The Bulletin of Symbolic Logic
Volume 2, Number 2, June 1996

HYPERFINITE LAW OF LARGE NUMBERS

YENENG SUN

Abstract. The Loeb space construction in nonstandard analysis is applied to the theory of processes to reveal basic phenomena which cannot be treated using classical methods. An asymptotic interpretation of results established here shows that for a triangular array (or a sequence) of random variables, asymptotic uncorrelatedness or asymptotic pairwise independence is necessary and sufficient for the validity of appropriate versions of the law of large numbers. Our intrinsic characterization of almost sure pairwise independence leads to the equivalence of various multiplicative properties of random variables.

Introduction. In probability theory and its applications, it is usual to study a large number of random variables with low intercorrelation. Since the continuum is commonly used to model the ideal situations for macroscopic phenomena, it is natural to consider a continuum of random variables (simply called a process) with low intercorrelation. In applications (especially in economic applications), *exact* equality between the mean (or distribution) of a sample function and the theoretical mean (or distribution) is obtained by assuming that a continuum of independent and identically distributed random variables satisfies the law of large numbers (see for example, [1], [2], [6], and [10]). This is often referred to as "*aggregation removes individual uncertainty*". There are, however, serious difficulties with this assumption. It has been pointed out by Doob that such a process, typically indexed by a Lebesgue interval, is not measurable and even has no measurable standard modification with respect to the relevant product measure (unless, of course, the common distribution is concentrated at a single point; see [3], p. 67). In fact, as noted in [6], there is a naturally constructed process with independent and identically distributed random variables (indexed by the unit Lebesgue interval) for which it is not true that almost all sample functions are measurable. If one considers a natural extension of the probability measure on the sample space so that the measurability problem disappears, then the set of sample realizations satisfying the property that the expectation or

Received December 20, 1995, and in revised form, February 1, 1996.
1991 *Mathematics Subject Classification*. Primary 03H05, 60F15; Secondary 60A05, 60G07.
The author is indebted to Yu Kiang Leong, Peter Loeb and Ali Khan for many helpful suggestions. The research is partially supported by NUS grant RP3920641.

1079-8986/96/0202-0003/$2.00

distribution of the sample function is the theoretical one has outer measure one and inner measure zero; this set is, therefore, not measurable. Thus there is no meaningful way to claim that almost all sample realizations will have the required properties even for the simplest case considered in [6]. Since the law of large numbers does in fact hold for the discrete case, it follows that the usual concept of a continuum of random variables is not a good model for the ideal limit of a triangular array or, more specifically, of a sequence of random variables with low intercorrelation.

This note outlines our successful treatment of the law of large numbers using hyperfinite index sets for general processes and presents the resulting discovery of many hidden characteristics of those processes. Our results suggest that for many applications, a general process should be indexed using a hyperfinite set endowed with Loeb measure, and in particular, it should be a process measurable with respect to the Loeb space of the internal product of a given hyperfinite internal probability space with another internal probability space. Such a process will often be called a hyperfinite process. We note that the external cardinality of hyperfinite sets in any nonstandard model obtained from an ultrapower construction on the natural numbers is the same as the cardinality of the continuum. That is, a hyperfinite number of random variables is a continuum of random variables.

We now fix some notation. Let T be a hyperfinite set, $\mathcal{T}$ the internal algebra of all internal subsets of T, and λ an internal finitely additive probability measure on $(T, \mathcal{T})$. Let $\big(T, L(\mathcal{T}), L(\lambda)\big)$ be the standardization, i.e., the Loeb space, formed from $(T, \mathcal{T}, \lambda)$; this standard probability space will be the hyperfinite parameter space for the processes to be considered here. Starting with another internal probability space $(\Omega, \mathcal{A}, P)$, we let the Loeb space $\big(\Omega, L(\mathcal{A}), L(P)\big)$ be our sample probability space. For basic properties of Loeb spaces, see [5] and [8].

