty, t_o)]$. Note that by a change of variables,

$$\int_G [\mu*\phi_A(f)/\mu*\phi(f)]d\mu = \int_{-\infty}^{t_o} \int_{A^1} I_{A^2}(t-f^1(\nu))\phi(t-f^1(\nu))d\mu^1(\nu)dt$$

$$= \int_{A^1} [\int_{-\infty}^{t_o-f^1(\nu)} I_{A^2}(u)\phi(u)du]d\mu^1(\nu)$$

$$= \int_{A^1} \Phi[\min\{a, t_o-f^1(\nu)\}]d\mu^1(\nu)$$

$$= \int_{A^1\cap B} \Phi(t_o-f^1(\nu))d\mu^1(\nu) + \mu^1(A^1\cap B^c)\mu^2(A^2)$$

where Φ is the distribution function of μ^2 and $B = \{\nu \in \Omega^1 | t_o-f^1(\nu) \leq a\}$. The first term may be interpreted as a convolution of ω^2 with f^1 restricted to $A^1 \cap B$, so

$$\int_{A^1 \cap B} \Phi(t_o - f^1(\nu))d\mu^1(\nu) = \mu\{[(A^1 \cap B) \times \mathbb{R}] \cap G\}$$

$$= \mu\{[(A^1 \cap B) \times A^2] \cap G\}$$

since $\omega^1 \in B$ and $(\omega^1, \omega^2) \in G$ implies $\omega^2 \leq a$. In addition, $\mu^1(A^1 \cap B^c) \cdot \mu^2(A^2) = \mu\{(A^1 \cap B^c) \times A^2\} = \mu\{[(A^1 \cap B^c) \times A^2] \cap G\}$ since $\omega^1 \in B^c$ and $\omega^2 \leq a$ implies $(\omega^1, \omega^2) \in G$. Since $\mu\{[(A^1 \cap B) \times A^2] \cap G\} + \mu\{[(A^1 \cap B^c) \times A^2] \cap G\} = \mu\{[A^1 \times A^2] \cap G\}$, the proof is complete.

The central result can now be proved.

Theorem 3.3: Under Assumption 3.1, σ^* is continuous with respect to the pointwise convergence metric on $\mathcal{F}^*$.

Proof: Let $f \in V^*$, $A^1 \in \mathcal{F}$, $a \in \mathbb{R}$, and $A = A^1 \times (-\infty, a)$. Let $\phi_A = I_{(-\infty,a)} \cdot \phi$, and $a_1, a_2, \ldots, a_I$ the points of discontinuity of ϕ_A, with $a_1 < a_2 < \ldots < a_I$. For ease of notation let $a_o = -\infty$, $a_{I+1} = +\infty$. Let $M = \sup \phi$.

Let Ψ_A be the distribution function of $f^1|_{A^1}$ with atoms (points of jump) $\{b_j\}$ (there can be at most countably many atoms), and write $\Psi_A = \Psi_c + \Psi_d$ where Ψ_c is continuous and Ψ_d is discrete.

