Available online at www.sciencedirect.com
SCIENCE DIRECT®

Games and Economic Behavior 47 (2004) 1–35

GAMES and Economic Behavior

www.elsevier.com/locate/geb

Aggregation and the law of large numbers in large economies

Nabil I. Al-Najjar

Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA

Received 26 August 2002

Abstract

This paper introduces a new model of environments with a large number of agents and stochastic characteristics. We consider sequences of finite but increasingly large economies that 'discretize' the continuum. In the limit we obtain a model that is continuum-like in important respects, yet it has a countable set of agents with a finitely additive, 'uniform' distribution. In this model, the law of large numbers is meaningful and holds on all subintervals. This framework provides, among other things, a new interpretation of the measurability problem and the failure of the law of large numbers in the continuum. It is also shown that the Pettis integral in the continuum coincides with the empirical frequencies in the discrete model almost surely. Finally, the model is used to study a mechanism design problem in a large economy with private information.

JEL classification: C7; C6; H41

1. Introduction

This paper introduces a new approach to model environments with a large number of agents and stochastic characteristics. The two standard approaches in the literature are based on either

(1) a sequence of finite but increasingly large sets of agents, or
(2) a continuum of agents.

E-mail address: al-najjar@nwu.edu.

0899-8256/$ – see front matter
doi:10.1016/S0899-8256(03)00175-1

The paper proposes a 'hybrid' model that eliminates many of the problems associated with (and provides a new interpretation of) these two common modeling approaches.

We consider a sequence $\{T_N\}_{N=1}^{\infty}$ of finite but increasingly large economies. The sequence is chosen so it 'discretizes' the continuum in a natural sense and yields, in the limit, a model with a *finitely additive distribution* on a *countable set of agents*. Although discrete, the limiting model we construct is 'continuum-like' in important ways: it preserves the interval and metric structures of the continuum, and the distribution of agents is 'uniform' and atomless. A rich class of independent stochastic processes of agent characteristics can be 'translated' back and forth between the continuum and our discrete model in an essentially unique manner. This class encompasses virtually all specifications used in applications, including all i.i.d., conditionally independent, and exchangeable distributions.

The law of large numbers states that the empirical frequency of independent random variables is almost surely equal to the population mean. It is well known that this statement is problematic in the continuum because the 'empirical frequency of a sample realization' cannot be meaningfully defined.[1] By contrast, in the discrete model introduced in this paper, the law of large numbers has a straightforward formal statement and its proof is a natural extension of the usual proof for a sequence of independent random variables. It is also straightforward to show that the law of large numbers holds simultaneously on all subintervals, and can be extended to non-identically distributed and/or correlated distributions.

Although new, the approach advocated here builds on two ideas already in the literature. First, Bewley (1986) and Guesnerie (1981, 1995) proposed modeling large economies as a sample of agents $\{t_1, \ldots\}$ randomly drawn from the continuum. The second idea, proposed by Feldman and Gilles (1985), is to use an infinite sequence of agents with a density charge.[2] This paper combines the strengths of the above authors' insights. Like Feldman and Gilles (1985), this paper uses density charges on a countable index set, and like Bewley (1986) and Guesnerie (1981), our model maintains a close connection to the continuum. The differences with these two approaches are substantial, however, and they are detailed in Section 5.

The usefulness of a formal model of large economies is judged by its ability to address economic questions. In Section 6, I use the framework of this paper to provide a large-economy formulation of a mechanism design problem in an environment with private information. Rob (1989) and Mailath and Postlewaite (1990) examined the effect of private information on efficiency in finite-agent economies with externalities and public goods. They showed that in a sequence of finite but increasingly large number of agents and independent types, the outcome of any incentive compatible and individually rational mechanism is asymptotically inefficient. In Section 6, I provide a large-economy formulation of this model and explain how the concepts of mechanism design theory extend to such environments. In this limiting model, exact (rather than asymptotic) inefficiency is easily established using the concept of influence in Al-Najjar and Smorodinsky (1996).

[1] Judd (1985) and Feldman and Gilles (1985). See Section 4 for a detailed discussion.

[2] That is, a uniform, finitely additive distribution on the integers.

The mechanism design example of Section 6 illustrates a number of important points. In the setting of that example, it would be unreasonable to restrict attention to anonymous mechanisms (ones where two agents of the same type are treated identically). As a result, a distributional model of a large economy—i.e., where the large economy is viewed as a measure on a space of characteristic—may be inappropriate because it eliminates agents' labels and thus forces mechanisms to be anonymous. Our framework preserves agents' labels and thus has no problem handling non-anonymous mechanisms. Second, the example illustrates the need for a strong law of large numbers in studying mechanism design problems in large economies with private information. In the example, types are independent, so many natural classes of mechanisms are not even well-defined without a strong law of large numbers (e.g., to compute aggregate transfers).

The well known technical problems associated with introducing idiosyncratic risk in the continuum did not slow down its wide-spread use in applications. At an intuitive level, the continuum does appear to 'make sense' as a model of large economies and it has indeed proved valuable in generating important economic insights. This paper provides some support for such practice. Specifically, I prove a simple characterization (Theorem 6) stating that the Pettis integral in the continuum and the empirical frequencies in the discrete model coincide almost surely. The Pettis integral was suggested by Uhlig (1996) as a way to formulate a weak law of large numbers for the continuum (see Sections 3.5 and 5). The problem is that the Pettis integral has no obvious meaning in terms of empirical frequencies of sample realizations. The above mentioned result provides a way to interpret the Pettis integral in a way suggestive of a strong law of large numbers for the continuum.

Finitely additive measures (and, in particular, density charges) are standard mathematical tools—see, for example, Dunford and Schwartz (1958) and Rao and Rao (1983). Although used extensively in certain research fields (such as Decision Theory; see, for instance, Fishburn, 1970), they are uncommon as models of large economies. In Section 2, I discuss two problems that might have led to the limited use of countable models of agents and a finitely additive measure, and explain how they can be remedied using the framework introduced here. The application to mechanism design in Section 6 further illustrate the applicability of our model to concrete economic settings.

Some of the earliest published work using finitely additive measure spaces to model large economies are Weiss (1981), Armstrong and Richter (1984, 1986). Weiss establishes the equivalence of core and competitive allocations in a model with a finitely additive space of agents. Armstrong and Richter (1984) provide a more general treatment of the same problem covering, among other things, countable spaces of agents with finitely additive measures. They point out that formulating coalitions in the continuum requires the unconvincing assumption that the set of coalitions be a σ-algebra. Armstrong and Richter (1986) prove a general result on the existence of competitive equilibria that covers the environments of their earlier paper. The argument they use in the proof consists of mapping the finitely additive model into a countably additive one. They then show that the existence proof in the countably additive model carries back to the original finitely additive environment they started with. Other subsequent uses of finitely additive measures in Economics and Game Theory include Feldman and Gilles (1985), Gilboa and Matsui (1992), Werner (1997), Marinacci (1997). See these papers for more exhaustive account of the literature.

Three approaches to modeling large economies are directly comparable to the one presented in this paper. These are:

(1) the sampling approach of Bewley (1986) and Guesnerie (1981, 1995);
(2) density charges on a countable set of agents (Feldman and Gilles, 1985); and
(3) linear methods relying on variants of the Pettis integral (Uhlig, 1996; Al-Najjar, 1995).

I provide a detailed comparison with these approaches in Section 5. A more recent interesting approach is that introduced by Sun (1998) and Khan and Sun (1999) based on hyperfinite Loeb spaces. In these papers, the space of agents is a hyperfinite internal probability space that includes non-standard entities (e.g., infinitesimals and hyper-real numbers). This approach appears to be sufficiently different from the one introduced here that a detailed comparison is difficult. The interested reader may refer to these works and to the references therein for further details.

2. Three models of large economies

2.1. The continuum and its discretization

Our starting point is the standard model of a large economy as a continuum of agents, $\overline{T} = [0, 1]$, with the uniform distribution $\bar{\lambda}$.[3]

We would like to think of the continuum $\overline{T}$ as an idealization of a sequence of finite models, $\{T_N\}_{N=1}^{\infty}$, where the Nth model has the form

$$T_N = \{t_1, \dots, t_{\#T_N}\} \subset \overline{T}, \quad N = 1, 2, \dots.$$

The uniform distribution on T_N, denoted λ_N, assigns to each subset $A \subset T_N$ a measure equal to its relative frequency:

$$\lambda_N(A) \equiv \frac{\#A}{\#T_N}.$$

Notational conventions. It is convenient to think of λ_N as a distribution on $\overline{T}$ with support T_N so λ_N and $\bar{\lambda}$ are defined on the same space. With this convention, the expression $\lambda_N([a, b])$ has an unambiguous meaning as $\lambda_N(T_N \cap [a, b])$. We may thus use λ_N and T_N interchangeably to talk about the Nth model, since T_N is just the support of λ_N.

We assume that the sequence $\{T_N\}_{N=1}^{\infty}$ asymptotically has the same interval and measure structures as the continuum $\overline{T}$:

Definition 1. A *discretizing sequence* $\{T_N\}_{N=1}^{\infty}$ is one satisfying:

(1) $T_N \subset T_{N+1}$ for every N; and
(2) The sequence of finite distributions $\{\lambda_N\}$ converges weakly to $\bar{\lambda}$.

[3] The continuum [0,1] is endowed with the σ-algebra $\overline{\mathcal{T}}$ of the Borel sets.

Using a standard characterization of weak convergence,[4] condition 2 is equivalent to the requirement: for every non-degenerate interval of agents $[a, b] \subset [0, 1]$,[5]

$$\lim_{N\to\infty} \lambda_N\big([a, b]\big) = \bar{\lambda}\big([a, b]\big). \tag{1}$$

That is, the mass of agents in T_N who fall within a given interval $[a, b]$ asymptotically coincides with the corresponding mass of agents in the continuum. Note that this implies that $\bigcup_{N=1}^{\infty} T_N$ must be dense in $\overline{T}$.[6]

2.2. *The large discrete model*

Fix a discretizing sequence $\{T_N\}_{N=1}^{\infty}$. Intuitively, as N becomes large, one would expect individual agents to have vanishing weight and for the law of large numbers to approximately hold as N goes to infinity. In the limit, when the number of agents is infinite, one would expect these properties to hold *exactly*.

What is the 'right limit' for a sequence of increasingly large, finite economies? Taking the continuum $\overline{T}$ to be that limit, each agent indeed has zero weight, but serious difficulties arise in stating the law of large numbers. Here I introduce a new model, T, as a limit of the sequence $\{T_N\}_{N=1}^{\infty}$. The set T satisfies $T_N \subset T \subset \overline{T}$ and has the special structure detailed below:[7]

Theorem 1. *Fix any discretizing sequence $\{T_N\}_{N=1}^{\infty}$ and define $T = \bigcup_{N=1}^{\infty} T_N$. Then there is a finitely additive probability measure λ on 2^T such that*

$$\lambda(A) = \lim_{N\to\infty} \lambda_N(A) \tag{2}$$

for every $A \subset T$ for which the limit exists.

A measure λ with the properties asserted in Theorem 1 is known as a *density charge* on T. Such measure assigns to every set its limiting frequency (the RHS of (2)) whenever this frequency is well defined. The theorem says that λ can be extended to every subset of agents in an additive fashion.

Definition 2. A *discretization* of $\overline{T}$ is a pair $(\{T_N\}_{N=1}^{\infty}, \lambda)$ satisfying the properties in Theorem 1.[8]

---

[4] See, for instance, Shiryayev (1984, Theorem 1, p. 309). To apply that theorem, it suffices to note that the boundary of the interval $[a, b]$ has $\bar{\lambda}$-measure 0.

[5] Throughout the paper, an 'interval' will refer to a set $[a, b]$ with $a < b$. This rules out degenerate intervals of the form $[a, a]$.

[6] Constructing a discretizing sequence is straightforward. For instance, one can choose T_N by induction as $T_{N+1} = \{k/(N+1)\}_{k=0}^{N+1} \cup T_N$.

[7] All proofs are in Appendix A. Except for a minor change, Theorem 1 is just the standard result that can be found, say, in Rao and Rao (1983, p. 41) or Feldman and Gilles (1985).

