ELSEVIER

Journal of Mathematical Economics 29 (1998) 419–503

JOURNAL OF Mathematical ECONOMICS

A theory of hyperfinite processes: the complete removal of individual uncertainty via exact LLN [1]

Yeneng Sun [a,b]

[a] *Department of Mathematics, National University of Singapore, Singapore 119260*
[b] *Cowles Foundation, Yale University, New Haven, CT 06520, USA*

Received 1 October 1996

Abstract

The aim of this paper is to provide a viable measure–theoretic framework for the study of random phenomena involving a large number of economic entities. The work is based on the fact that processes which are measurable with respect to hyperfinite Loeb product spaces capture the limiting behaviors of triangular arrays of random variables and thus constitute the 'right' class for general stochastic modeling. The primary concern of the paper is to characterize those hyperfinite processes satisfying the *exact* law of large numbers by using the basic notions of conditional expectation, orthogonality, uncorrelatedness and independence together with some unifying multiplicative properties of random variables. The general structure of the processes is also analyzed via a biorthogonal expansion of the Karhunen–Loéve type and via the representation in terms of the simpler hyperfinite Loeb counting spaces. A universality property for atomless Loeb product spaces is formulated to show the abundance of processes satisfying the law. Generalizations to a hyperfinite number of continuous (or discrete) parameter stochastic processes are considered. The various necessary and sufficient conditions for the validity of the law provide a rather complete understanding about the cancelation of individual risks or uncertainty in

[1] Some of the results were presented at the 23rd Conference on Stochastic Processes and Their Applications of the Bernoulli Society for Mathematical Statistics and Probability in June 1995. The paper was revised while the author was visiting the Cowles Foundation for Research in Economics, Yale University in 1996–1997.

PII S0304-4068(97)00036-0

general settings. Some explicit asymptotic interpretations are also given.

JEL classification: C1; C6; D8; E0; G0

Keywords: Law of large numbers; Individual risks; Uncorrelatedness and independence; Karhunen–Loéve biorthogonal expansion; Multiplicative properties; Hyperfinite factor model; Loeb spaces

1. Introduction

The aim of this paper is to provide a viable measure–theoretic framework for the study of the ideal limit of a large number of random variables. Since large economic systems with uncertain environments are often modeled by a large number of random variables, our results can be used along with other standard measure–theoretic methods to study various mass economic phenomena in a rigorous and effective way. The main concern of our work is the characterization of macroscopic stability as formalized by the law of large numbers via certain microscopic conditions involving the basic notions of conditional expectation, orthogonality, uncorrelatedness and independence as well as some unifying multiplicative properties of random variables.

When one models an economic situation involving a large number of random entities, it is natural to consider a continuum of random variables (simply called a process) since the continuum is commonly used to idealize a large number of objects. However, it will be seen below that if a large number of random variables has low intercorrelation, one may have difficulties using a continuum of random variables in the usual sense to model the ideal limit.

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. [2] This is often referred to as 'aggregation removes individual uncertainty', i.e., risks or uncertainty are allowed for the individual entities but there is no uncertainty from macroscopic point of view. The required exact equality between the sample mean (or distribution) and the theoretical mean (or distribution) allows the formulation of conceptually simpler and analytically more tractable models.

Now the question is: is there really a law of large numbers for processes with a continuum of random variables? There are two serious difficulties with the validity of the law in this case. The first is the well known measurability problem. As

[2] See, for example, Bertola (1994), Bewley (1986), Diamond and Dybvig (1983), Lucas (1980), Lucas and Prescott (1974), and Noe and Rebello (1994).

already pointed out by Doob 6 decades ago, the sample functions of such a process are usually too irregular to be useful (see Doob, 1937, 1953, p. 67 and p. 148, respectively). For example, it was shown that if the random variables of a continuous parameter process are independent and have a common distribution (not concentrated at a single point), then the process is not measurable and even has no measurable standard modification with respect to the relevant product measure (see Doob, 1953, p. 67). In fact, as noted by Judd (1985), 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, one still has to face the second difficulty. That is, the set of sample realizations satisfying the property that the expectation or distribution of the sample function is the theoretical one has outer measure one and inner measure zero; this set is, therefore, not measurable. Of course, one can again extend the measure on the sample space so that the law is satisfied by this very special process. But it has to be noted that the measure can also be extended in such a way that almost no sample realizations have the required properties. Thus, even if one obtains the law as described, the honesty of the results based on such a law is necessarily under strong suspicion, since one can also claim the almost sure failure of the relevant results if the measure is extended differently.