Fix $\varepsilon > 0$ and choose $K > 0$ and $G \in \mathcal{F}$ with $\mu(G) > 1 - \varepsilon/8$ such that for $\omega \in G$, $\mu^*\phi(f(\omega)) > K$. Choose $\delta > 0$ with $\delta < \varepsilon/8$ such that for any nonnegative reals x,y,z,w with $y > K$, $|x-z| < \delta$, and $|y-w| < \delta$, it follows that $|(x/y) - (z/w)| < \varepsilon/4$. There exists J such that $\mu^1\{\omega^1 \in \Omega^1 | f^1(\omega^1) = b_j$ for some $j > J\} < \delta/(8MI)$. Since Ψ_c is uniformly continuous (being nondecreasing) and f is absolutely continuous, there exists $\tau > 0$ with $\tau < (a_{i+1} - a_i)/2$ for $i=1, 2, \ldots, I-1$ such that if $F = \{\omega \in \Omega \mid |f(\omega) - a_i - b_j| \geq 3\tau$ for every $i=1, 2, \ldots, I$ and $j=1, 2, \ldots, J\}$, then $\mu(F) > 1 - \delta$, and $|\Psi_c(x) - \Psi_c(x')| < \delta/(8MI)$ for every $x, x' \in \mathbb{R}$ with $|x - x'| < 3\tau$. On each interval $[a_i + \tau, a_{i+1} - \tau]$ and on $(-\infty, a_1 - \tau]$ and $[a_I + \tau, \infty)$, ϕ_A is uniformly continuous. Then there exists $\eta > 0$ with $\eta < \tau$ such that for every $x, x' \in \mathbb{R}$ with $|x - x'| < \eta$ and $a_i + \tau < x, x' < a_{i+1} - \tau$ for some $i = 0, 1, \ldots, I$, then $|\phi_A(x) - \phi_A(x')| < \delta/(2I)$. Then for any $t, t' \in \mathbb{R}$ with $|t - t'| < \eta$ and $t \in f(F)$, letting $\underline{t} = \min\{t, t'\}$, $\bar{t} = \max\{t, t'\}$,

$$|\mu*\phi_A(t) - \mu*\phi_A(t')| \leq \sum_{i=0}^{I} \int_{\bar{t}-a_{i+1}+\tau}^{\underline{t}-a_i-\tau} |\phi_A(t-x) - \phi_A(t'-x)|\, d\Psi_A(x)$$

$$+ \sum_{i=1}^{I} \int_{\underline{t}-a_i-\tau}^{\bar{t}-a_i+\tau} |\phi_A(t-x) - \phi_A(t'-x)|\, d\Psi_A(x)$$

$$\leq \delta/2 + 2M \sum_{i=1}^{I} [\Psi_c(t'-a_i+\tau) - \Psi_c(t-a_i-\tau)]$$

$$+ 2M \sum_{i=1}^{I} [\Psi_d(t'-a_i+\tau) - \Psi_d(t-a_i-\tau)] \qquad \leq \delta/2 + \delta/4 + \delta/4 = \delta$$

since $t' - a_i + \tau$ and $t - a_i - \tau$ do not lie on opposite sides of any b_j for $j=1, \ldots, J$. Since the set of points of discontinuity of ϕ is contained in the set of points of discontinuity of ϕ_A, for the same t,t' as above, $|\mu*\phi(t) - \mu*\phi(t')| < \delta$.

Let $f' \in V^*$ with $\theta(f, f') < \eta$, so for some $H \in \mathcal{F}^1$ with $\mu^1(H) > 1 - \eta$, it follows that for $\omega \in H$, $|f^1(\omega^1) - f^{1'}(\omega^1)| < \eta$. For $\omega \in F \cap G \cap H$, it follows by the above that $\mu^*\phi(f(\omega)) > K$, $|\mu^*\phi_A(f(\omega)) - \mu^*\phi_A(f'(\omega))| < \delta$, and $|\mu^*\phi(f(\omega)) - \mu^*\phi_A(f'(\omega))| < \delta$. Therefore by Lemma 3.1, $|P[A|\sigma(f)] - P[A|\sigma(f')]| < \varepsilon/4$. Therefore

$$\|P[A|\sigma(f)] - P[A|\sigma(f')]\| \leq \varepsilon/4 + 2\mu(F^c) + 2\mu(G^c) + 2\mu^1(H^c) < \varepsilon$$

so since all sets of the form $A^1 \times (-\infty, a)$ generate $\mathcal{F}$, the proof is complete by Proposition 2.2 of Cotter (1985).

Note that Theorem 3.3 fails when $\mathcal{F}^*$ is given the Boylan metric. This observation was made by Allen (1983, Section 14). The following example is also Example 6.1 of Cotter (1985).