[8] When $\{T_N\}$ and/or λ is clear from the context, we may refer to (T, λ) or just T as a discretization. Any sequence $\{T_N\}_{N=1}^{\infty}$ gives rise to a unique T; conversely, a given T has meaning in this paper only if it arises from a sequence of finite models $\{T_N\}$ satisfying Eq. (2).

This construction has several implications: First, any such measure must be *atomless*[9] so, as with the continuum, λ assigns zero mass to any individual agent. Second, λ cannot be countably additive since T is countable, $\lambda(t) = 0$ for every $t \in T$, but $\lambda(T) = 1$.[10] Third, the value of λ is completely pinned down for sets $A \subset T$ whose asymptotic frequency is unambiguously defined and these include, by Eq. (1), the set of all intervals. On the other hand, λ is not uniquely defined for sets without a well-defined limiting frequency. Later we introduce additional structure that makes the indeterminacy of λ irrelevant (see Theorem 2).

2.3. Integration

Functions of the form $f : T \to \mathbb{R}$ will be used to represent agents' endowments, transfers, types, etc. We are interested in integrating (averaging out) such functions to obtain average endowment, transfers and so on. The integral $\int_T f \,\mathrm{d}\lambda$ with respect to a density charge λ is well defined and extensively studied; its development closely follows that of the usual integral.[11] For a *simple* function f, i.e., a function with finite range $\{x_1, \ldots, x_K\}$, define

$$\int_T f \,\mathrm{d}\lambda \equiv \sum_{k=1}^{K} x_k \lambda\bigl(\{t\colon\ f(t) = x_k\}\bigr). \tag{3}$$

That is, the integral is just the average of its values weighted by their frequencies. Note that this is well defined for *any* simple function, since every subset of T is measurable.

For more general functions, we have:

Definition 3. The integral of a bounded function $f : T \to \mathbb{R}$ is

$$\int_T f \,\mathrm{d}\lambda = \lim_{n\to\infty} \int_T f_n(t) \,\mathrm{d}\lambda,$$

where $f_n : T \to \mathbb{R}$, $n \in \mathbb{N}$, is any sequence of simple functions such that f_n converges to f uniformly.

Standard results (e.g., Rao and Rao, 1983, Chapter 4, or Fishburn, 1970, Chapter 10) show that this integral is well defined for all bounded functions, does not depend on the particular approximating sequence $\{f_n\}$, and satisfies the basic properties of integrals.

[9] The limiting frequency of any finite set $\{t_1, \ldots, t_m\}$ is zero, hence λ assigns weight zero to such sets.

[10] In fact, λ is *purely* finitely additive: If η is any *countably* additive measure on $(T, 2^T)$ such that $\eta(A) \leqslant \lambda(A)$ for every $A \subset T$, then $\eta(A) = 0$ for every $A \subset T$. See Rao and Rao (1983, p. 240). The fact that λ is purely finitely additive follows from Theorem 10.3.2 in Rao and Rao.

[11] Dunford and Schwartz (1958) develop the theory integration for finitely additive measures, then introduce countable additivity as a special case. Rao and Rao (1983) is a more specialized reference.

It is convenient to define integration with respect to λ_N, which reduces to taking the simple average of a function over T_N:[12]

$$\int_T f \,\mathrm{d}\lambda_N \equiv \frac{1}{\#T_N} \sum_{t \in T_N} f(t).$$

Notational convention. Using our convention to view λ_N as a measure on $\overline{T}$ with support T_N, given $f : A \to \mathbb{R}$ it is legitimate to write:

$$\int_{T_N} f \,\mathrm{d}\lambda_N = \int_B f \,\mathrm{d}\lambda_N, \quad \text{whenever } T_N \subset A \subset B \subset \overline{T}.$$

For example, $\int_a^b f \,\mathrm{d}\lambda_N$ is meaningful for $f : T \to \mathbb{R}$ (here $B = [a, b]$ and $A = T \cap B$).

2.4. *Functions with well-behaved frequencies*

From Section 2.3 we know that *any* bounded function is integrable with respect to the finitely additive probability λ. Despite this, models with finitely additive probabilities attracted little interest in the literature because the freedom they offer comes at a price. Problems include that finitely additive probabilities allow 'mass to disappear'; they can be hard to interpret as limits of large finite models; and basic results such as Fubini's theorem might fail.

In Section 6, I illustrate these problems in the context of mechanism design in large economies. Here, I simply introduce a class of extremely well-behaved functions, $\mathcal{F}$, and show that they do not suffer from many of the problems common in models with finitely additive probabilities. Another, independent reason for considering $\mathcal{F}$ is that every typical realization of an independent stochastic process must belong to it (see Theorem 4).

2.4.1. *Functions with well-defined distributions*

For a function on the continuum $\bar{f} : \overline{T} \to \mathbb{R}$, and $r \in \mathbb{R}$, we have the standard definition of a cumulative distribution function (cdf):

$$\overline{F}(r, [a, b]) \equiv \frac{1}{|a - b|} \bar{\lambda}\big(\big\{t \in [a, b]\colon\ \bar{f}(t) \leqslant r\big\}\big),$$

which must be non-decreasing and right-continuous in r. These properties of cdf's are essential in applications and follow as a by-product of countable additivity. It is useful to note that right-continuity in the continuum follows from a more basic property, namely that for every $r \in \mathbb{R}$ and interval $[a, b]$:

$$\bar{\lambda}\big(\big\{t \in [a, b]\colon\ \bar{f}(t) = r\big\}\big) = \inf_{r' > r} \overline{F}\big(r', [a, b]\big) - \sup_{r' < r} \overline{F}\big(r', [a, b]\big). \qquad (*)$$

[12] The integral f on a subset A is covered by this notation, since $\int_A f \,\mathrm{d}\lambda \equiv \int_T \chi_A f \,\mathrm{d}\lambda$, where χ_A is the indicator function of A.

Things are quite different in the discrete model where it is no longer guaranteed that mass behaves continuously under limits. Let

$$F\big(r,[a,b]\big) \equiv \frac{1}{|a-b|}\lambda\big(\big\{t \in [a,b]\colon\ f(t) \leqslant r\big\}\big).$$

Definition 4. A function $f : T \to \mathbb{R}$ has *well-behaved distribution function* if for every interval $[a,b]$ and $r \in \mathbb{R}$,

$$\lambda\big(\big\{t \in [a,b]\colon\ f(t) = r\big\}\big) = \inf_{r'>r} F\big(r',[a,b]\big) - \sup_{r'<r} F\big(r',[a,b]\big). \tag{4}$$

Without this restriction, $F(\cdot,[a,b])$ may fail to be right-continuous. To see this, consider the following example:

Example 1. Define $f : T \to \mathbb{R}$ by: $f(t) = 1$ for $t \in T_1$, and $f(t) = 1/N$ for $t \in T_N - T_{N-1}$, $N > 2$. Then for every $r > 0$, $\lambda\{f(t) \leqslant r\} = 1$, yet $\lambda\{f(t) \leqslant 0\} = 0$. In particular, $\int_T f\,\mathrm{d}\lambda = 0$.[13]

In this example, $F(r,[0,1]) = 1$ for every $r > 0$, but $1 = \lim_{r\downarrow 0} F(r,[0,1]) \neq F(0,[0,1]) = 0$, so F is *not* a distribution function. The problem is that points at which f is less than r has mass 1 for every $r > 0$, but this mass disappears in the limit when $r = 0$.

Mass may also escape "from below," as seen by examining the function $-f$, which has a well defined distribution function, but mass disappears, this time *upward* to zero. In both cases, mass accumulates at a point, but the mass of the point itself is zero. In the continuum, discontinuity at the limit is *indirectly* ruled out through countable additivity. With finitely additive probabilities, this discontinuity must be ruled out explicitly, as we do in Definition 4.

2.4.2. Functions with asymptotic frequencies

Another issue is the relationship between the limiting model T and large finite models, T_N. Define the cdf relative to T_N as

$$F_N\big(r,[a,b]\big) \equiv \frac{1}{|a-b|}\lambda_N\big(\big\{t \in [a,b]\colon\ f(t) \leqslant r\big\}\big).$$

The following example shows that F_N and F need not be related.

Example 2. Fix $0 < \epsilon < \frac{1}{2}$ and an increasing sequence of integers $\{N_k\}$ such that $N_{k+1} > N_k/\epsilon$. For $k \geqslant 2$, let $f(t) = 0$ if $t \in T_{N_{k+1}} - T_{N_k}$ for odd k and $f(t) = 1$ for even k. Then

$$\liminf_{N\to\infty} \int_T f(t)\,\mathrm{d}\lambda_N < \epsilon < 1 - \epsilon < \limsup_{N\to\infty} \int_T f(t)\,\mathrm{d}\lambda_N.$$

[13] To prove these claims, $\lambda\{f(t) \leqslant 0\} = 0$ is obvious since f is strictly positive; $\lambda\{f(t) \leqslant r\} = 1$ follows from the facts that $f(t) < r$ for $t \in T - T_N$ for N large enough so the set $\{f(t) > r\}$ is finite and hence has λ-measure zero. Finally, $\int_T f\,\mathrm{d}\lambda = 0$ follows from similar reasoning.

Although the integral $\int_T f(t)\,\mathrm{d}\lambda$ in this example is a well-defined number in [0, 1], it bears no connection to the integrals in large finite models, $\int_T f(t)\,\mathrm{d}\lambda_N$, which fluctuate wildly as N changes. In the example, the problem is that the limit f reflects poorly the asymptotic behavior of the distributions in finite models.

To address the issue raised by this example, we introduce the restriction:

Definition 5. A function $f : T \to \mathbb{R}$ has *asymptotic frequencies* if for every interval $[a, b]$,

$$F_N\bigl(\cdot, [a, b]\bigr) \to F\bigl(\cdot, [a, b]\bigr) \quad \text{weakly.}$$

Note that $\lim_{N\to\infty} \lambda_N(A)$ exists if and only if the indicator function $\chi_A : T \to \mathbb{R}$ has asymptotic frequencies on [0, 1].

2.5. Integration revisited

We combine the last two definitions to obtain

Definition 6. $\mathcal{F}$ is the set of all functions with well-defined distribution functions and asymptotic frequencies.

2.5.1. Integration of functions in $\mathcal{F}$ is extremely well behaved

Theorem 2. *For any $f \in \mathcal{F}$ and interval $[a, b]$,*

(1) *$F(\cdot, [a, b])$ is a distribution function.*
(2) *Change of variables*:

$$\frac{1}{|a-b|} \int_a^b f\,\mathrm{d}\lambda = \int_{\mathbb{R}} r\,\mathrm{d}F\bigl(r, [a, b]\bigr). \tag{5}$$

(3) *The integral in the limit is the limit of integrals in finite models*:

$$\int_a^b f\,\mathrm{d}\lambda = \lim_{N\to\infty} \int_a^b f\,\mathrm{d}\lambda_N. \tag{6}$$

In particular, the integral $\int_a^b f\,\mathrm{d}\lambda$ of a function $f \in \mathcal{F}$ does not depend on the particular extension λ chosen in Theorem 1.

2.5.2. Piecewise uniformly continuous functions

An important class of functions we shall use is

Definition 7. The set of *piecewise uniformly continuous functions*, denoted $\mathcal{C}$, is the set of all functions $f : T \to \mathbb{R}$ such that for some real numbers $0 = a_1 < \cdots < a_M = 1$, f is uniformly continuous on every 'interval' $(a_{m-1}, a_m) \cap T$.[14]

We define $\overline{\mathcal{C}}$ similarly for functions $\bar{f} : \overline{T} \to \mathbb{R}$.[15]

In Section 4.2 we show that there exists between $\mathcal{C}$ and $\overline{\mathcal{C}}$ a natural (essentially) one-one and onto mapping that preserves integration. For the moment we simply note:

Theorem 3. $\mathcal{C} \subset \mathcal{F}$.

In summary, $\mathcal{F}$ is a rich and extremely well-behaved class of functions: It includes all piecewise uniformly continuous functions; the integral of any $f \in \mathcal{F}$ can be computed by integrating the corresponding cdf; and any such function has a natural interpretation as a limit of a sequence of functions in the finite models. Our interest in $\mathcal{F}$ stems from the fact that a typical realization of an independent random process must belong to $\mathcal{F}$ (Theorem 4), yet such realizations cannot even be approximated by functions in $\mathcal{C}$ (Section 4).