Note that the internal product space $(T \times \Omega, \mathcal{T} \otimes \mathcal{A}, \lambda \otimes P)$ is also an internal probability space. The corresponding Loeb space is denoted by $\big(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P)\big)$; it will be referred to as the Loeb product space. As usual, a measurable function of two variables of the type considered here is called a process. Given a process f on the Loeb product space, for each $t \in T$, and $\omega \in \Omega$, f_t denotes the function $f(t, \cdot)$ on Ω and f_ω denotes the function $f(\cdot, \omega)$ on T. The functions f_t are usually called the random variables of the process f, while the f_ω form the sample functions of the process.

A real-valued measurable process f on the Loeb product space is said to satisfy the law of large numbers if for $L(P)$-almost all $\omega \in \Omega$,

$$\int_T f_\omega(t)\, dL(\lambda)(t) = \iint_{T \times \Omega} f\, dL(\lambda \otimes P),$$

i.e., for almost all sample realizations $\omega \in \Omega$, the mean of the sample function f_ω on the parameter space is equal to the mean of f as a random variable on the joint space of parameters and samples. For the case that f takes values in some metric space X, we say that f satisfies the law of large numbers in distribution if for $L(P)$-almost all $\omega \in \Omega$, the empirical distribution of the sample function f_ω is the same as the distribution of f as a random variable on the Loeb product space.

The results. In general, averages of a mass phenomenon stabilize at some fixed value as the number of observations increases. The same limiting value is obtained if the averages are evaluated over a previously given subsequence of the observations. This property has been characterized with the aphorism "*No betting system can beat the house*"; it means that a gambler cannot change the expectation of his return by timing his betting. In the following definition, we formalize and generalize this intuitive observation to the hyperfinite setting. This notion of consistent satisfiability of the law of large numbers will prove to be crucial in our theory.

DEFINITION 1. Let f be a real-valued process on a Loeb product space $\big(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P)\big)$. We say that f satisfies the consistent law of large numbers (or simply the consistency law) if for any internal set $A \in \mathcal{T}$ with $L(\lambda)(A) > 0$, the process f^A on the reduced Loeb product space $\big(A \times \Omega, L(\mathcal{T}^A \otimes \mathcal{A}), L(\lambda^A \otimes P)\big)$ still satisfies the law of large numbers. This means that for $L(P)$-almost all $\omega \in \Omega$,

$$\int_A f_\omega^A \, dL(\lambda^A) = \iint_{A\times\Omega} f^A \, dL(\lambda^A \otimes P),$$

where f^A is the restriction of f to $A \times \Omega$, $\mathcal{T}^A$ is the collection of all internal subsets of A, and λ^A is the internal probability measure on $(A, \mathcal{T}^A)$ rescaled from λ.

To characterize processes satisfying the consistency law, we need first to recall the following easily established fact about nonstandard product spaces: Given the standard measure spaces $\big(T, L(\mathcal{T}), L(\lambda)\big)$ and $\big(\Omega, L(\mathcal{A}), L(P)\big)$, the standard product measure space $\big(T \times \Omega, L(\mathcal{T}) \otimes L(\mathcal{A}), L(\lambda) \otimes L(P)\big)$ is contained in the Loeb product space $\big(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P)\big)$. By taking T to be a hyperfinite set and Ω its internal power set, both endowed with the Loeb counting probability measures, one obtains a product space for which the inclusion is proper (see [7]). The theorem presented below can, in fact, be used to show much more: the stated inclusion is proper if and only if neither $L(\lambda)$ nor $L(P)$ is purely atomic.

Let $\mathcal{U}$ denote the product σ-algebra $L(\mathcal{T}) \otimes L(\mathcal{A})$. For an integrable real-valued process f on the Loeb product space, we let $E(f|_{\mathcal{U}})$ denote the conditional expectation of f with respect to $\mathcal{U}$. Note that such a construction which involves both a product σ-algebra and a significant extension of it (such as the $\mathcal{U}$ and the Loeb product algebra $L(\mathcal{T} \otimes \mathcal{A})$ here) is unique in the sense that it has no natural counterpart in standard mathematical practice or in nonstandard mathematics using only internal entities.