<u>Example 3.4</u>: Let $(\Omega^1, \mathcal{F}^1, \mu^1) = (\Omega^2, \mathcal{F}^2, \mu^2) = (T, \mathcal{T}, \lambda)$ (see Example 2.1). For each n define $f_n \in V^*$ to be $f_n(\omega^1, \omega^2) = \omega^1/n + \omega^2$. Then $\{f_n\}$ converges in probability to f, where $f(\omega^1, \omega^2) = \omega^2$. All of the hypotheses of Theorem 3 are satisfied. However, as shown in Cotter (1985, Example 6.1), $\{\sigma(f_n)\}$ does not converge to $\sigma(f)$ in the Boylan metric, since for any n the join of $\sigma(f_n)$ and $\sigma(f)$, $\sigma(f_n) \vee \sigma(f)$, equals $\mathcal{F}$, while $\sigma(f) \vee \sigma(f) = \sigma(f) = \{\emptyset, \Omega^1\} \times \mathcal{F}^2$. Recall that the join operation is jointly continuous in the Boylan metric (Allen, 1983, Lemma 14.1).

Note in fact that $\{f_n\}$ converges to f in L^∞, and that this example can be modified to allow Ω^1 and Ω^2, as well as the supports of μ^1 and μ^2, to equal $\mathbb{R}$.

Note also that $\{f_n^1\}$ need not be chosen to converge to 0. There seems to be no way to "fix up" Assumption 3.1 to obtain continuity of σ^* with respect to the Boylan metric.

4. Convergence in distribution of expectations

The results of Section 3 are not very useful for studying problems of the optimal choice of information. The problem is that with respect to the topology of convergence in probability, compact subsets of V^1 are very small unless $\mathcal{F}^1$ is finite. Weaker topologies on V^1 such as the weak (L^1,L^∞) topology are not likely to lead to continuity results analogous to Theorem 3.3 since weak convergence of a sequence of random variables does not imply convergence of a function (such as the absolute value) of that sequence.

One solution to the compactness problem is possible when the state-dependent objective function is measurable with respect to $\mathcal{F}^2$, which is the case of a fixed real-valued decision-relevant state and noise terms in V^1. In this case the set of error terms can be expanded to cover signals which are sometimes completely uninformative by allowing error terms to be infinite with positive probability. Let $\bar{V}$ be the set of all random variables on $\bar{\mathbb{R}}$, the one-point compactification of $\mathbb{R}$. Define $L^1(\mathcal{F}^2)$ to be the set of integrable $\mathcal{F}^2$-measurable random variables. Note that for any two error terms in $\bar{V}$ which have the same image measure on $\bar{\mathbb{R}}$, the corresponding conditional expectations of any random variable in $L^1(\mathcal{F}^2)$ as defined in Section 3 have the same image measure on $\mathbb{R}$. In that case the value of information derived from an optimization problem under uncertainty depends only on the probability distribution of the error term, so the distribution of conditional expectation and therefore the value of information can be taken as well-defined functions of error probability distributions on $\bar{\mathbb{R}}$. The main result of Section 4 is that both functions are continuous with respect to weak convergence of distributions. Since the set of probability distributions on a compact set is compact with respect to the weak topology, these results can be used to show that in a wide variety of situations in which all error terms are restricted to taking values in a compact set (including the extended real line) the value of information in a maximization problem with an $\mathcal{F}^2$-measurable utility function has a maximum on that subset.

The specification of information by the distribution of error terms does have some drawbacks. In particular, comparisons between agents cannot be made by specifying only marginal distributions of errors, and the addition of random variables is not continuous with respect to convergence in distribution. Another problem is that the join of two information fields is no longer well-defined.

Random variables in $L^1(\mathcal{F}^2)$ may be written with no confusion as Borel-measurable functions on $\mathbb{R}$. Assumption 3.1 is taken to hold throughout, and as in Section 3, let $a_1, a_2, \ldots, a_I$ be the points of discontinuity of the density function ϕ of the decision-relevant space. Note that the set of bounded uniformly continuous functions on $\mathbb{R}$, denoted C_U, is dense in $L^1(\mathcal{F}^2)$. Let $C_0 = \{h \in C_U : \lim_{|x| \to \infty} h(x) = 0\}$, also used to denote those continuous functions h on $\bar{\mathbb{R}}$ for which $h(\infty) = 0$.