3. The law of large numbers

We will consider models where each agent t has a random characteristic (e.g., his type, endowment, and so on), modeled as a random variable $\tilde{s}(t)$ taking values in some set of characteristics S.

We first consider the special case $S = \{0, 1\}$ to illustrate the paper's main ideas in the simplest possible setting. In Section 3.8, we show how to extend the analysis to the case where S is an arbitrary finite set.

3.1. The law of large numbers in the discrete model

3.1.1. States, events, and probabilities

It is convenient to have a single state space Ω rich enough to represent randomness in the continuum and discrete models simultaneously.

Formally, let $\Omega = S^{\overline{T}}$, i.e., the set of all functions $\omega : [0, 1] \to \{0, 1\}$. We find it convenient to define agents' characteristics as a function $s : T \times \Omega \to \{0, 1\}$, with $s(t, \omega)$

[14] That is, for every $\epsilon > 0$ there is $\delta > 0$ such that for every $t, t' \in (a_{m-1}, a_m)$, $|t - t'| < \delta$ implies $|f(t) - f(t')| < \epsilon$.

[15] While every continuous function on $\overline{T} = [0, 1]$ is uniformly continuous, this is not true in T since it is not compact. For example, choose $0 < \bar{t} < 1$ such that $\bar{t} \notin T$ and consider the function $f : T \to \mathbb{R}$ given by $f(t) = 0$ for $t < \bar{t}$ and $f(t) = 1$ for $t > \bar{t}$. Then f is continuous relative to T, but not uniformly so. Note that any extension of f to the continuum $\overline{T}$ would necessarily be discontinuous at $\bar{t}$. Roughly, the problem is that f's "discontinuity point" $\bar{t}$ falls outside its domain T.

representing the realization of agent t at state ω (with this notation, $s(t,\omega) \equiv \omega(t)$). We call the function $s(\cdot,\omega): T \to \{0,1\}$, or $t \mapsto s(t,\omega)$ for short, the *sample realization* at a state ω.

Notational conventions. We use a '$\tilde{}$' in referring to a random variable when explicit reference to the state is omitted. Thus, we shall write $s(t,\omega)$, without '$\tilde{}$', but write $\tilde{s}(t)$ to emphasize that the latter refers to a random outcome.

Let Σ be the σ-algebra on Ω generated by the random variables $\{\tilde{s}(t):\ t \in T\}$. Then Σ is the smallest σ-algebra of events with respect to which each $\tilde{s}(t)$, $t \in T$, is measurable.[16] This formulation enables us to model randomness when the space of agents is T using the large state space Ω since Σ treats as identical two states that agree on T. Similarly, to model randomness when the space of agents is T_N yet maintain Ω as the state space, we introduce $\Sigma_N \subset \Sigma$, the sub-σ-algebra generated by the collection of random variable s $\{\tilde{s}(t):\ t \in T_N\}$.

Randomness in characteristics can now be introduced as a probability distribution on the relevant sets of events. In the discrete space of agents T, this is a distribution P on (Ω, Σ). Since $\Sigma_N \subset \Sigma$, such distribution also defines a probability distribution on T_N by restriction.

For a probability distribution P, let $\mu: T \to [0,1]$ denote its expectation function (that is, $\mu(t) \equiv E_P\tilde{s}(t)$).

3.1.2. Main result

Definition 8. A distribution P on (Ω, Σ) is *independent* if for every finite subset $\{t_1, \ldots, t_K\} \subset T$ the random variable s $\{\tilde{s}(t_1), \ldots, \tilde{s}(t_K)\}$ are independent.

Theorem 4 (Strong law of large numbers). *Suppose that P is independent and $\mu \in \mathcal{C}$. Then for P-almost every ω, the sample realization $t \mapsto s(t,\omega)$ belongs to $\mathcal{F}$, and*

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \mu(t)\,\mathrm{d}\lambda, \quad \textit{for every } [a,b]. \tag{7}$$

That is, the integral of the sample realization, $\int_a^b s(t,\omega)\,\mathrm{d}\lambda$, equals the integral of the expectations on every subinterval almost surely. The fact that $t \mapsto s(t,\omega)$ belongs to $\mathcal{F}$ assures that this integral corresponds to the limit of averages of sample realizations in large but finite models T_N:

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \lim_{N\to\infty} \int_a^b s(t,\omega)\,\mathrm{d}\lambda_N \equiv \lim_{N\to\infty} \frac{1}{\#T_N \cap [a,b]} \sum_{T_N \cap [a,b]} s(t,\omega).$$

3.2. Simple examples

The following examples illustrate the theorem.

[16] This is a minimal requirement that must be satisfied in order to assign probabilities to agents' characteristics.

Example 3 (*The i.i.d. case*). Suppose that P is i.i.d. with mean α. Then $\int_0^1 \mu(t)\,\mathrm{d}\lambda = \alpha$. By Theorem 4, for every interval $[a, b]$:

$$\frac{1}{|a-b|}\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \alpha, \quad P\text{-a.s.}$$

Here we have the usual statement of the strong law of large numbers, namely that, with probability one, the sample distribution of agent characteristics is equal to the population mean.

Example 4. Suppose that P is independent, with $\mu(t) = 1 - t$. Then, for every interval $[a, b]$,

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \mu(t)\,\mathrm{d}\lambda = \frac{(b-a)^2}{2} + (b-a)\,(1-b), \quad P\text{-a.s.}$$

In this example, $\int_0^1 \mu(t)\,\mathrm{d}\lambda = 0.5$, and the theorem indeed implies that $\int_0^1 s(t,\omega)\,\mathrm{d}\lambda = 0.5$, P-a.s. But the theorem delivers a finer conclusion: it indicates how the integral changes after conditioning on agents being in a particular subinterval. This is a consequence of the fact that the law of large numbers in our model holds on all subintervals, as we see next.

3.3. The law of large numbers holds on subintervals

Feldman and Gilles (1985) suggested that it is desirable for the law of large numbers to hold not just on [0, 1], but also on any subinterval $[a, b]$. Each such subinterval contains infinitely many agents, so one would expect the aggregation of uncertainty implied by the law of large numbers to hold there as well. Feldman and Gilles then showed that this conclusion fails when the space of agents is the continuum $\overline{T}$.

We shall see below that this problem does not appear in the discrete model of this paper. Before stating this formally, it is instructive to formulate Feldman and Gilles's idea more generally:

Definition 9. Suppose that P is independent and $\mu \in \mathcal{C}$. Then the law of large numbers *holds simultaneously* on a collection of subsets $\mathcal{A} \subset 2^T$ if there is $\Omega' \in \Sigma$, with $P(\Omega') = 1$, such that for every $\omega \in \Omega'$ and $A \in \mathcal{A}$:

$$\int_A s(t,\omega)\,\mathrm{d}\lambda = \int_A \mu(t)\,\mathrm{d}\lambda.$$

What are the collections of sets $\mathcal{A}$ on which the law of large numbers holds simultaneously? Theorem 4 tells us that one such collection is the set of all non-degenerate subintervals:

Corollary 1. *Suppose that P is independent with $\mu \in \mathcal{C}$, then the law of large numbers holds simultaneously on all subintervals.*

It is easy to see that the law of large numbers does not hold simultaneously on *all* infinite subsets, 2^T.[17] Theorem 4 and Corollary 1 thus illustrate the role of the interval structure that T inherits from the continuum.

3.4. The law of large numbers in the continuum

3.4.1. States, events, and probabilities

When the space of agents is the continuum $\overline{T} = [0, 1]$, we continue to use $\Omega = S^{\overline{T}}$ as state space. However, to ensure that agents' random characteristics are measurable, we enrich the set of events to $\overline{\Sigma}$, the σ-algebra generated by the random variable s $\{\tilde{s}(t)\colon\ t \in \overline{T}\}$. Then $\overline{\Sigma}$ is the smallest σ-algebra of events with respect to which each $\tilde{s}(t)$, $t \in \overline{T}$, is measurable.

A stochastic structure is a probability distribution $\overline{P}$ on $\overline{\Sigma}$. Since $\Sigma_N \subset \Sigma \subset \overline{\Sigma}$, any such $\overline{P}$ defines, by restriction, probability distributions on the discrete models T and T_N. As before, define $\bar{\mu} : \overline{T} \to [0, 1]$ by $\bar{\mu}(t) \equiv E_{\overline{P}}\tilde{s}(t)$.

Definition 10. A distribution $\overline{P}$ on $(\Omega, \overline{\Sigma})$ is *independent* if for every finite subset $\{t_1, \ldots, t_K\} \subset \overline{T}$ the random variable s $\{\tilde{s}(t_1), \ldots, \tilde{s}(t_K)\}$ are independent.

3.4.2. How should the law of large numbers be defined in the continuum?

One would like to have an analogue of Theorem 4 for the continuum, namely a result asserting that the 'average' of a sample realization equals the integral of the expectations almost surely. How should such claim be stated? Recall the usual statement of the law of large numbers in the standard case of an i.i.d. sequence of random variables $\{\tilde{x}_1, \ldots\}$, namely that with probability 1;

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{x}_n = E\tilde{x}.$$

In the continuum, it is natural to replace averages by integrals, yielding a statement of the form

$$\int_a^b s(t, \omega)\, \mathrm{d}\bar{\lambda} = \int_a^b \bar{\mu}(t)\, \mathrm{d}\bar{\lambda}, \quad \text{for every } [a, b], \quad \overline{P}\text{-a.s.} \tag{8}$$

The integral on the RHS represents the population mean and is always well-defined whenever $\bar{\mu} \in \overline{\mathcal{C}}$ (an assumption we shall maintain throughout). The problem is the integral

[17] Example: let P be an i.i.d. process with mean 0.5, and Ω' a set of P-probability 1 for which the conclusion of Theorem 4 holds. Obviously, $(1/\lambda(A)) \int_A \mu(t)\, \mathrm{d}\lambda = 0.5$ for every $A \subset T$. For $\omega \in \Omega$ define $A_\omega = \{t \in T\colon\ s(t, \omega) = 0\}$. Then, for every $\omega \in \Omega'$, $(1/\lambda(A_\omega)) \int_{A_\omega} s(t, \omega)\, \mathrm{d}\lambda = 0$. Thus, for every $\omega \in \Omega'$, the law of large numbers fails on some subset A_ω of T. The example in fact proves that the law of large numbers fails to hold simultaneously even on the collection of sets with well-defined frequencies.

of the sample realizations, "$\int_a^b s(t,\omega)\,\mathrm{d}\bar{\lambda}$". One is tempted to define this as the Lebesgue integral of the sample realization $t \mapsto s(t,\omega)$. From the literature (see Section 4), we know this to be problematic because a typical sample realization is not measurable so this integral is not meaningfully defined. In Section 4, I argue that this lack of measurability is an inevitable feature of the continuum, one with important substantive implications.

One solution suggested by the framework of this paper is to replace $\bar{\lambda}$ with λ in the LHS of Eq. (8), obtaining:

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \bar{\mu}(t)\,\mathrm{d}\bar{\lambda} \quad \overline{P}\text{-a.s.}$$

That is, we look at the integral of the sample realization on a discretization of the space of agents. Note that, although $s(\cdot,\omega)$ is defined on all of $\overline{T}$, it is averaged out only over points in the discretization T, and the a.s. statement is made with respect to the probability law $\overline{P}$ on the continuum.[18]

3.4.3. Main result

Theorem 5. *For every independent $\overline{P}$ with $\bar{\mu} \in \overline{\mathcal{C}}$, any discretization T, and any interval $[a,b]$:*

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \bar{\mu}(t)\,\mathrm{d}\bar{\lambda}, \quad \overline{P}\text{-a.s.} \tag{9}$$

In particular,

$$\int_a^b \bar{\mu}(t)\,\mathrm{d}\bar{\lambda} = \lim_{N\to\infty} \int_a^b s(t,\omega)\,\mathrm{d}\lambda_N, \quad \overline{P}\text{-a.s.} \tag{10}$$

That is, although a typical sample realization does not have a well-defined integral in the continuum, it has, nonetheless, considerable structure: its integral in the discrete model is well-behaved and satisfies the law of large numbers .