From another point of view, the law of large numbers is a result on macroscopic stability based on microscopic conditions; that is, whenever the microscopic conditions are satisfied by a process, then almost all sample functions should stablize in some sense. Finding a particular process to satisfy both the microscopic conditions and macroscopic stability is simply not enough. Since the law of large numbers does hold approximately for the discrete case, it follows that the usual concept of a continuum of random variables is a poor model for the ideal limit of a triangular array or, more specifically, of a sequence of random variables with low intercorrelation. Thus, to model large random economic phenomena with individual risks or uncertainty in an ideal setting, one has to go beyond the usual continuum framework.

This paper presents our successful treatment of the law of large numbers using hyperfinite processes which are measurable with respect to Loeb product spaces and also reveals their many hidden characteristics. This class of processes can be justified to be the ‘right’ class for modeling random phenomena with a large number of economic entities on the following three grounds.

First, to study the asymptotic properties of a large finite collection of random economic entities, there is virtually no need to consider processes beyond the class of hyperfinite processes, since all the asymptotic properties are already embedded in the limit setting. Consider m individual economic entities moving simultaneously. Their actions are not deterministic, and hence the system can be modeled by m random variables. When m is *large*, one can use a sequence of finite collections of random variables in the form $\{x_i^n : 1 \le i \le n\}$, $n = 1,2,\ldots,$ (called a triangular array of random variables) to describe some asymptotic properties of the large economic system. By transfer, these asymptotic properties are equivalent to some properties of internal processes with a hyperfinite index set, which in turn are equivalent to some properties of processes that are measurable with respect to the Loeb space of the relevant internal product of a hyperfinite internal probability space with another internal probability space (the Loeb product space) by the standardization procedure (i.e., by rounding off some infinitesimals). Thus, those Loeb product measurable processes, which form the class of hyperfinite processes studied here, capture the nature of all the asymptotic properties of a large finite probabilistic system as described by a triangular array of random variables. In fact, by appealing to the ultrapower construction on natural numbers (see, for example, Hurd and Loeb, 1985), the hyperfinite processes in this context are simply equivalence classes of triangular arrays of random variables, and the external cardinality of their index sets is the same as the cardinality of the continuum, which means that a hyperfinite number of random variables is a continuum of random variables.

Second, a general hyperfinite process f is usually not reducible to a continuous parameter process measurable with respect to the relevant product measure. It can be done so if and only if a sort of essential continuity in term of the index variable t must be imposed on the random variables f_t (see, for example, Keisler, 1988, p. 118, and also Lemma A.5 in Appendix A). As noted in Sections 3.2 and 4.3 below, f usually has a non-trivial part with low intercorrelation, or in other words, the process admits randomness on the individual level and thus even not measurable with respect to the relevant product space of Loeb spaces (the essential continuity certainly cannot be satisfied). Thus, if a typical large random economic system is modeled by a continuous parameter process, then some economic phenomena associated with the large system are already lost by the choice of the continuum model. In fact, to model a probabilistic phenomenon involving a large number of random variables in situations where there is no natural topological structure on the set indexing these random variables, it is also not meaningful to use a Lebesgue interval instead of a hyperfinite set as the index set. On the other hand, when a time parameter (which allows a natural topological structure) is explicitly introduced into a model with many economic entities, one can use a hyperfinite number of stochastic processes to model the economic situation (see Sections 5.4 and 8).

Third, *exact* results on hyperfinite processes can be routinely translated to large finite results and vice versa (this is also relevant to the first point). This type of property, called *asymptotic implementability* in Khan and Sun (1997a), essentially comes from the basic requirement on a non-standard model being elementarily equivalent to the standard model, i.e., they have the same collection of true sentences in a formal language. The property guarantees that the pathological phenomena as discussed in the early part of this introduction will not arise in the hyperfinite setting. Such kind of translations were already used by Brown and Robinson (1972) to obtain approximate results on the core of large finite economies from some internal results. We also illustrate such a translation in Section 9 to relate some *exact* results on hyperfinite processes to approximate results on triangular arrays of random variables.