Our first theorem shows that f satisfies the consistency law if and only if $E(f|_{\mathcal{U}})$ is independent of particular samples. It can be seen from the theorem that a process satisfying the consistency law is already not measurable with respect to $\mathcal{U}$ except the trivial case that the random variables in the process are constants.

THEOREM 1. *Let f be an integrable real-valued process on a Loeb product space $\big(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P)\big)$. Then the following are equivalent*:

(1) *f satisfies the consistency law*;

(2) *for $L(\lambda \otimes P)$-almost all $(t, \omega) \in T \times \Omega$, $E(f|_{\mathcal{U}})(t, \omega) = h(t)$, where h is an integrable function on $\big(T, L(\mathcal{T}), L(\lambda)\big)$.*

The above characterization of processes satisfying the consistency law can be refined if the process is assumed to be square integrable. The equivalence of (1) and (2) below shows that almost sure uncorrelatedness of random variables f_t forming a square integrable process f is the weakest condition needed to assure that f satisfies the consistency law. A special case of this condition occurs when the random variables f_t are independent and have identical distributions. The equivalence of (2) and (3) below indicates a sort of duality between the random variables and the sample functions in a process. If square integrable processes f and g satisfy the property described below in (4), then we will say that those processes themselves are almost surely uncorrelated. Thus the equivalence of (2) and (4) says, then, that a square integrable process has almost surely uncorrelated random variables if and only if it is almost surely uncorrelated with any other square integrable processes.

THEOREM 2. *Let f be a square integrable real-valued process on a Loeb product space $\big(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P)\big)$. Then the following are equivalent*:

(1) *f satisfies the consistency law*;

(2) *the random variables f_t are almost surely uncorrelated, i.e., for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, f_{t_1} and f_{t_2} are uncorrelated*;

(3) *the mean function*

$$m(\omega) = \int_T f_\omega \, dL(\lambda)$$

and the covariance function

$$G(\omega_1, \omega_2) = \int_T \big(f_{\omega_1}(t) - m(\omega_1)\big)\big(f_{\omega_2}(t) - m(\omega_2)\big)\, dL(\lambda)(t)$$

of the sample functions f_ω are constant functions;

(4) *given another square integrable process g on $\big(T\times\Omega, L(\mathcal{T}\otimes\mathcal{A}), L(\lambda\otimes P)\big)$, we have for $L(\lambda\otimes\lambda)$-almost all $(t_1,t_2)\in T\times T$,*

$$\int_\Omega f_{t_1}\cdot g_{t_2}\,dL(P)=\int_\Omega f_{t_1}\,dL(P)\int_\Omega g_{t_2}\,dL(P).$$

We have seen that the conditional expectation $E(f|_{\mathcal{U}})$ of a process f can be used to test whether f satisfies the law of large numbers. It will be desirable to have a more concrete description of the process $E(f|_{\mathcal{U}})$ in terms of the original process f. Assume f is a square integrable real-valued process on the Loeb product space. Let

$$R(t_1,t_2)=\int_\Omega f(t_1,\omega)f(t_2,\omega)\,dL(P)$$

be the associated autocorrelation function. By serving as a kernel, the autocorrelation function defines an integral operator K on the space $L^2\big(L(\lambda)\big)$ of functions square integrable on $\big(T, L(\mathcal{T}), L(\lambda)\big)$. That is,

$$K(h)(t_1)=\int_T R(t_1,t_2)h(t_2)\,dL(\lambda)(t_2)$$

for $h\in L^2\big(L(\lambda)\big)$. One can also define the autocorrelation function S of sample functions f_ω by letting

$$S(\omega_1,\omega_2)=\int_T f(t,\omega_1)f(t,\omega_2)\,dL(\lambda).$$

Let L be the integral operator with S serving as the kernel function. Now K and L are compact, self-adjoint and semi-definite operators. Let γ_1, γ_2, ... be the nonincreasing sequence of all the positive eigenvalues of K with each eigenvalue being repeated up to its multiplicity. Let ψ_1, ψ_2, ... be the corresponding eigenfunctions adjusted to form an orthonormal family. This sequence of functions is called a complete eigensystem for K. The following biorthogonal theorem gives a representation for $E(f|_{\mathcal{U}})$ in terms of the eigensystems of K and L (see [9], p. 144, for such a biorthogonal expansion for processes which are continuous in quadratic means on an interval). Note that if there are only m positive eigenvalues, then we should replace the infinite sum by a finite sum of m terms.