Note that the theorem asserts that this procedure works for any discretization, any interval, and any independent probability law with $\bar{\mu} \in \overline{\mathcal{C}}$. Finally, Eq. (10) states that this is the 'right' integral in the sense that this is what we get by looking at averages of sample realizations in finite but increasingly large discrete models.

3.5. The Pettis integral and the law of large numbers

For two random variable f, f' on $(\Omega, \overline{\Sigma})$, define their inner product as $(f \mid f') = \int_\Omega f \cdot f'\,\mathrm{d}\overline{P} = \mathrm{cov}(f,f') + EfEf'$. Also, f and f' are *equivalent* if they agree with probability 1. The following definition is standard.[19]

[18] This makes sense since any such law induces a law in the model where the set of agents is T.

[19] Diestel and Uhl (1977) is a basic reference.

Definition 11. A random variable $[\int_{[a,b]}^{(P)} \tilde{s}(t)\,\mathrm{d}\bar{\lambda}] : \Omega \to \mathbb{R}$ is a *version of the Pettis integral* of $\{\tilde{s}(t)\colon\ t \in [a,b]\}$ if for every random variable $\tilde{x}$,

$$\left(\tilde{x} \,\Big|\, \int\limits_{[a,b]}^{(P)} \tilde{s}(t)\,\mathrm{d}\bar{\lambda}\right) = \int\limits_{[a,b]} (\tilde{x} \mid \tilde{s}(t))\,\mathrm{d}\bar{\lambda}, \tag{11}$$

where the integral on the RHS in Eq. (11) is the ordinary Lebesgue integral.[20]

The *Pettis integral* of the collection of random variables $\{\tilde{s}(t)\colon\ t \in [a,b]\}$ is the equivalence class of all such versions.

Uhlig (1996) suggested the Pettis integral as a way to formulate a *weak* version of the law of large numbers for the continuum (see the literature discussion in Section 5). A difficulty with this approach is that the Pettis integral has no transparent interpretation in terms of states and limiting averages of sample realizations. In particular, it is not clear what the Pettis integral evaluated at a state ω, $[\int_{[a,b]}^{(P)} \tilde{s}(t)\,\mathrm{d}\bar{\lambda}](\omega)$, has to do with the sample realization $t \mapsto s(t,\omega)$.

The framework of this paper provides a simple characterization the Pettis integral that supports its interpretation as a *strong* law of large numbers:

Theorem 6 (The Pettis integral as a strong law of large numbers). *Suppose that $\overline{P}$ is independent and $\bar{\mu} \in \overline{\mathcal{C}}$. Then on any interval $[a,b]$, the random variable $\omega \to \int_a^b s(t,\omega)\,\mathrm{d}\lambda$ is a version of the Pettis integral.*

That is, the sample average in the discrete model and the Pettis integral in the continuum coincide almost surely.

3.6. The role of our assumptions

The results of the last two subsections show that a law of large numbers for the continuum can be stated as a *natural extension* of the standard law for sequences. This is made possible through the framework developed in this paper. Two examples below show that our conclusions fail without this framework.

The first example shows that just taking an arbitrary countable dense set of agents is not enough:

Example 5. Suppose that Σ and P are defined relative to some dense sequence of agents $\{t_1, t_2, \ldots\}$ in $\overline{T}$.[21] Suppose also that P is independent with $\mu(t) = 0.7$ for $t \leqslant 0.5$ and $\mu(t) = 0.3$ otherwise. Then,

[20] Equation (11) should be interpreted as saying: the RHS integral exists and its value is equal to $(\tilde{x} \mid \int_{[a,b]}^{(P)} \tilde{s}(t)\,\mathrm{d}\bar{\lambda})$.

[21] So $\{t_1, t_2, \ldots\}$ is not necessarily a discretization of $\overline{T}$ in the sense of Definition 1.

(1) The dense sequence of agents $\{t_1, t_2, \ldots\}$ may be chosen so that, for almost every ω, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} s(t_n, \omega)$ exists but is different from 0.5;
(2) The dense sequence of agents $\{t_1, t_2, \ldots\}$ may be chosen so that, for almost every ω, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} s(t_n, \omega)$ does not exist.

The example illustrates that the way we build the discretizing sequence $\{T_N\}$ in Definition 1 is critical: a dense sequence of agents can be found so the conclusion of the law of large numbers can fail either because sample realizations have well-defined limit of averages but this limit violates the law of large numbers, or because the limit of averages does not even exist.

The second example shows that the use of finitely additive probabilities on the space of agents is critical.

Example 6. Suppose that Q is a dense subset of $\overline{T}$ (for instance, the set of rationals) and that P is i.i.d. with mean 0.5. Let λ' be any *countably additive* probability measure on Q. Define the limiting averages under λ' as: $\lim_{N\to\infty} \sum_{n=1}^{N} \lambda'(t_n) s(t_n, \omega)$. Note that $\int_0^1 \mu(t)\,\mathrm{d}\lambda' = 0.5$.

Then there is a set of states, $S \subset \Omega$, with $P(S) > 0$, such that

$$\int_0^1 s(t,\omega)\,\mathrm{d}\lambda' \neq \int_0^1 \mu(t)\,\mathrm{d}\lambda', \quad \text{for every } \omega \in S.$$

That is, the conclusion of the law of large numbers for the discrete model, Eq. (7), fails. In fact, the RHS of the above equation is a constant 0.5 while the LHS is random.[22]

In this example, countable additivity forces λ' to put most of its weight on a large finite set of points. Since the average outcome on any finite set of points is random, the limiting average under λ' will also be random and the law of large numbers cannot hold. This problem does not arise in our construction because the density charge λ assigns zero mass to any finite set of points.

3.7. Correlation

Our analysis extends to cover the type of correlated processes frequently used in the literature, namely processes where individual characteristics are independent conditional on a common aggregate parameter. To make this formal, let (Ω, Σ) be the state space in Section 3 and let $\Theta = \{\theta_1, \ldots, \theta_K\}$ be a finite set of aggregate parameters. For every $\theta \in \Theta$ let P^θ be a probability distribution on Ω. We shall write $\mu^\theta \equiv E_{P^\theta} \tilde{s}(t)$.

[22] To see this, arrange the elements of Q in any weakly decreasing order relative to the weights given by λ'. Fix $0 < \epsilon < 0.5$; countable additivity implies that we can find N large enough that $\lambda'(\{t_1, \ldots, t_N\}) \geqslant 1 - \epsilon$. Consider the positive probability event Ω' that $s(t_n, \omega) = 1$ for $1 \leqslant n \leqslant N$. For any $\omega \in \Omega'$, $\liminf_{N\to\infty} \sum_{n=1}^{N} \lambda'(t_n) s(t_n, \omega) \geqslant 1 - \epsilon$. Thus, on a positive probability event, the limiting average is either undefined, or different from the population mean.

Definition 12. P is *conditionally independent* if there is a probability distribution ν on Θ such that for every $S \in \Sigma$, $P(S) = \sum_\theta P^\theta(S)\nu(\theta)$ and for every $\theta \in \Theta$, P^θ is independent with $\mu^\theta \in \mathcal{C}$.

An example of how correlation may arise is the case where P is"exchangeable:" each P^θ is i.i.d. with mean α^θ, and these means are distinct. Realizations are independent given knowledge of the value of θ. Without this knowledge, however, learning the outcome $\tilde{s}(t)$ is informative about the underlying aggregate parameter θ and hence informative about the random outcome $\tilde{s}(t')$ of some other agent t'.

Theorem 7 (Strong law of large numbers in the correlated case). *Suppose that P is conditionally independent. Then for $P \times \nu$-almost every (ω, θ):*

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \mu^\theta(t)\,\mathrm{d}\lambda, \quad \textit{for every } [a,b]. \tag{12}$$

As in the independent case, the integral of the sample realization is almost surely equal to the population mean. The difference is that now the population mean $\int_a^b \mu^\theta(t)\,\mathrm{d}\lambda$ is itself random, and so unknown ex ante. All we can say is that, almost surely, the integral of the sample realizations takes the same value as the population mean, whatever that value may be.

3.8. Finite outcomes

Let S be an arbitrary finite set of characteristics. Then an agent's characteristic is a random vector $\tilde{s}(t) : \Omega \to S$ taking values in S. The results on the law of large numbers of Section 3 directly generalize. Here we only briefly sketch the changes to be made.

The state space again is $\Omega = S^{\overline{T}}$, i.e., the set of all functions $\omega : [0,1] \to S$, and Σ is the σ-algebra on Ω generated by the random variables $\{\tilde{s}(t) : t \in T\}$. Randomness in characteristics is represented by a probability distribution P on (Ω, Σ).

Let Δ denote the unit simplex representing the set of all probability distributions on the finite set S. It is convenient to identify S with the vertices of Δ, so $\tilde{s}(t) : \Omega \to \Delta$. With this notation, $\mu(t) \equiv E_P\tilde{s}(t)$ is a point in Δ, representing the probability of each element of S.

The definition of $\mathcal{C}$ is the natural multi-dimensional generalization of our earlier definition. Then an analogue of Theorem 4 for the multi-dimensional case can be stated and proved exactly as before component by component. The obvious proof is omitted.

4. Understanding the measurability problem

The difficulties in formulating a law of large numbers for the continuum revolve around whether sample realizations $t \mapsto s(t,\omega)$ are measurable, and whether the law of large numbers holds on all subintervals. See Judd (1985) and Feldman and Gilles (1985).[23] This section addresses two questions: Does the measurability problem have a substantive content? And, why does not the discrete model, despite preserving much of the character of the continuum, suffer from the same problem?

4.1. *Ex ante similarity vs. ex post variability*

Economic models appealing to the law of large numbers typically assume that agents have *ex ante similar* characteristics, but that their *realized characteristics* display a degree of stochastic independence. A good example is the common assumption that characteristics are independent and identically distributed, which is a strong form of ex ante similarity.[24]

One way to formalize ex ante similarity is to require that the *distribution* of characteristics be piecewise uniformly continuous. Theorem 9 in Section 4.2 below shows that, under this interpretation of similarity of characteristics, the continuum and discrete models have *identical* similarity structures: there is a natural, essentially one-one and onto correspondence between $\mathcal{C}$ and $\overline{\mathcal{C}}$. This means that any of the commonly used specifications of distributions in the continuum can be translated to the discrete model, and conversely.

The problem is that independence requires *ex ante similar* agents to be wildly *dissimilar ex post*. It is here where the discrete and continuum models are strikingly different. Recall the formulation of the law of large numbers in the continuum Eq. (8) as:

$$\int_a^b s(t,\omega)\,\mathrm{d}\bar{\lambda} = \int_a^b \bar{\mu}(t)\,\mathrm{d}\bar{\lambda}, \quad \text{for every } [a,b], \quad \overline{P}\text{-a.s.} \tag{13}$$

As discussed earlier, this is not meaningful because $\int_a^b s(t,\omega)\,\mathrm{d}\bar{\lambda}$ is not well-defined. In the discrete model, by contrast, Theorem 4 provides a straightforward formulation of the law of large numbers:

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \mu(t)\,\mathrm{d}\lambda, \quad \text{for every } [a,b], \quad P\text{-a.s.} \tag{14}$$

Why does Eq. (14) make sense, while Eq. (13) does not?

---

[23] Here is a brief formal statement of the problem. Consider an i.i.d. process with mean α, and let $\overline{P}$ denote the joint distribution on the state space Ω. Judd pointed out that the set of sample realizations $t \mapsto s(t,\omega)$ is non-measurable with respect to $\overline{\Sigma}$. Say that $t \mapsto s(t,\omega)$ satisfies the law of large numbers on every subinterval if $1/(a-b)\int_a^b s(t,\omega)\,\mathrm{d}\bar{\lambda} = \alpha$ on every subinterval $[a,b]$. Feldman and Gilles showed that any sample realization must be either non-measurable or does not satisfy the law of large numbers on every subinterval.

[24] Correlation does not remove the problem in the continuum, provided that some degree of conditional independence remains (as in Section 3.7, for instance).

We first note that, for a fixed ω, a sample realization $t \mapsto s(t,\omega)$ in the continuum is integrable if and only if it is approximately continuous, in the sense that:[25]

$$\inf_{\bar{f}\in\bar{\mathcal{C}}} \int_{\bar{T}} \left| s(t,\omega) - \bar{f}(t) \right| \mathrm{d}\bar{\lambda} = 0. \tag{15}$$

This means that any statement of the law of large numbers along the lines of Eq. (13) makes the implicit requirement that agents' *realized* characteristics must be approximately continuous. This, in essence, is the point made by Feldman and Gilles (1985).