The paper contains various versions of the law of large numbers for processes over Loeb product spaces. [3] There are also some unexpected findings in our work. In particular, we show that almost sure uncorrelatedness and almost sure pairwise independence are not only sufficient, but in contrast to the previous literature, also necessary for the validity of appropriate versions of the law of large numbers. Such intrinsic characterizations are then used to unify many other conditions involving characteristic functions, generating functions, method of moments, maximum of random variables, etc. The reinterpretation of some of the results to the large finite setting shows 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. For a general real-valued square integrable hyperfinite process, we analyze its structure in Sections 4.2 and 4.3 below. The asymptotic interpretation of the structural result (called a hyperfinite factor model) roughly says that a process formed by a large finite number of random variables can be expressed as the sum of a relatively small number of product functions together with a residual term which has low intercorrelation. Moreover, linear combinations of random variables in the residual term in general satisfies the law of large numbers approximately. This is another instance of ‘aggregation removes individual uncertainty’, which also indicates the central importance of the particular topic of law of large numbers in a general theory of hyperfinite processes.

Most of our results can be understood in the aspect that these results are valid on a special kind of standard measure spaces (i.e., the Loeb spaces) but not valid on the traditional measure spaces. [4] In fact, there is no hope of developing an analog of our theory for processes with a Lebesgue interval parameter set (see Appendix A). Our results differ from those in the literature of non-standard theory of stochastic processes in the aspect that both the conditions we use and the results we derive involve standard notions of conditional expectation, orthogonality,

[3] Note that Theorems 3.17 and 4.2 as well as some parts of Theorems 4.6 and 7.6 in this paper are announced in Sun (1996a).

[4] Since Loeb spaces are standard measure spaces, one can understand almost all the results and many proofs by regarding Loeb spaces as some special measure spaces endowed with certain desired properties not shared by other measure spaces. Note that when Lebesgue measure on the real line is used, one can usually ignore the Dedekind set-theoretic construction of real numbers and the particular construction of Lebesgue measure. For applications of the theory developed here, the particular construction of hyperfinite sets as well as that of Loeb spaces can also be ignored.

uncorrelatedness and independence rather than the transferred version of these notions. [5] In fact, the study of non-standard processes has been focused on those internal processes or their standardizations, which are reducible to the usual continuous parameter processes (see, for example, Albeverio et al., 1986).

Once the conceptual foundations are properly set, the proofs of many of our results on Loeb product spaces are not particularly difficult. Except for a few special results whose proofs depend on non-standard arguments, most other proofs are based on the application of these special results through standard measure–theoretic methods. [6] On the other hand, by comparing the asymptotic results with their exact versions in the Loeb space setting, we observe that complex approximations in the large finite case are reduced to simple truth in the ideal setting. These two facts together show the power of the chosen analytical framework as well as the regularity of the objects being studied. Thus, additional power is injected into probabilistic and measure–theoretic methods via the special measure spaces–Loeb spaces.

We conclude this introduction by summarizing the economic relevance of the results in this paper. First, as already emphasized earlier, hyperfinite processes form a suitable class for modeling large random economic phenomena. Second, on a more specific level, the meaning of individual risks, as documented in Anderson (1991), Feldman and Gilles (1985) and Green (1994), is that these risks are borne by individual economic entities but not by ‘large’ coalitions; thus, our characterizations of those hyperfinite processes satisfying the relevant versions of the law of large numbers by the various microscopic conditions provide a rather complete understanding about the description and cancellation of individual risks or uncertainty in general situations. Third, the results in this paper as well as their generalizations in Sun (1997b) are already applied to particular economic models. The endogenous hyperfinite factor model as summarized in Corollary 4.8, which derives factors endogenously through the Karhunen–Loéve type biorthogonal expansion and allows the *complete* removal of individual risks through diversification, are used in Khan and Sun (1997b) to single out three types of risks in a large asset market and then to unify some basic ingredients in the capital-asset-pricing model of Sharpe (1964) and Lintner (1965), and the arbitrage pricing theory of Ross (1976). The law of large numbers for set-valued processes is developed in Sun (1997b) and then used to establish stochastic consistency in general equilib-

[5] These standard notions are very helpful for the discovery and the proof of those results which have no analog in classical probability theory as well as in nonstandard probability theory developed so far. Indeed, as noted above, if we interpret the results asymptoticly, we can obtain many new results about triangular arrays (or sequences) of random variables.