Since elements of $\bar{V}$ are interpreted here as error terms, let e be an element of $\bar{V}$. Recall that a sequence of random variables on $\bar{\mathbb{R}}$ converges in distribution if their image measures on $\bar{\mathbb{R}}$ converge weakly. Equivalently, $\{e_n\}$ converges in distribution to e if and only if for every $h \in C_0$, $\lim_{n \to \infty} |E[h(e_n) - h(e)]| = 0$. Then f_n, f are defined in a way analogous to the definition in Section 3: $f_n(\omega^1, \omega^2) = f_n^{\,1}(\omega^1) + \omega^2$ and $f(\omega^1, \omega^2) = f^1(\omega^1) + \omega^2$.

It is clear that in general, if e, $e' \in V^1$ have the same distribution, then $E[g|\sigma(f)]$ and $E[g|\sigma(f')]$ do not have the same distribution for all $g \in L^1$. However, if $g \in L^1(\mathcal{F}^2)$ then $E[g|\sigma(f)]$ and $E[g|\sigma(f')]$ do have the same distribution (this follows from Theorem 4.1 below), so the possibility exists that $\{E[g|\sigma(f_n)]\}$ converges to $E[g|\sigma(f)]$ in distribution. This is indeed the case.

Theorem 4.1: Suppose Assumption 3.1 holds, and let $\{e_n\}$ be a sequence in $\bar{V}$ converging to $e \in \bar{V}$ in distribution. Define f_n, f as above. Then for every $g \in L^1(\mathcal{F}^2)$, $\{E[g|\sigma(f_n)]\}$ converges to $E[g|\sigma(f)]$ in distribution.

Proof: Let Ψ_n and Ψ be the subdistribution functions on $\mathbb{R}$ of $f_n^{\,1}$ and f^1 respectively. The density function of f_n restricted to $\mathbb{R}$ is $\Psi_n * \phi$, where $\Psi_n * \phi(t) = \int_{\mathbb{R}} \phi(t-x)d\Psi_n(x)$, and the density of f restricted to $\mathbb{R}$ is $\Psi * \phi$. Given $\varepsilon > 0$ there exists $\hat{\phi} \in C_U$ such that $|\hat{\phi}(z) - \phi(z)| < \varepsilon/3$ for all $z \neq a_i$ for any i. Let $t_o \in \mathbb{R}$ be such that $t_o \neq a_i + b$ for any i and any b a point of discontinuity of any Ψ_n or Ψ.
Let $S = \{a_1 + t_o, a_2 + t_o, \ldots, a_I + t_o\}$. Then for each n,

$$|\Psi_n * \phi(t_o) - \Psi_n * \hat{\phi}(t_o)| \leq \int_S |\phi(t_o - x) - \hat{\phi}(t_o - x)|d\Psi_n(x)$$

$$+ \int_{S^c} |\phi(t_o - x) - \hat{\phi}(t_o - x)|d\Psi_n(x) < 0 + \varepsilon/3 = \varepsilon/3$$

since no element of S is an atom of Ψ_n. Similarly, $|\Psi * \phi(t_o) - \Psi * \hat{\phi}(t_o)| < \varepsilon/3$. By the definition of weak convergence of measures, there exists N such that for all $n \geq N$, $|\Psi_n * \hat{\phi}(t_o) - \Psi * \hat{\phi}(t_o)| < \varepsilon/3$. It follows by the triangle inequality that $|\Psi_n * \phi(t_o) - \Psi * \phi(t_o)| < \varepsilon$, so $\{\Psi_n * \phi\}$ converges to $\Psi * \phi$ almost everywhere with respect to Lebesgue measure. Since

$\int_{\mathbb{R}}\Psi_n * \phi(t)dt = \int_{\mathbb{R}}\Psi * \phi(t)dt \leqslant 1$ for each n, it follows by Exercise 12, p. 144 of Wheeden and Zygmund (1977) that $\lim_{n\to\infty} \int_{\mathbb{R}}|\Psi_n * \phi(t) - \Psi * \phi(t)|dt = 0$, so by their Exercise 9, p. 85, $\{\Psi_n * \phi\}$ converges to $\Psi * \phi$ in Lebesgue measure. Then by Exercise 10, p. 100 of Chung (1974), for every bounded Borel measurable function F on $\mathbb{R}$, $\{E[F(f_n)]\}$ converges to $E[F(f)]$.