But there is a fundamental incompatibility between stochastic independence and requiring nearby agents to have similar characteristics *ex post*. This incompatibility is not an artifact of the continuum; in fact, it appears in full force in our discrete model, as shown in the following result:

Theorem 8. *Suppose that P is i.i.d. with mean $\alpha = 0.5$. Then*

$$\inf_{f\in\mathcal{C}} \int_{T} \left| s(t,\omega) - f(t) \right| \mathrm{d}\lambda \geqslant 0.5 > 0 \quad P\text{-a.s.}$$

The theorem says that the discrete model is no different from the continuum in that independence also requires an irreducible discrepancy between sample realizations and their best continuous approximations.

If the conflict between stochastic independence and continuity of realized characteristics is the same in the discrete model as in the continuum, why does not the former suffer from the measurability problem? Theorem 8 shows that a typical realization of an i.i.d. process cannot be approximated by a continuous function. This, however, does not raise the same problems in the discrete model because we identified a broad class of functions, $\mathcal{F} \supset \mathcal{C}$, that are not approximately continuous, but nevertheless have a perfectly sensible integral. Roughly, the discrete model inherits the similarity structure of the continuum, but not its inability to integrate complicated functions (ones that cannot be approximated by continuous functions). In the continuum, by contrast, the only functions we can integrate are those that are approximately continuous.

The other question raised at the beginning of this section is: Does the measurability problem have a substantive content? Needless to say, independence is a very common assumption in economic models; models with ex ante identical agents whose characteristics are determined by independent random draws (representing wealth, information, types, etc.) abound. Independence in these models often plays a crucial, substantive role. In models with a finite number of agents, including auctions, public goods and other mechanism design problems, dropping independence may drastically alter the analysis. There is no reason to expect independence to be less critical in models with a large number of agents. In summary, extending the models of Information Economics to study large economies requires that one reconciles two conflicting goals: to be able to aggregate (integrate) individual choices and outcomes into economy-wide aggregates while, at the same time, allowing

[25] In fact, approximation can be obtained using only continuous functions (see, for instance, Royden, 1968). We use the class $\bar{\mathcal{C}}$ for expository convenience.

for stochastic independence of characteristics. The measurability problem has substantive rather than purely technical content in so far as it is the mathematical consequence of the impossibility of reconciling aggregation and independence in the continuum.

4.2. Piecewise uniformly continuous functions on T and $\overline{T}$ coincide

Here we show that piecewise uniformly continuous functions can be 'transferred' back and forth between the continuum and the discrete model. This transfer is essentially unique and preserves integration.

4.2.1. Transfers between $\mathcal{C}$ and $\overline{\mathcal{C}}$

The *discretization* of $\bar{f} \in \overline{\mathcal{C}}$ is the function $f : T \to \mathbb{R}$ given by $f(t) = \bar{f}(t)$, $\forall t \in T$. In this case, we also refer to $\bar{f}$ as an *extension* of f. Starting with $\bar{f} \in \overline{\mathcal{C}}$, its discretization f must be in $\mathcal{C}$. Conversely, given $f \in \mathcal{C}$ and a component interval $(a_{m-1}, a_m) \cap T$, f has a unique continuous extension to the interval $[a_{m-1}, a_m]$ in the continuum.[26] This defines a unique piecewise uniformly continuous extension $\bar{f} : \overline{T} \to \mathbb{R}$, except possibly at the end points a_m, $m = 1, \ldots, M$, where we set $\bar{f}$ arbitrarily. The resulting $\bar{f}$ is uniquely defined up to a finite set of points (which has measure zero under both λ and $\bar{\lambda}$). We thus have:

Theorem 9. *For any $\bar{f} \in \overline{\mathcal{C}}$, its discretization f belongs to $\mathcal{C}$. Conversely, any $f \in \mathcal{C}$ has an essentially unique extension $\bar{f} \in \overline{\mathcal{C}}$: if $\bar{f}_1$, $\bar{f}_2$ are two such extensions, then $\bar{f}_1(t) = \bar{f}_2(t)$ for $t \in \overline{T}$ except for at most finitely many points.*

4.2.2. Transfer preserves integration

Theorem 9 establishes a natural and essentially one-to-one correspondence between piecewise uniformly continuous functions in the continuum and the discrete model. The next result shows that this correspondence preserves integration:

Theorem 10. *For any $\bar{f} \in \overline{\mathcal{C}}$,*

$$\int_a^b \bar{f}\, \mathrm{d}\bar{\lambda} = \int_a^b f\, \mathrm{d}\lambda \quad \text{on any interval } [a, b].$$

4.2.3. How rich is $\overline{\mathcal{C}}$?

The class of piecewise uniformly continuous functions is clearly an important class that includes all continuous functions and all step functions that are constant on intervals. In a sense, $\overline{\mathcal{C}}$ is as rich a class of functions on the continuum as one can hope for. The only (bounded) functions tractable enough in the continuum are those that can be integrated. As we pointed out earlier, the integrability of (a bounded function) $\bar{g} : \overline{T} \to \mathbb{R}$ implies

$$\inf_{\bar{f} \in \overline{\mathcal{C}}} \int_{\overline{T}} \left| \bar{g}(t) - \bar{f}(t) \right| \mathrm{d}\bar{\lambda} = 0. \tag{16}$$

[26] See, for instance, Royden, 1968, Proposition 11, p. 136.

Thus, in the continuum the only function that can be used in practice are those that are approximately piecewise uniformly continuous in the sense of Eq. (16). Theorem 9 may be loosely interpreted to mean that, up to closure, the discrete model captures anything we can deal with in the continuum.

5. Related literature

I briefly discuss the differences with three closely related approaches.

5.1. The sampling approach

Bewley (1986) and Guesnerie (1981, 1995) proposed that one defines the integral as the limiting average over a randomly drawn infinite sample of agents $\{t_1, \ldots\}$.

One way to think of our discretizing sequence is that it represents a typical sample of agents drawn from the continuum. The approach proposed in this paper differs fundamentally from the sampling approach, however: We choose a *fixed set of agents* T as our primitive. We want the same set to work *simultaneously* for all functions and probability distributions we build on it subsequently. In the sampling approach proposed by Bewley and Guesnerie, one draws a random sample from the continuum each time an integral is to be evaluated. The underlying model then is the continuum, and so the difficulties this model raises are not circumvented.

5.2. Countable models

Another approach, pioneered by Feldman and Gilles (1985), is to take an infinite sequence of agents $\{t_1, \ldots\}$ and define a density charge on it. The major difference with the discrete model developed in this paper is that the latter preserves key features of the continuum. The metric and interval structures are preserved, as well as piecewise uniformly continuous functions and their integrals.[27]

The fact that our discrete model is continuum-like means that it enjoys many of the features that make continuum models tractable and convenient to use in economic applications. It also means that intuitions and insights can be translated back and forth between continuum and discrete models, using whichever is more convenient for the specific application at hand.

5.3. Linear methods

Uhlig (1996) proposed the use of the Pettis integral as a way to formulate the law of large numbers in the continuum. Using this approach, Uhlig (1996) shows that the Pettis integral of an i.i.d. process is a random variable that equals the population mean almost surely.

[27] Theorems 9 and 10 provide a one-to-one mapping between piecewise uniformly continuous (including step) functions in the continuum and the discrete model that preserves integration.

Al-Najjar (1995) showed that stochastic processes on the continuum can be decomposed into aggregate and idiosyncratic components, and that the limiting average on a typical sample converges to the aggregate component of the process. The approach common to both papers is to convert the original stochastic process to an L^2-valued mapping on $\overline{T}$ (by viewing each random variable as a vector in L^2). See Section 3.5 for details.

A difficulty with this formulation is that once a stochastic process is converted to a L^2-valued mapping, the underlying state space no longer plays any role in the analysis. Indeed, the definition of the Pettis integral has no immediate interpretation in terms of sample realizations, limiting averages, and so forth. Thus, although the Pettis integral seems to provide the intuitively correct answer, its interpretation as a law of large numbers remained problematic. By contrast, the present paper produces a strong law of large numbers in the traditional sense: we consider the sample path at each state and assert that the integral of this path equals the population mean almost surely.

The discretization approach provides a new interpretation of the Pettis integral as a law of large numbers. Theorem 6 shows that the Pettis integral coincides almost surely, state-by-state, with the empirical frequencies of *any* discretization of the continuum. As the examples in Section 3.6 show, this interpretation is possible only because of the structure introduced in this paper (e.g., finitely additive density charges, the transfer between the discrete and continuum models, etc.).

6. Application: mechanism design and public goods in large economies

In this final section, I illustrate the framework of this paper by applying it to a mechanism design problem in a large economy with private information. Rob (1989) and Mailath and Postlewaite (1990) considered the problem of externalities and public goods in a sequence of finite, but increasingly large models of agents.

The goal of this section is to use this familiar economic context to "test" whether our framework makes sense. In particular, I show that that this mechanism design problem has a straightforward formulation in a large, atomless public good economy. I then prove a sharp inefficiency result using the concept of 'influence' introduced in Al-Najjar and Smorodinsky (1996).

6.1. The primitives

The space of agents is T. Each agent has a type represented by a valuation for the public good in a finite set $S = \{0 < s_1 < \cdots < s_M\}$. We use as state space the set $\Omega = S^T$. Agents' random types are represented by a function $s : T \times \Omega \to S$. A '$\tilde{}$' will denote a random variable when explicit reference to ω is suppressed. As before, the σ-algebra Σ used is the one generated by all sets of the form $\{\tilde{s}(t) = s\}$, $s \in S$, $t \in T$.

Provision of the public good is modeled as a function $\delta : \Omega \to \{0, 1\}$, where the public good is provided iff $\tilde{\delta} = 1$.[28] Given a transfer $c \in \mathbb{R}$ and a provision outcome $\tilde{\delta}$, the payoff of an agent of type s is $s\tilde{\delta} - c$.

We assume that types are independent; that there is α such that $P\{\tilde{s}(t) = 0\} > \alpha > 0$ for every t; and that $t \mapsto P\{\tilde{s}(t) = s\}$ belongs to $\mathcal{C}$ for every $s \in S$.

6.2. *Defining mechanisms*

Appealing to the Revelation Principle, we focus on direct revelation mechanisms where each agent reports his type (incentive compatibility constraints are introduced later). A mechanism in this context has two components. First, a mechanism includes a *provision function* $\delta : \Omega \to [0, 1]$. We make the obvious assumption that

$$\tilde{\delta} \text{ is a random variable,}^{29} \tag{17}$$

a requirement also needed in finite models. Second, the mechanism includes a *transfer function*:

$$c : T \times \Omega \to \mathbb{R},$$

so $c(t, \omega)$ is the transfer paid by agent t in state ω. At a minimum, we need the average expected transfer, $\int Ec(t, \omega)\, \mathrm{d}\lambda$, to be well-defined. We thus require that:

$$\text{for every agent } t, \quad \tilde{c}(t) \text{ is a random variable,} \tag{18}$$

so the expectation $E\tilde{c}(t)$ is meaningful. We also require that:

$$\text{the function } t \mapsto E\tilde{c}(t) \text{ belongs to } \mathcal{F}, \tag{19}$$

ensuring that $\int E\tilde{c}(t)\, \mathrm{d}\lambda$ is well-behaved.

Conditions (18) and (19) are rather weak, and they are obviously needed. Much less obvious are the implications of the phenomenon of ‘disappearing mass’ discussed in Section 2.4.1. There we displayed a simple example of a function f that is strictly positive, yet $\int f(t)\, \mathrm{d}\lambda = 0$ (Example 1). To appreciate the implications of this example on the problem of defining efficient mechanisms, let $c'(t, \omega) \equiv c(t, \omega) - f(t)$. Since f is strictly positive, every agent is strictly better off under $\tilde{c}'$ relative to $\tilde{c}$. But since $\int f(t)\, \mathrm{d}\lambda = 0$, the aggregate payment collected under $\tilde{c}'$ is the same as under $\tilde{c}$.