[6] Except that of Theorem 3.1, all the proofs in Sections 3, 4, 7 and 8 can be understood without knowledge about nonstandard analysis, provided that either Theorem 3.1 or a Fubini theorem for Loeb measures is assumed.

rium models with idiosyncratic shocks in endowments and preferences; [7] a counterpart in the setting of large games is also considered by Khan and Sun (1996b). Finally, we note that a main contribution of this paper as well as that of Sun (1996a,b, 1997a,b) is to obtain special probabilistic and measure–theoretic properties of Loeb spaces which have direct significance for problems in economic theory, game theory and finance, where traditional measure theory is not applicable.

2. The mathematical background

As usual in the probabilistic literature, in this paper, a measurable function of two variables will be referred to as a process. When processes with the classical parameter sets such as the set $\mathbb{Z}^+$ of positive integers or a Lebesgue interval are involved, we shall use the term 'stochastic processes' to describe them since, in this case, we are more interested in the finite dimensional distributions of the processes generated from the parameter sets rather than the means or distributions of the sample functions. When a continuum (or a hyperfinite number in our setting) of stochastic processes are concerned, we may also call the collection a hyperprocess. For a continuum of continuous parameter stochastic processes, the first continuum is an index set and the second continuum often refers to time.

We shall now give a brief discussion about Loeb spaces and refer the interested readers to Khan and Sun (1997a) for constructions of Loeb spaces based on sequences (for details, see Albeverio et al., 1986; Hurd and Loeb, 1985). Let T and Ω be internal sets, $\mathcal{T}$ and $\mathcal{A}$ be some internal algebras of internal subsets of T and Ω respectively. Let λ and P be finitely additive internal probability measures on $(T, \mathcal{T})$ and $(\Omega, \mathcal{A})$ respectively. Define a real-valued set function $^{\circ}P$ on $(\Omega, \mathcal{A})$ such that for each $A \in \mathcal{A}$, $^{\circ}P(A)$ is standard part $^{\circ}(P(A))$ of the finite non-standard real number $P(A)$. The $\aleph_1$-saturation property implies that a countably infinite union of disjoint non-empty sets in $\mathcal{A}$ is never itself in $\mathcal{A}$, and hence $^{\circ}P$ is trivially countably additive. Thus the Carathéodory extension theorem shows that $^{\circ}P$ can be extended to a standard probability measure $L(P)$ on the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$. Let $(\Omega, L(\mathcal{A}), L(P))$ be the completion of the space $(\Omega, \sigma(A), L(P))$. This completion is usually referred to as the Loeb space. We can also obtain the corresponding Loeb space $(T, L(\mathcal{T}), L(\lambda))$ in the same way.

[7] A sequential version was first formulated by Hildenbrand (1971). General equilibrium models with indivisible goods as studied by Mas-Collel (1977) and Khan and Yamazaki (1981) are also considered. In contrast to the approach based on the special connection between two Loeb spaces as used here and in Sun (1997b), the papers of Sun (1996b, 1997a) focus on the special properties of one Loeb space, and these properties are applied to non-cooperative game theory in Khan and Sun (1996a, 1997a).

Note that the internal product space $(T \times \Omega, T \otimes \mathscr{A}, \lambda \otimes P)$ is also an internal probability space. The corresponding Loeb space is denoted by $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$; it is referred to as the Loeb product space. In this paper, all processes on the Loeb product space are assumed to be measurable with respect to the Loeb product algebra. 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. When we interpret T as the parameter space and Ω as the sample space of the process f, the f_t are usually called the random variables of the process, 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 (or simply the law) if for $L(P)$-almost all $\omega \in \Omega$,

$$\int_T f_\omega(t) \mathrm{d} L(\lambda)(t) = \int\int_{T \times \Omega} f \mathrm{d} L(\lambda \otimes P),$$

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 (or simply the law 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.

As proposed in Section 1, it is appropriate to use hyperfinite sets to index processes. Since asymptotic versions of hyperfinite sets are sequences of finite sets, there is no probabilistic reason to use a general internal set T as a parameter space of processes. In fact, from the point of view of scientific modeling, it is even not necessary to consider sample spaces beyond hyperfinite sets, since processes with hyperfinite sample spaces already capture the asymptotic as well as continuous nature of the phenomena being modeled. The main reason to allow the sample space Ω to be a general internal set is to relate our results on Loeb spaces to the part of probability theory involving a triangular array of random variables on a *common* probability space (see Section 9). Since Ω is allowed to be non-hyperfinite, we also allow T to be so for symmetric treatment of the mathematical results as well as for the sake of completeness and generality.