Let $g' \in C_U$, and define $\hat{g}_n: \mathbb{R} \to \mathbb{R}$ to be $\hat{g}_n(t) = \Psi_n * (g'\phi)(t)/\Psi_n * \phi(t)$, and $\hat{g}: \mathbb{R} \to \mathbb{R}$ to be $\hat{g}(t) = \Psi * (g'\phi)(t)/\Psi * \phi(t)$. Extend $\hat{g}_n$, $\hat{g}$ to $\bar{\mathbb{R}}$ by defining $\hat{g}_n(\infty) = \hat{g}(\infty) = E[g']$. Then by an argument similar to the proof of Theorem 3.2, $E[g'|\sigma(f_n)] = \hat{g}_n(f_n)$, $E[g|\sigma(f)] = \hat{g}(f)$. By the argument of the previous paragraph, $\{\Psi_n * (g'\phi)\}$ converges to $\Psi * (g'\phi)$ a.e. (Lebesgue measure) and in Lebesgue measure.

Choose $h \in C_0$ and $\varepsilon > 0$. Let $\delta > 0$ be such that for $x, x' \in \mathbb{R}$ with $|x - x'| < \delta$, it follows that $|h(x) - h(x')| < \varepsilon/4$. Let M be an upper bound (in absolute value) of h and g'. Choose $K \subset \bar{\mathbb{R}}$ to be compact such that

(1) $\mu\{\omega|f_n(\omega) \in K\} > 1 - \varepsilon/(\delta M)$ for each n

(2) there exists $m > 0$ such that $\Psi * \phi(t) > m$ for $t \in K$

Let $\eta > 0$ be such that for x, y, z, w nonnegative reals with $y > m$, $|x - z| < \eta$, and $|y - w| < \eta$, it follows that $|(x/y) - (z/w)| < \delta$. Then by the absolute continuity of f_n and f, there exists by Proposition 24, p. 72 of Royden (1968) some $K' \subset K$ and $N > 0$ such that for $n \geq N$ and $t \in K'$,

(1′) $\mu\{\omega|f_n(\omega) \in K'\} > 1 - \varepsilon/(8/M)$

(2′) $|\Psi_n * (g'\phi)(t) - \Psi * (g'\phi)(t)| < \eta$

(3′) $|\Psi_n * \phi(t) - \Psi * \phi(t)| < \eta$

taking note that as $t \to \infty$, the left-hand sides of (2) and (3) go to 0, so for such n and t, $|\hat{g}_n(t) - \hat{g}(t)| < \delta$ and therefore $|h \circ \hat{g}_n(t) - h \circ \hat{g}(t)| < \varepsilon/4$. Then easily $|E[h \circ \hat{g}_n(f_n) - h \circ \hat{g}(f_n)]| \leq E[|h \circ \hat{g}_n(f_n) - h \circ \hat{g}(f_n)|] < \varepsilon/4 + \varepsilon/4 = \varepsilon/2$. By taking N to be larger if necessary, $|E[h \circ \hat{g}(f_n) - h \circ \hat{g}(f)]| < \varepsilon/2$, so $|E[h(E[g'|\sigma(f_n)]) - h(E[g'|\sigma(f)])]| < \varepsilon$.

For $g \in L^1(\mathcal{F}^2)$, choose $g' \in C_u$ with $\| g - g' \| < \delta\,\varepsilon/2$. Then for $n \geq N$ as above,

$$
\begin{aligned}
&|E[h(E[g|\sigma(f_n)]) - h(E[g|\sigma(f)])]| \\
&\leq |E[h(E[g|\sigma(f_n)]) - h(E[g'|\sigma(f_n)])]| + \varepsilon \\
&+ |E[h(ELg'|\sigma(f)]) - h(E[g|\sigma(f)])]|
\end{aligned}
$$

and since $\| E[g|\sigma(f_n)] - E[g'|\sigma(f_n)] \| < \delta\,\varepsilon/2$ and $\| E[g|\sigma(f)] - E[g'|\sigma(f)] \| < \delta\,\varepsilon/2$, a standard argument shows that $|E[h(E[g|\sigma(f_n)]) - h(E|g') \sigma(f_n)])]| < \varepsilon$ and $|E[h(E[g|\sigma(f)]) - h(E[g'|\sigma(f)])]| < \varepsilon$, completing the proof.