This example points out that without additional restrictions, optimal mechanisms never exist, since any mechanism can be improved by using a function like f to generate a ‘money pump’ that increases every agent’s payoff. This, perhaps, is one reason why models with a countable number of agents and finitely additive measures received relatively little interest in the literature on large economies. Using the structure developed in this paper, however, this problem is easily remedied by requiring:

$$\text{the function } t \mapsto c(t, \omega) \text{ belongs to } \mathcal{F}, \quad P\text{-a.s.} \tag{20}$$

We can now state our formal definition of a mechanism:

[28] The choice of S^T as state space here is not without loss of generality as it requires that public good provision and transfers to agents depend deterministically on the vector of agents’ reports.

[29] That is, $\delta : \Omega \to [0, 1]$ is Σ-measurable.

Definition 13. A *mechanism* is a pair $(\tilde{\delta}, \tilde{c})$, where $\tilde{\delta}$ is a provision function and $\tilde{c}$ is a transfer function such that conditions (17)–(20) hold.

An obvious requirement is that the expected value of aggregate transfer, $\int_a^b c(t,\omega)\,\mathrm{d}\lambda$, is equal to the integral of the expected transfers:

$$E\int_a^b c(t,\omega)\,\mathrm{d}\lambda = \int_a^b Ec(t,\omega)\,\mathrm{d}\lambda, \quad \text{for every } [a,b]. \tag{21}$$

This amounts to a version of Fubini's theorem. In general, Fubini's Theorem may fail for finitely additive probabilities (see, for instance, Marinacci, 1997). The following lemma shows that Fubini's Theorem holds for mechanisms satisfying the rather mild requirements in Definition 13.

Lemma 1. *Let* $c : T \times \Omega \to \mathbb{R}$ *be any bounded function satisfying conditions* (18)–(20). *Then Eq.* (21) *holds on every interval* $[a,b]$.

6.3. Individual rationality and incentive compatibility

The conditions of (interim) individual rationality and incentive compatibility can be formulated here in a way that is virtually identical to finite-agent mechanism design problems:

Definition 14. The mechanism $(\tilde{\delta}, \tilde{c})$ is *individually rational* (*IR*) if

$$sE\big(\tilde{\delta} \mid \tilde{s}(t) = s\big) - E\big(\tilde{c}(t) \mid \tilde{s}(t) = s\big) \geqslant 0 \quad \text{for all } t \in T \text{ and } s \in S. \tag{22}$$

Definition 15. The mechanism $(\tilde{\delta}, \tilde{c})$ is *incentive compatible* (*IC*) if for every $t \in T$ and $s, s' \in S$,

$$sE\big(\tilde{\delta} \mid \tilde{s}(t) = s\big) - E\big(\tilde{c}(t) \mid \tilde{s}(t) = s\big) \geqslant sE\big(\tilde{\delta} \mid \tilde{s}(t) = s'\big) - E\big(\tilde{c}(t) \mid \tilde{s}(t) = s'\big). \tag{23}$$

6.4. Inefficiency in large public good economies with private information

We now consider the problem where producing the public good entails a per capita cost of $\beta > 0$. One way to formulate feasibility is to require that budget balance holds ex post:

$$\int_0^1 c(t,\omega)\,\mathrm{d}\lambda \geqslant \beta\delta(\omega), \quad P\text{-a.s.} \tag{24}$$

A weaker requirement is that budget balance holds on average, which we obtain by taking expectations of both sides of Eq. (24):

$$E\int_0^1 c(t,\omega)\,\mathrm{d}\lambda \geqslant \beta E\tilde{\delta}. \tag{25}$$

We refer to this condition as ex ante feasibility (see Mailath and Postlewaite, 1990) and interpret it as reflecting the availability of a risk-neutral lender.

We can now state our main inefficiency result:

Theorem 11. *For any per capita cost $\beta > 0$, the probability of provision of the public good is zero*:

$$P\{\omega\colon\ \delta(\omega) = 1\} = 0,$$

under any mechanism $(\tilde{\delta}, \tilde{c})$ satisfying incentive compatibility, individual rationality, and ex ante feasibility.

The result is an atomless, large-economy version of the asymptotic inefficiency results in Rob (1989) and Mailath and Postlewaite (1990). Its proof appeals to a large-economy generalizations of the concepts of "influence" and "pivotalness" introduced by Al-Najjar and Smorodinsky (1996) and which are of independent interest.

6.5. Influence and non-pivotalness theorem in large economies

Rearrange the IC constraint, Eq. (23), to obtain:

$$E\big(\tilde{c}(t) \mid \tilde{s}(t) = s\big) - E\big(\tilde{c}(t) \mid \tilde{s}(t) = s'\big) \leqslant s \underbrace{\big[E\big(\tilde{\delta} \mid \tilde{s}(t) = s\big) - E\big(\tilde{\delta} \mid \tilde{s}(t) = s'\big)\big]}_{V}. \tag{26}$$

The term V above represents agent t's influence on the probability of provision as he changes his report from s to s'. Following Al-Najjar and Smorodinsky (1996), define the *influence* of agent t relative to $\tilde{\delta}$ as:

$$V(t, \tilde{\delta}) = \max_{s,s' \in S} \big| E\big(\tilde{\delta} \mid \tilde{s}(t) = s\big) - E\big(\tilde{\delta} \mid \tilde{s}(t) = s'\big)\big|.$$

Equation (26) tells us that an agent's expected contribution is bounded by his influence, $V(t, \tilde{\delta})$. Note that one can always find $\tilde{\delta}$ under which $V(t, \tilde{\delta})$ is high for a particular agent t. For example, a 'dictatorial' $\tilde{\delta}$ that ignores the reports of all but a particular agent $\bar{t}$ is an example where one agent has large influence. The key point is that *average* influence must be zero for any choice of $\tilde{\delta}$.

Theorem 12. *For any $\delta : \Omega \to [0, 1]$,*

$$\int_0^1 V(t, \tilde{\delta})\, \mathrm{d}\lambda = 0.$$

The proof, found in Appendix A, adapts the argument in Al-Najjar and Smorodinsky to the discrete model T.[30]

[30] Since every bounded function is integrable, the integral $\int_0^1 V(t, \tilde{\delta})\, \mathrm{d}\lambda$ is always well defined. We are not asserting, however, that the function $t \mapsto V(t, \tilde{\delta})$ belongs to $\mathcal{F}$.

Proof of Theorem 11. Condition IR, Eq. (22), evaluated at $s = 0$ implies that $E(\tilde{c}(t) \mid \tilde{s}(t) = 0) = 0$. Averaging over the types of agent t, Eq. (26) implies

$$E\tilde{c}(t) \leqslant s_M V\big(t, \tilde{\delta}\big). \tag{27}$$

This says that the expected contribution of any agent is bounded above by a constant times the maximum influence of that agent under $\tilde{\delta}$. Integrating over T, we obtain

$$\int_0^1 E\tilde{c}(t)\,\mathrm{d}\lambda \leqslant s_M \int_0^1 V\big(t, \tilde{\delta}\big)\,\mathrm{d}\lambda. \tag{28}$$

From Eq. (28) and Theorem 12, we have $\int_0^1 E\tilde{c}(t)\,\mathrm{d}\lambda = 0$. From ex ante feasibility, Eq. (25),

$$P\big\{\omega\colon \delta(\omega) = 1\big\} \equiv E\tilde{\delta} \leqslant \frac{1}{\beta}\int_0^1 E\tilde{c}(t)\,\mathrm{d}\lambda = 0. \qquad \square$$

6.6. *Mechanism design and the law of large numbers*

The law of large numbers is central to formulating mechanism design problems like the one described above. As an illustration consider the problem of verifying whether feasibility is satisfied. Such verification requires computing the integral:

$$\int_T c(t, \omega)\,\mathrm{d}\lambda. \tag{29}$$

Our framework ensures that the value of this integral is well-behaved provided that condition (20) is satisfied. This condition requires that $\tilde{c}$ has well-behaved distribution and that $\int_T c(t, \omega)\,\mathrm{d}\lambda$ is the limit of integrals in large finite models.

The contrast with the continuum should be emphasized: there, whether or not an integral like that in expression (29) is meaningful depends crucially on the mechanism being considered. An important class of mechanisms for which the integral in (29) *cannot* be defined is that of anonymous mechanisms. In such mechanisms, the transfer of an agent depends only on his type, not his label. With an i.i.d. distribution of types, for instance, the problem of defining the integral in (29) reduces to that of integrating a typical draw of an i.i.d. process, which is not meaningful in the continuum.

This is not to say that *no* mechanism is well defined in the continuum. Rather, the question is: *How severe a restriction do we need to impose on the class of allowable mechanisms?* The example of anonymous mechanisms indicates that the restrictions in the continuum have to be very severe indeed. In the discrete model we also need to restrict the class of allowable mechanisms, namely by requiring conditions (17)–(20). Of these, conditions (17) and (18) are innocuous, and are needed in the continuum as well as in mechanism design problems with a finite number of agents. Conditions (19) and (20), on the other hand, considerably weaken the measurability requirement that one would have to impose in the continuum.

It may be possible to find an *indirect* way to use the continuum to model interesting mechanism design problems in large economies (e.g., using the Pettis integral). But any such approach will have to forego references to what happens at each state. The discretization approach presented here has the advantage of being similar enough to the continuum, yet consistent with a straightforward large-economy formulation of finite mechanism design problems.

Acknowledgments

This paper is dedicated to Marcel K. Richter. It will be reprinted in the forthcoming volume *Foundations of Economic Theory, Essays in Honor of Marcel K. Richter*, Springer-Verlag.

I am grateful to an Associate Editor and a referee for their valuable feedback. I also thank Larry Blume, Jeff Ely, Massimo Marinacci, Rafi Rob, and Asher Wolinsky for their valuable comments, and Lineu Vargas for proofreading earlier versions.

Appendix A

Proof of Theorem 1. Let l_∞ denote the set of all bounded sequences with the supremum norm. That is, for $x = (x_1, \ldots) \in l_\infty$, $\|x\| = \sup_n x_n$. We use the fact (see Rao and Rao, 1983, p. 39–41) that there exists a function $T : l_\infty \to \mathbb{R}$ that is:

(1) *linear*: $T(cx + dy) = cT(x) + dT(y)$;
(2) *non-negative*: $T(x) \geqslant 0$ if $\inf_n x_n \geqslant 0$;
(3) *preserves identity*: $T(1, 1, 1, \ldots) = 1$;
(4) *translation invariant*: $T(x_1, x_2, \ldots) = T(x_2, x_3, \ldots)$.

Any such function is known as a Banach limit. The above reference shows that

$$\liminf_n x_n \leqslant T(x) \leqslant \limsup_n x_n;$$

in particular, $T(x) = \lim_{n\to\infty} x_n$ whenever the limit exists. Thus, one may think of T as a generalized limit.

Fix one such function T. For an arbitrary set $A \subset X$, define

$$\lambda(A) = T\big(\lambda_1(A), \lambda_2(A), \ldots\big).$$

It is clear that λ thus defined is a finitely additive probability on X. Furthermore, Eq. (2) must hold since T assigns to every convergent sequence its limit. □

Proof of Theorem 2. We drop reference the interval $[a, b]$ in what follows to simplify notation.

1. That F is non-decreasing follows from the additivity of λ, so it only remains to show that F is right-continuous. Suppose that r is a point at which F is not right-continuous. Using the fact that F is non-decreasing, there is a sequence $r_n \downarrow r$ such that $F(r_n) \downarrow F^+ > F(r) \geqslant \sup_{r'<r} F(r') \equiv F^-$.

By the additivity of λ and Eq. (4), in Definition 4, we have $F(r) \equiv \lambda\{f(t) \leqslant r\} \geqslant \lambda\{f(t) = r\} + \lambda\{f(t) \leqslant r'\} \equiv (F^+ - F^-) + F(r')$ for any $r' < r$. For any $\epsilon > 0$, there is $r' < r$ such that $F^- - F(r') < \epsilon$. Thus, given any $\epsilon > 0$ and choosing r' as in the last sentence, we have $F(r) \geqslant (F^+ - F^-) + F^- - \epsilon = F^+ - \epsilon$, so $F(r) = F^+$, which is a contradiction.