THEOREM 3. *Let $\{\psi_n\}_{n=1}^{\infty}$ be a complete eigensystem for the integral operator K defined by the autocorrelation function R of a square integrable process f. Then*

$$E(f|_{\mathcal{U}})(t,\omega)=\sum_{n=1}^{\infty}\lambda_n\varphi_n(\omega)\psi_n(t),$$

where

$$\lambda_n = \gamma_n^{1/2}, \qquad \varphi_n(\omega) = \frac{1}{\lambda_n}\int_T f_\omega(t)\psi_n(t)\,dL(\lambda),$$

and γ_n is the corresponding eigenvalue for both ψ_n and φ_n. Moreover, the functions φ_n are orthonormal, forming a complete eigensystem for the integral operator L defined by the autocorrelation function S of the sample functions in the process f.

The following definition is an analog of Definition 1 for the case of the law of large numbers in terms of empirical distributions.

DEFINITION 2. Let f be a process from a Loeb product space $\big(T\times\Omega, L(\mathcal{T}\otimes\mathcal{A}), L(\lambda\otimes P)\big)$ to a metric space X. We say that f satisfies the consistent law of large numbers in distribution (or simply the consistency law in distribution) if for any internal set $A\in\mathcal{T}$ with $L(\lambda)(A)>0$, the process f^A from $\big(A\times\Omega, L(\mathcal{T}^A\otimes\mathcal{A}), L(\lambda^A\otimes P)\big)$ to X satisfies the law in distribution. In other words, for $L(P)$-almost all $\omega\in\Omega$, the distribution μ_ω^A induced by the random variable f_ω^A from $\big(A, L(\mathcal{T}^A), L(\lambda^A)\big)$ to X is equal to the distribution μ^A induced by the random variable f^A from $\big(A\times\Omega, L(\mathcal{T}^A\otimes\mathcal{A}), L(\lambda^A\otimes P)\big)$ to X. Here f^A is the restriction of f to $A\times\Omega$, $\mathcal{T}^A$ is the collection of all internal subsets of A, and λ^A is the internal probability measure on $(A,\mathcal{T}^A)$ rescaled from λ.

Our final result uses separating classes for collections of distributions. It is clear from the definition given below that a class is separating for a fixed collection of distributions if and only if it has a countable subclass which is separating for the collection of distributions.

DEFINITION 3. Let $\mathcal{D}$ be a collection of Borel probability measures on a separable metric space X and let $\mathcal{E}$ be a class of real or complex valued Borel functions on X such that for any given pair $(\phi,\mu)\in\mathcal{E}\times\mathcal{D}$, ϕ is μ-integrable. The class $\mathcal{E}$ is said to be separating for $\mathcal{D}$, if there is a sequence $\{\phi_n\}_{n=1}^\infty$ of functions in $\mathcal{E}$ which distinguishes the members of $\mathcal{D}$. That is, μ and ν in $\mathcal{D}$ are equal if

$$\int_X \phi_n\,d\mu = \int_X \phi_n\,d\nu$$

for all $n\geq 1$.

Even though independence has long been a primary focus of probability theory, it remains as a source of problems in current research when it is not assumed. The assumption of independence yields crucial multiplicative properties involving characteristic functions, generating functions, method of moments, etc. It is known that for a fixed finite collection of random variables, independence is strictly stronger than these properties. A natural and fundamental problem remains open: what is the precise relationship between independence and multiplicative properties when a large number of random variables is involved?