To model the role of information in economic decision-making, let $X \subset \mathbb{R}^\ell$ for some ℓ be closed, and $C(X; \mathbb{R})$ be the set of continuous functions $V: X \to \mathbb{R}$, with the topology of uniform convergence on compacta. In particular, $C(X; \mathbb{R})$ is a separable metric space. Convergence in distribution of random variables taking values in $C(X; \mathbb{R})$ is defined precisely as before.

<u>Theorem 4.2</u>: Let $U: \Omega^2 \to C(X; \mathbb{R})$ be Borel-measurable such that for some $r \in L^1(\mathcal{F}^2)$, $\sup_{x \in X} |U(\omega^2)(x)| \leq r(\omega^2)$ for a.e. ω^2. Then if $\{e_n\} \subset \bar{V}$ converges to $e \in \bar{V}$ in distribution, then $\{E[U|\sigma(f_n)]\}$ converges to $E[U|\sigma(f)]$ in distribution.

<u>Proof</u>: Let $P_n = \mu \circ (E[U|\sigma(f_n)])^{-1}$ and $P = \mu \circ (E[U|\sigma(f)])^{-1}$ be the

corresponding image measures on $C(X; \mathbb{R})$. By Lemma 8.3 of Cotter (1985), $\{P_n\}$ is tight and therefore relatively compact in the weak topology of measures by Theorem 6.1 of Billingsley (1968).

Let $u \in C(X; \mathbb{R})$. Then a closed sphere about u consists of a compact $K \subset X$ and $\varepsilon > 0$ such that the sphere is $\{v \in C(X;\mathbb{R}) \mid |u(x) - v(x)| \leq \varepsilon$ for all $x \in K\}$. Let $\{x_j\} \subset K$ be dense in K; then the above sphere is the intersection over all j of the sets $\{v \in C(X; \mathbb{R}) \mid |u(x_j) - v(x_j)| \leq \varepsilon$ for $j=1,\ldots,J\}$. Since $C(X; \mathbb{R})$ is separable, any open set is a countable union of closed spheres. Therefore by Billingsley (1968, p. 15), the sets of the form $\Pi^{-1}_{\{x_1,\ldots,x_J\}}(H)$, where $\Pi_{\{x_1,\ldots,x_J\}}: C(X; \mathbb{R}) \to \mathbb{R}^J$ is the projection evaluation map and $H \subset \mathbb{R}^J$ is closed, form a determining class for the set of measures on $C(X; \mathbb{R})$.

For $\{x_1,\ldots,x_J\} \subset X$, $P_n \circ \Pi^{-1}_{\{x_1,\ldots,x_J\}} = \mu^2 \circ (\Pi_{\{x_1,\ldots,x_J\}} \circ E[U|\sigma(f_n)])^{-1} =$ $\mu^2 \circ (E[\Pi_{\{x_1,\ldots,x_J\}} \circ U|\sigma(f_n)])^{-1} = \mu^2 \circ (E[U(x_1)|\sigma(f_n)], E[U(x_2)|\sigma(f_n)],\ldots$ $E[U(x_j)|\sigma(f_n)])^{-1}$ (with an abuse of notation), and similarly for P. By Theorem 6.1 of Billingsley (1968), weak convergence of $P_n \circ \Pi^{-1}_{\{x_1,\ldots,x_J\}}$ is equivalent to convergence in distribution of $\{\sum_{j=1}^{J} \alpha_j E[U(x_j)|\sigma(f_n)]\}$ to $\sum_{j=1}^{J} \alpha_j E[U(x_j)|\sigma(f)]$ for every $\alpha_1,\alpha_2,\ldots,\alpha_J \in \mathbb{R}$, but this follows immediately by Theorem 4.1. Since every subsequence of $\{P_n\}$ has a convergent further subsequence whose finite dimensional projections converge to those of P, and the finite dimensional sets are a determining class, it follows that each such subsequence converges to P, hence $\{P_n\}$ converges to P, completing the proof.