2. *Change of variables.* Since f is bounded, we may restrict attention to some fixed interval $[r^{\min}, r^{\max}]$, so $F(r^{\min}) = 0$ and $F(r^{\max}) = 1$. Choose $\{r_1, \ldots, r_M\}$ to be equally spaced points such that $r^{\min} = r_0 < \cdots < r_M = r^{\max}$ and consider the step function $g_M : [r^{\min}, r^{\max}] \to [r^{\min}, r^{\max}]$ that takes the constant value r_m on the interval $(r_m, r_{m+1}]$.

Clearly, the sequence of simple functions $\{g_M(f(\cdot))\}$ converges uniformly to f, so by the definition of the integral (Definition 3), we have $\lim_{M\to\infty} \int_a^b g_M(f(\cdot))\,\mathrm{d}\lambda = \int_a^b f(\cdot)\,\mathrm{d}\lambda$.

For any interval $(r_m, r_{m+1}]$, $m = 1, \ldots, M$, we have:

$$\frac{1}{|a-b|}\int_a^b \chi_{(r_m, r_{m+1}]}\big(f(t)\big)\,\mathrm{d}\lambda(t) = \int_{\mathbb{R}} \chi_{(r_m, r_{m+1}]}(r)\,\mathrm{d}F(r).$$

By linearity of the integral, this extends to all step functions g_M constructed in the last paragraph:

$$\frac{1}{|a-b|}\int_a^b g_M\big(f(t)\big)\,\mathrm{d}\lambda = \int_{\mathbb{R}} g_M\,\mathrm{d}F(r).$$

Taking limits of both sides as M goes to infinity establishes the claim.

3. *The integral is the limit of integrals in finite models.* If $f \in \mathcal{F}$ then, by Definition 5, $F_N(\cdot) \to F(\cdot)$ weakly. Clearly, $\int_a^b f\,\mathrm{d}\lambda_N = \int_{\mathbb{R}} r\,\mathrm{d}F_N(r)$, and by part (2) above, $\int_a^b f\,\mathrm{d}\lambda = \int_{\mathbb{R}} r\,\mathrm{d}F(r)$. Weak convergence then implies that $\lim_{N\to\infty} \int_{\mathbb{R}} r\,\mathrm{d}F_N(r) = \int_{\mathbb{R}} r\,\mathrm{d}F(r)$. □

Proof of Theorem 3. It is enough to prove the claim for the case in which f is uniformly continuous. We prove that $\bar{f}$ has asymptotic frequencies; this then implies that f does too. Since $\bar{f}$ is continuous, we can find a countable set of disjoint (possibly empty) open intervals $\{I_k\}_{k=1}^{\infty}$ such that $I = \{t\colon\ \bar{f}(t) > r\} = \bigcup_{k=1}^{\infty} I_k$. Clearly, the boundary of I, denoted ∂I, consists of the end points of the intervals I_k, and is thus countable. In particular, $\bar{\lambda}(\partial I) = 0$.

By definition,

$$1 - F_N(r) \equiv \lambda_N(I) \quad \text{and} \quad 1 - \overline{F}(r) \equiv \bar{\lambda}(I).$$

Using a standard characterization of weak convergence (see, for instance, Shiryayev (1984), Theorem 1, p. 309), the sequence of measures $\{\lambda_N\}$ converges weakly to $\bar{\lambda}$ if and only if for every measurable set A with $\bar{\lambda}(\partial A) = 0$, $\lambda_N(A) \to \bar{\lambda}(A)$. In our case weak convergence implies that $\lambda_N(I) \to \bar{\lambda}(I)$. This implies $F_N(r) \to \overline{F}(r) = F(r)$ for every $r \in \mathbb{R}$, and the result follows.[31] □

[31] This proves more than just weak convergence, which requires only convergence for r's that are continuity points of F.

Proof of Theorem 10. The result follows from the change of variables formula in Theorem 2. □

To prove Theorem 4 we note that the assumption $\mu \in \mathcal{C}$ is unnecessary; the result (stated below as Theorem 4*) holds for $\mu \in \mathcal{F}$. We chose the more restrictive assumption to make it easier to compare this result with those for the continuum (Theorems 5 and 6 for which functions in $\mathcal{F}\backslash\mathcal{C}$ do not make sense).

An example of a process with $\mu \in \mathcal{F}\backslash\mathcal{C}$ is the following: Consider an i.i.d. process on T that takes the values $1/3$ and $2/3$ with equal probability. Theorem 4 guarantees that almost every realization has the property that its integral is 0.5 on every subinterval $[a,b]$. Let μ be one such realization, and note that $\mu \in \mathcal{F}\backslash\mathcal{C}$ (this follows from Theorem 8). The independent process with mean μ is the desired example. It is interesting to note that integrals of sample realization on intervals cannot separate the process with mean μ from the 50-50 i.i.d. process.

Theorem 4*. *Suppose that P is independent and $\mu \in \mathcal{F}$. Then for P-almost every realization ω, on every interval $[a,b]$, the integral of the sample realization equals the integral of the expectations.*

Proof. Fix an interval $[a,b]$. Since $\mu \in \mathcal{F}$, by Theorem 2,

$$\int_a^b \mu(t)\,\mathrm{d}\lambda = \lim_{N\to\infty} \frac{1}{\#T_N \cap [a,b]} \sum_{T_N\cap[a,b]} \mu(t).$$

Using the version of the strong law of large numbers in Shiryayev (1984, Theorem 2, p. 364), there is a set $\overline{\Omega}_{[a,b]} \subset \Omega$ with $P(\Omega_{[a,b]}) = 1$ such that for every $\omega \in \Omega_{[a,b]}$

$$\lim_{N\to\infty} \frac{1}{\#T_N \cap [a,b]} \sum_{T_N\cap[a,b]} s(t,\omega) = \lim_{N\to\infty} \frac{1}{\#T_N \cap [a,b]} \sum_{T_N\cap[a,b]} \mu(t)$$

which proves the claim for a single interval $[a,b]$.

Define $\Omega' = \cap\Omega_{[a,b]}$, where the intersection is taken over all intervals $[a,b]$ with rational endpoints. Since there are countably many such intervals, we have $P(\Omega') = 1$, and the law of large numbers is satisfied simultaneously on every such interval in T.

To prove the general claim, let $[a,b] \subset [0,1]$ be an arbitrary interval where a, b are irrational. Consider any sequences of rational numbers $\{a_n\} \uparrow a$ and $\{b_n\} \downarrow b$ and fix any state $\omega \in \Omega'$. Belonging to Ω' implies that

$$\int_{a_n}^{b_n} s(t,\omega)\,\mathrm{d}\lambda = \int_{a_n}^{b_n} \mu(t)\,\mathrm{d}\lambda$$

for every n. To simplify notation, write $A_n = [a_n, b_n] - [a,b]$. Then,

$$\int_{a_n}^{b_n} s(t,\omega)\,\mathrm{d}\lambda = \int_a^b s(t,\omega)\,\mathrm{d}\lambda + \int_{A_n} s(t,\omega)\,\mathrm{d}\lambda \quad \text{and}$$

$$\int_{a_n}^{b_n} \mu(t)\,\mathrm{d}\lambda = \int_a^b \mu(t)\,\mathrm{d}\lambda + \int_{A_n} \mu(t)\,\mathrm{d}\lambda;$$

substituting, we obtain:

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_{a_n}^{b_n} \mu(t)\,\mathrm{d}\lambda - \int_{A_n} s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \mu(t)\,\mathrm{d}\lambda + \int_{A_n} \mu(t)\,\mathrm{d}\lambda - \int_{A_n} s(t,\omega)\,\mathrm{d}\lambda.$$

The claim now follows by noting that the term $\int_{A_n} \mu(t)\,\mathrm{d}\lambda - \int_{A_n} s(t,\omega)\,\mathrm{d}\lambda$ goes to zero as $\lambda(A_n) = a_n - a + b - b_n \to 0$. □

Proof of Theorem 5. By Theorem 10, we may replace $\int_a^b \bar{\mu}(t)\,\mathrm{d}\bar{\lambda}$ in Eq. (9) by $\int_a^b \mu(t)\,\mathrm{d}\lambda$, so this equation is equivalent to:

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \mu(t)\,\mathrm{d}\lambda \quad \overline{P}\text{-a.s.} \tag{A.1}$$

Since λ puts unit mass on T, any pair of states ω, ω' that agree on T must also satisfy $\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b s(t,\omega')\,\mathrm{d}\lambda$. This means that the random variable $\omega \mapsto \int_a^b s(t,\omega)$ on $(\Omega, \overline{\Sigma})$ is in fact Σ-measurable. Thus, Eq. (A.1) also holds P-a.s., and it is thus equivalent to:

$$\int_a^b s(t,\omega)\,\mathrm{d}\lambda = \int_a^b \mu(t)\,\mathrm{d}\lambda \quad P\text{-a.s.}$$

This, of course, is true by Theorem 4. □

Proof of Theorem 6. Let $\tilde{x}$ be any random variable. Then

$$\left(\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right) = \left([\tilde{x} - E\tilde{x}] + E\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right)$$
$$= \left(\tilde{x} - E\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right) + \left(E\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right)$$

By Theorem 4, $\int_a^b \tilde{s}(t)\,\mathrm{d}\lambda$ is constant almost surely, so its covariance with any random variable is zero. This and $E[\tilde{x} - E\tilde{x}] = 0$ imply:

$$\left(\tilde{x} - E\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right) = \operatorname{cov}\left(\tilde{x} - E\tilde{x}, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right) + E[\tilde{x} - E\tilde{x}]E\int_a^b \tilde{s}(t)\,\mathrm{d}\lambda = 0.$$

On the other hand, from the definition of the Pettis integral, and again using Theorem 4,

$$\left(E\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right) = E\tilde{x}\,E\int_a^b \tilde{s}(t)\,\mathrm{d}\lambda = E\tilde{x}\int_a^b \mu(t)\,\mathrm{d}\lambda.$$

In summary, we have

$$\left(\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right) = E\tilde{x}\int_a^b \mu(t)\,\mathrm{d}\lambda. \qquad (*)$$

Next we compute $\int_a^b (\tilde{x} \mid \tilde{s}(t))\,\mathrm{d}\lambda$. It is convenient to write

$$\int_a^b \big(\tilde{x} \mid \tilde{s}(t)\big)\,\mathrm{d}\lambda = \int_a^b \big(\tilde{x} \mid [\tilde{s}(t) - E\tilde{s}(t)] + E\tilde{s}(t)\big)\,\mathrm{d}\lambda.$$

Since the $\tilde{s}(t)$'s are independent, the $[\tilde{s}(t) - E\tilde{s}(t)]$'s are orthogonal with bounded L^2-norm. Then, for any enumeration of $\{t_1, \ldots\}$ we must have $\sum_n^\infty (\tilde{x} \mid [\tilde{s}(t_n) - E\tilde{s}(t_n)])^2 < \infty$ (see Dunford and Schwartz, 1958, IV.4.10, p. 251). In particular, for every $\epsilon > 0$, $(\tilde{x} \mid \tilde{s}(t) - E\tilde{s}(t)) < \epsilon$ for all but finitely many values of t. Then the sequence of functions $\{f_N\}$, where $f_N = (\tilde{x} \mid \tilde{s}(t) - E\tilde{s}(t))$ if this value exceeds $1/N$ and $f_N = 0$ otherwise, converges to the function $t \mapsto (\tilde{x} \mid \tilde{s}(t) - E\tilde{s}(t))$ uniformly. Furthermore, $\int_a^b f_N\,\mathrm{d}\lambda = 0$. By the definition of the integral (Definition 3), we have:

$$\int_a^b \big(\tilde{x} \mid \tilde{s}(t) - E\tilde{s}(t)\big)\,\mathrm{d}\lambda = 0.$$

On the other hand, $(\tilde{x} \mid E\tilde{s}(t)) = E\tilde{x}\,E\tilde{s}(t)$. Combining these facts, we have

$$\int_a^b \big(\tilde{x} \mid \tilde{s}(t)\big)\,\mathrm{d}\lambda = E\tilde{x}\int_a^b E\tilde{s}(t)\,\mathrm{d}\lambda. \qquad (**)$$

Combining $(*)$ and $(**)$, we obtain

$$\left(\tilde{x} \,\Big|\, \int_a^b \tilde{s}(t)\,\mathrm{d}\lambda\right) = \int_a^b \big(\tilde{x} \mid \tilde{s}(t)\big)\,\mathrm{d}\lambda$$

which is the desired conclusion, Eq. (11). □

Proof of Theorem 7. The result follows since, by Theorem 4, Eq. (12) holds for each θ that has positive probability. □

Proof of Theorem 9. Clearly, the restriction of a piecewise uniformly continuous function $\bar{f} : \overline{T} \to \mathbb{R}$ is a function that is piecewise uniformly continuous on T.