The following theorem says that almost sure pairwise independence is necessary and sufficient for the satisfiability of the consistency law in distribution. We also obtain the unexpected result that the almost sure versions of various multiplicative properties are all equivalent to almost sure pairwise independence. To see this, we note first that the complex exponentials e^{iux} form a separating class for all distributions on $\mathbb{R}$, the class of functions $\{ z^x : z \in (-1, 1) \}$ is separating for distributions of random variables taking values in natural numbers, and the class of functions $\{ x^k : k \in \mathbb{N} \}$ is separating for a collection of distributions satisfying some moment conditions. Applying Theorem 4 below to these separating classes leads to the equivalence of independence and the presence of multiplicative properties involving characteristic functions, generating functions and the method of moments.

THEOREM 4. *Let $\mathcal{D}$ be a collection of distributions on a separable metric space X, $\mathcal{E}$ a separating class for $\mathcal{D}$ of real or complex valued Borel functions, and f a process from a Loeb product space $\big(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P)\big)$ to X such that the distributions of all the sample functions f_ω as well as the distribution of the process f viewed as a random variable are all in $\mathcal{D}$. Assume that $\varphi(f)$ is a square integrable process for each $\varphi \in \mathcal{E}$. Then the following are equivalent*:

(1) *f satisfies the consistency law in distribution*;

(2) *the random variables f_t are almost surely pairwise independent, i.e., for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, f_{t_1} and f_{t_2} are independent*;

(3) *for each $\varphi \in \mathcal{E}$, the $\big(\varphi(f)\big)_t$ are almost surely uncorrelated, i.e., for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$,*

$$E\Big(\big(\varphi(f)\big)_{t_1}\overline{\big(\varphi(f)\big)_{t_2}}\Big) = E\big(\varphi(f)\big)_{t_1} E\overline{\big(\varphi(f)\big)_{t_2}}.$$

REMARK. It can be shown that if one is given any complete separable metric space and any two atomless Loeb spaces, then one can construct a process with almost surely pairwise independent random variables from the relevant Loeb product space to the metric space, and moreover, the random variables in the process can be required to have any variety of distributions. This property, which guarantees the abundance of processes with almost surely pairwise independent random variables, will be called the universality of atomless Loeb product spaces.

Concluding remarks. Note that standard models and the corresponding nonstandard models have the same collection of true sentences in a suitable formal language. The transfer of a standard result to the nonstandard model cannot be regarded as a new result. As indicated by Henson and Keisler in [4], however, nonstandard constructions allow one to use higher order sets more conveniently and efficiently, and this in turn helps one to discover genuinely new results when interpreting nonstandard results in the standard models. In our setting, we can see by reinterpreting parts of Theorems 2 and 4 in the large finite settings that asymptotic uncorrelatedness or asymptotic pairwise independence is necessary and sufficient for a triangular array (or a sequence) of random variables to satisfy appropriate versions of the law of large numbers. Moreover, the "backwards transfer" to standard models of the equivalence of (2) and (3) in Theorem 4 shows that asymptotic pairwise independence is equivalent to the asymptotic versions of various multiplicative properties.

Next note that the asymptotic properties of stochastic processes on large finite probability spaces are equivalent to certain properties of internal processes on hyperfinite probability spaces. These properties in turn are usually equivalent to some properties of processes which are measurable with respect to the relevant Loeb product spaces. Thus, to understand the asymptotic nature of processes on large finite probability spaces (which correspond to the large finite phenomena being modeled), one only needs to consider processes which are measurable with respect to Loeb product spaces. By applying some results formulated usually in the framework of continuous mathematics (for example, conditional expectations on infinite spaces), results about processes on Loeb product spaces can be obtained, which also correspond to some asymptotic results on large finite probability spaces (see [7] and [11] for some other special standard properties of Loeb spaces).