It follows easily that the optimal behavior resulting from the perturbed observation of the decision-relevant state has the same continuity with respect to information properties as do conditional expectations. The following theorem is generic, and includes the case of prices and demand.

Theorem 4.3: Let $U: \Omega^2 \to C(X; \mathbb{R})$ be Borel-measurable such that for some $r \in L^1(\mathcal{F}^2)$, $\sup_{x\in X}|U(\omega^2)(x)| \leq r(\omega^2)$ for a.e. ω^2. Suppose further that for a.e. ω^2, $U(\omega^2)$ is strictly concave. Let S be a metric space, and let $\gamma: S \twoheadrightarrow X$ be a nonempty, convex-valued, compact-valued upperhemicontinuous correspondence. Let $z: S \times \bar{V} \to L^1$ be defined by $z(S,e)(\omega) = \text{argmax } \{E[U|\sigma(f)](\omega)(x)|x\in\gamma(s)\}$. Then if $\{(s_n,e_n)\}$ is a sequence such that $\lim_{n\to\infty} s_n = s$ and $\{e_n\}$ converges in distribution to e, then $\{Z(s_n,e_n)\}$ converges in distribution to $Z(s,e)$.

Proof: Let $C^* \subset C(X; \mathbb{R})$ consist of the strictly concave continuous functions, with the subspace topology. Then $\hat{z}: S \times C^* \to X$ defined by $\hat{z}(s,u) = \text{argmax } \{u(x)|x\in\gamma(s)\}$ is continuous (Hildenbrand, 1974, pp. 29-30). The result follows by easy application of the definition of convergence in distribution.

Corollary 4.4: Let $u^*: S \times \bar{V} \to \mathbb{R}$ be the value of information defined by $u^*(s,e) = \int_\Omega U(\omega)(z(s,e)(\Omega))d\mu(\omega)$. Then if $\{(s_n,e_n)\}$ is a sequence such that $\lim_{n\to\infty} s_n = s$ and $\{e_n\}$ converges in distribution to e, then $\lim_{n\to\infty} u^*(s_n,e_n) = u^*(s,e)$.

Corollary 4.5: Let $M(\bar{\mathbb{R}})$ be the set of probability measures on $\bar{\mathbb{R}}$ with the weak topology. Let $u^*: S \times M(\bar{\mathbb{R}}) \to \mathbb{R}$ be the value of information defined by $u^*(s,\nu) = u^*(s,e)$ for any $e \in \bar{V}$ whose probability distribution is given by ν. Then u^* is continuous.

References

B. Allen, 1983, Neighboring Information and Distributions of Agents' Characteristics Under Uncertainty, Journal of Mathematical Economics 12, 63-101.

B. Allen, 1984a, The Existence of Fully Rational Expectations Approximate Equilibria With Noisy Price Observations, Discussion Paper No. 155, University of Bonn.

B. Allen, 1984b, The Existence of Rational Expectations Equilibria in a Large Economy With Noisy Price Observations, Journal of Mathematical Economics, forthcoming.

R. Anderson and H. Sonnenschein, 1982, On the Existence of Rational Expectations Equilibrium, Journal of Economic Theory 26, 261-278.

P. Billingsley, 1968, Convergence of Probability Measures (New York: Wiley).

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

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

K. Cotter, 1985, Similarity of Information and Behavior With a Pointwise Convergence Topology, unpublished manuscript, Northwestern University.

H. Kudō, 1974, A Note on the Strong Convergence of s-Algebras," Annals of Probability 2, 76-83.

H. Royden, 1968, Real Analysis, 2nd ed. (New York: Macmillan).