In the other direction, fix a function f that is piecewise uniformly continuous on T. Consider any interval $(a_m, a_{m+1}) \cap T$ on which f is uniformly continuous. Then f has a unique continuous extension to the closure of this interval in $\overline{T}$, namely $[a_m, a_{m+1}]$. Repeating this process for all m, we obtain a uniquely defined function $\bar{f}$ on the set $\overline{T} - \{a_1, \ldots, a_M\}$. The result follows by defining $\bar{f}$ arbitrarily on the finite set $\{a_1, \ldots, a_M\}$. □

Proof of Theorem 8. Let g be a step function (i.e., constant on the component intervals (a_m, a_{m+1}). Let $s(t, \omega)$ be a typical sample realization (i.e., satisfies the conclusion of Theorem 4). On each component A_m that has positive mass and where g assumes the value g_m, we have

$$\frac{1}{\lambda(A_m)} \int\limits_{A_m} \left|s(t,\omega) - g(t)\right| \mathrm{d}\lambda = 0.5 * \left[|1 - g_m| + |g_m|\right] \geqslant 0.5.$$

Clearly, $\int_T |s(t,\omega) - g(t)| \,\mathrm{d}\lambda \geqslant 0.5$.

Next, consider now any function $f \in \mathcal{C}$. Clearly, for every $\epsilon > 0$, we can find a step function g that is uniformly ϵ-close to f. Then

$$\begin{aligned} 0.5 &\leqslant \int\limits_T \left|s(t,\omega) - g(t)\right| \mathrm{d}\lambda \leqslant \int\limits_T \left[\left|s(t,\omega) - f(t)\right| + \left|f(t) - g(t)\right|\right] \mathrm{d}\lambda \\ &= \int\limits_T \left|s(t,\omega) - f(t)\right| \mathrm{d}\lambda + \int\limits_T \left|f(t) - g(t)\right| \mathrm{d}\lambda \leqslant \int\limits_T \left|s(t,\omega) - f(t)\right| \mathrm{d}\lambda + \epsilon. \end{aligned}$$

That is, $\int_T |s(t,\omega) - f(t)| \,\mathrm{d}\lambda \geqslant 0.5 - \epsilon$, for every $\epsilon > 0$, and the claim of the theorem follows. □

Lemma A.1. *Suppose that $f : T \to \mathbb{R}$ is any function in $\mathcal{F}$ such that $\lambda\{t\colon f(t) < 0\} = 0$ and $\lambda\{t\colon f(t) > 0\} > 0$. Then $\int_{[0,1]} f \,\mathrm{d}\lambda > 0$.*

Proof. In what follows, references to $[a, b]$ are suppressed and the interval is assumed to be $[0, 1]$. Condition (4) states that for every r,

$$\lambda\big(f^{-1}(r)\big) = \inf_{r' > r} F(r') - \sup_{r' < r} F(r').$$

The assumption $\lambda\{t\colon f(t) < 0\} = 0$ implies $\sup_{r'<0} F(r') = 0$. Further, $\lambda\{t\colon f(t) > 0\} > 0$ implies that $\lambda(f^{-1}(0)) < 1$. Thus, for some $r' > 0$, $F(r') < 1$. The result now follows by the change-of-variables part of Theorem 2. □

Lemma A.2. *For any allocation x, the function $\omega \mapsto \int_a^b c(t,\omega)\,\mathrm{d}\lambda$ is Σ-measurable* (*hence a random variable on* (Ω, Σ)).

Proof. Note that $\int_a^b c(t,\omega)\,\mathrm{d}\lambda_N$ is a finite sum for any state ω, any N and any subinterval $[a, b]$. This means that the function $\omega \mapsto \int_a^b c(t,\omega)\,\mathrm{d}\lambda_N$ is Σ-measurable. By the assumption that for P-almost every state ω, the function $\omega \mapsto c(t,\omega)$ belongs to $\mathcal{F}$, Theorem 2 implies that for almost every state,

$$\lim_{N\to\infty} \int_a^b c(t,\omega)\,\mathrm{d}\lambda_N = \int_a^b c(t,\omega)\,\mathrm{d}\lambda.$$

Thus, the sequence of measurable functions $\{\int_a^b c(t,\omega)\,\mathrm{d}\lambda_N\}$ converges almost everywhere to the function $\int_a^b c(t,\omega)\,\mathrm{d}\lambda$, implying that this is a Σ-measurable function as required. □

Proof of Lemma 1. Note that $\int_a^b c(t,\omega)\,\mathrm{d}\lambda_N$ is a finite sum for any state ω, any N and any subinterval $[a,b]$. Thus,

$$E\int_a^b c(t,\omega)\,\mathrm{d}\lambda_N = \int_a^b Ec(t,\omega)\,\mathrm{d}\lambda_N.$$

By our assumption that $t \mapsto Ec(t,\omega)$ belongs to $\mathcal{F}$, we have $\lim_{N\to\infty}\int_a^b Ec(t,\omega)\,\mathrm{d}\lambda_N = \int_a^b Ec(t,\omega)\,\mathrm{d}\lambda$. Taking limits, we obtain:

$$\lim_{N\to\infty} E\int_a^b c(t,\omega)\,\mathrm{d}\lambda_N = \int_a^b Ec(t,\omega)\,\mathrm{d}\lambda.$$

So it only remains to show that

$$\lim_{N\to\infty} E\int_a^b c(t,\omega)\,\mathrm{d}\lambda_N = E\int_a^b c(t,\omega)\,\mathrm{d}\lambda.$$

From the proof of Lemma A.2, we know that the sequence of measurable functions $\{\int_a^b c(t,\omega)\,\mathrm{d}\lambda_N\}$ converges almost everywhere to $\int_a^b c(t,\omega)\,\mathrm{d}\lambda$ (both viewed as functions of ω, of course). Furthermore, this sequence is dominated by a constant since c is bounded by assumption. The claim of the Lemma now follows by applying the Dominated Convergence Theorem. □

The proof of Theorem 12 follows easily from the following lemma, which in turn follows from Theorem 1 in Al-Najjar and Smorodinsky (1996). Essentially, we modify the problem to an equivalent problem with a finite number of agents where we can appeal to the results Al-Najjar and Smorodinsky established for finite agent environments.

Lemma A.3. *For every $\alpha > 0$, there is a number K_α^* such that for any $\delta : \Omega \to [0,1]$, no more than K_α^* agents have influence exceeding α.*

Proof. For a fixed, finite set of K agents, T', define

$$\delta' : \prod_{t\in T'} S \to [0,1]$$

as follows: for a vector of types of agents in T', $\{\bar{s}(t)\colon t \in T'\}$,

$$\tilde{\delta}'\big(\big\{\tilde{s}(t) = \bar{s}(t)\colon t \in T'\big\}\big) = E\big(\tilde{\delta} \bigm| \big\{\tilde{s}(t) = \bar{s}(t)\colon t \in T'\big\}\big).$$

Note that for $t \in T'$, $V(t, \tilde{\delta}) = V(t, \tilde{\delta}')$. Thus, $\tilde{\delta}'$ is a mechanism in an environment with a finite set of agents T', and where each agent has the same influence as in the original problem.

Suppose the claim of the lemma were false. Then there is $\alpha > 0$ such that for every $K > 0$ there is a finite set T' of K agents such that $V(t, \tilde{\delta}) > \alpha$ for every $t \in T'$. From the observation made in the last paragraph, we have a mechanism $\tilde{\delta}'$ in a finite environment T' such that $V(t, \tilde{\delta}') > \alpha$ for agents in T'. In particular, average influence, $1/K \sum_{t \in T'} V(t, \tilde{\delta}') > \alpha$. This is a contradiction with Theorem 1 in Al-Najjar and Smorodinsky (1996) which states that average influence goes to 0 as K goes to infinity, uniformly over all mechanisms and all distribution of signals where $P\{\tilde{s}(t) = s\} > \epsilon$ for every t and s. □

Proof of Theorem 12. Define the function $f_N : T \to \mathbb{R}$ by $f_N(t) = V(t, \tilde{\delta})$ for $t \in T_N$, and $f_N(t) = 0$ otherwise. From the lemma, it follows that given $\alpha > 0$ there is N such that $V(t, \tilde{\delta}) < \alpha$ for every $t \notin T_N$. This implies that the sequence of functions $\{f_N\}$ converges to $t \mapsto V(t, \tilde{\delta})$ uniformly. From the definition of the integral (Definition 3), we have $\int_0^1 f_N(t)\,\mathrm{d}\lambda \to \int_0^1 V(t, \tilde{\delta})\,\mathrm{d}\lambda$ as $N \to \infty$. On the other hand, $\int_0^1 f_N(t)\,\mathrm{d}\lambda = 0$ for every N, and the result follows. □

References

Al-Najjar, N., Smorodinsky, R., 1996. Pivotal players and the characterization of influence. J. Econ. Theory 92, 318–342.

Al-Najjar, N.I., 1995. Decomposition and characterization of risk with a continuum of random variables. Econometrica 63, 1195–1224. Corrigendum: Econometrica 67 (1999), 919–920.

Armstrong, T., Richter, M., 1984. The core–Walras equivalence. J. Econ. Theory 33, 116–151.

Armstrong, T., Richter, M., 1986. Existence of nonatomic core–Walras allocations. J. Econ. Theory 38, 137–159.

Bewley, T., 1986. Stationary monetary equilibrium with a continuum of independently fluctuating consumers. In: Hildenbrand, W., Mas-Colell, A. (Eds.), Contributions to Mathematical Economics: In honor of Gerard Debreu. North-Holland, New York.

Diestel, J., Uhl, J., 1977. Vector Measures. In: Math. Surv., vol. 15. AMS, Providence, RI.

Dunford, N., Schwartz, J., 1958. Linear Operators, Part I. Interscience, New York.

Feldman, M., Gilles, C., 1985. An expository note on individual risk without aggregate uncertainty. J. Econ. Theory 35, 26–32.

Fishburn, P., 1970. Utility Theory for Decision Making. Wiley, New York.

Gilboa, I., Matsui, A., 1992. A model of random matching. J. Math. Econ. 21, 185–197.

Guesnerie, R., 1981. On taxation and incentives: further reflections on the limits of redistribution. École des hautes études en sciences sociales.

Guesnerie, R., 1995. A Contribution to the Pure Theory of Taxation. Cambridge Univ. Press, Cambridge.

Judd, K.L., 1985. The law of large numbers with a continuum of i.i.d. random variables. J. Econ. Theory 35, 19–25.

Khan, M., Sun, Y., 1999. Non-cooperative games on hyperfinite Loeb spaces. J. Math. Econ. 31, 455–492.

Mailath, G., Postlewaite, A., 1990. Asymmetric information bargaining problems with many agents. Rev. Econ. Stud. 57, 351–367.

Marinacci, M., 1997. Finitely additive and epsilon Nash equilibria. Int. J. Game Theory 26, 315–333.

Rao, K., Rao, M., 1983. Theory of Charges. Academic Press, New York.
Rob, R., 1989. Pollution claim settlements under private information. J. Econ. Theory 47, 307–333.
Royden, H.L., 1968. Real Analysis, 2nd Edition. MacMillan, New York.
Shiryayev, A.N., 1984. Probability. Springer-Verlag, New York.
Sun, Y., 1998. A theory of hyperfinite processes: The complete removal of individual uncertainty via exact LLN. J. Math. Econ. 29, 419–503.
Uhlig, H., 1996. A law of large numbers for large economies. Econ. Theory 8, 41–50.
Weiss, E.A., 1981. Finitely additive exchange economies. J. Math. Econ. 8, 221–240.
Werner, J., 1997. Diversification and equilibrium in securities markets. J. Econ. Theory 75, 89–103.