On the other hand, the classical theory of stochastic processes with a Lebesgue interval as the parameter space is usually only capable of dealing with processes which are measurable with respect to the relevant product measure. Such a process can be lifted to a process which is measurable with respect to the relevant product space of Loeb spaces (see, for example, [7]). This means that the classical theory is included in the theory of processes which are measurable with respect to the product space of Loeb spaces, i.e., with respect to the $\mathcal{U}$. It is thus no wonder that the results presented here are beyond the scope of classical analysis since they involve processes which are measurable with respect to the Loeb product spaces but usually have nontrivial singular parts with respect to the $\mathcal{U}$. Moreover, as argued in the above paragraph, there is essentially no loss of scientific information, if we model a probabilistic phenomenon by a hyperfinite process on a Loeb product space. This means that there is virtually no need in scientific modeling to consider processes beyond those which are Loeb product measurable. What we have presented here, therefore, provides a missing part in the general theory of processes. Indeed, there is no hope of developing a similar theory for processes with a Lebesgue interval parameter set.

In many branches of mathematics and its applications, there are various ways to model large but finite phenomena. One can, for example, often create models using differential equations. The many properties of differential equations which correspond to properties of difference equations lead one to numerical solution by computers. It is clear, however, that differential equations provide a better mathematical framework than difference equations for the qualitative study of relevant problems. When working with a large finite phenomenon for which the rate of convergence is not essential, continuous mathematics usually provides conceptually simpler and analytically more tractable models, making further study possible and efficient. For models involving topological structures, one can usually obtain asymptotic results from the continuum case with the help of compactness. For measure theoretic structures, however, there is often no direct way to relate the usual continuum models to the asymptotic cases. The results reported here strongly suggest that for the purpose of scientific modeling, hyperfinite Loeb spaces can be used systematically to replace the traditional measure spaces.

Full details and proofs of the theorems together with many other results are contained in [13] (see also [12] and [14] for the study of the law of large numbers in the setting of set-valued or vector valued processes).

REFERENCES

[1] R. M. Anderson, *Non-standard analysis with applications to economics*, ***Handbook of mathematical economics*** (W. Hildebrand and H. Sonnenschein, editors), vol. IV, North-Holland, New York, 1991.

[2] T. F. Bewley, *Stationary monetary equilibrium with a continuum of independently fluctuating consumers*, ***Contributions to mathematical economics: in honor of Gerard Debreu*** (W. Hildebrand and A. Mas-Colell, editors), North-Holland, New York, 1986.

[3] J. L. Doob, ***Stochastic processes***, Wiley, New York, 1953.

[4] C. W. Henson and H. J. Keisler, *On the strength of nonstandard analysis*, ***The Journal of Symbolic Logic***, vol. 51 (1986), pp. 377–386.

[5] A. E. Hurd and P. A. Loeb, ***An introduction to nonstandard real analysis***, Academic Press, Orlando, Florida, 1985.

[6] K. L. Judd, *The law of large numbers with a continuum of iid random variables*, ***Journal of Economic Theory***, vol. 35 (1985), pp. 19–25.

[7] H. J. Keisler, *Infinitesimals in probability theory*, ***Nonstandard analysis and its applications*** (N. J. Cutland, editor), Cambridge University Press, Cambridge, 1988.

[8] P. A. Loeb, *Conversion from nonstandard to standard measure spaces and applications in probability theory*, ***Transactions of the American Mathematical Society***, vol. 211 (1975), pp. 113–122.

[9] M. Loéve, ***Probability theory II***, 4th ed., Springer-Verlag, New York, 1977.

[10] R. E. Lucas and E. C. Prescott, *Equilibrium search and unemployment*, ***Journal of Economic Theory***, vol. 7 (1974), pp. 188–209.

[11] Y. N. Sun, *Distributional properties of correspondences on Loeb spaces*, to appear in

Journal of Functional Analysis.
[12] ———, *The law of large numbers for set-valued processes and stationary equilibria*, National University of Singapore, preprint.
[13] ———, *A theory of hyperfinite processes*, National University of Singapore, preprint.
[14] ———, *Vector valued processes and the law of large numbers*, to appear.

DEPARTMENT OF MATHEMATICS
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 0511

E-mail: matsuny@math.nus.sg