Note that standard models and the corresponding non-standard models have the same collection of true sentences in a suitable formal language. This fact allows the formulation of the transfer principle, which precisely states that the truth values of the interpretation of a sentence in the standard and non-standard models are the same. Thus it is important to note that the transfer of an existing result in the standard model will not give us anything new, other than a reinterpretation of the same result. In the case of the law of large numbers, one can transfer a version of the standard weak law of large numbers in the form involving a second moment (see, for example, Chung (1974), pp. 228–229) to obtain immediately a law of large numbers for a hyperfinite sequence of $*$-*independent* random variables with a hyperfinite counting probability measure on the index set (see, for example, Anderson, 1991; Keisler, 1977; Nelson, 1987 and Stroyan and Bayod, 1986). In fact, one can even obtain the stronger central limit theorem for a hyperfinite sequence of $*$-independent random variables by interpreting some standard version of the central limit theorem in non-standard models (see Albeverio et al., 1986; Anderson, 1976; Nelson, 1987; and Stroyan and Bayod, 1986).

In spite of the fact that the law of large numbers obtained by transfer is mathematically the same as its classical counterpart, we note that *$*$-independence* is not a robust condition for probabilistic modeling, since any infinitesimal error will destroy the property. From another aspect, to relate to traditional measure–theoretic results, one needs to work with a general standard process on a Loeb product space. Then $*$-independence is not a condition on the standard process but a condition on some lifting of the process, which is unfortunately not unique. Even if it is satisfied by one lifting, it will not be satisfied by almost all the rest. Thus, even though one can obtain a process satisfying both the macroscopic stability and the independence condition through transfer and pushing-down, this is not really a law of large numbers for processes on a Loeb product space as argued in the introduction in the context of the example in Judd (1985). Moreover, the existence of an internal sequence of $*$-independent random variables on a hyperfinite set Ω with a hyperfinite index set T imposes restriction on the comparative size of T and Ω. For example, let f be an internal process from $(T \times \Omega, \mathcal{T} \otimes \mathcal{A}, \lambda \otimes P)$ to $\{-1,1\}$ such that the f_t are identically distributed with equal distribution on -1 and 1. Let $|T|$ and $|\Omega|$ be the internal cardinalities of T and Ω respectively. If the f_t are $*$-independent, then, for any given internal sequence $\{i_t\}_{t \in T}$ of -1 and 1, $P(\{\omega\colon \forall t \in T,\ f_t(\omega) = i_t\} = 1/2^{|T|}$. That set is non-empty, and hence Ω is the union of $2^{|T|}$ many disjoint non-empty sets. Therefore, $2^{|T|} \le |\Omega|$. Even if we only assume the random variables f_t to be pairwise independent, similar problem still exists. In this case, the f_t, $t \in T$ are not equal to each other. Since there are only $2^{|\Omega|}$ many different internal functions from Ω to $\{-1,1\}$, we know that $|T| \le 2^{|\Omega|}$. Thus, results with the assumption of $*$-independence or even pairwise independence rule out many possible underlying spaces. It will be seen in Theorem 6.2 below that the almost sure pairwise independence condition proposed in this paper has a sort of strong universality. Moreover, almost sure uncorrelatedness and almost sure pairwise independence will prove to be intrinsic pertaining to the study of the law of large numbers (see Theorems 3.17, 4.6 and 7.6).

Now we recall the following easily established fact about non-standard product spaces: Given the standard measure spaces $(T, L(\mathcal{T}), \mathrm{L}(\lambda))$ and $(\Omega, L(\mathcal{A}), L(P))$, as first noted in Anderson (1976), the standard product measure space $(T \times \Omega, L(\mathcal{T}) \otimes L(\mathcal{A}), L(\lambda) \otimes L(P))$ is contained in the Loeb product space $(T \times \Omega, \mathrm{L}(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$ (see also Cutland, 1983; Stroyan and Bayod, 1986). 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 $\mathcal{U}$.

The difference between the two different types of non-standard product spaces needs special attention in the main body of this paper. Those non-standard product spaces which are obtained by taking the internal product first and the Loeb operation later have been called the Loeb product spaces, while spaces which are obtained by taking the Loeb operation first and the standard product later will be called the product Loeb spaces. On the other hand, the difference between $(T \times \Omega, L(\mathscr{T}) \otimes L(\mathscr{A}), L(\lambda) \otimes L(P))$ and its completion will not be elaborated in this paper. Both will be referred to as product Loeb spaces and denoted by $\mathscr{U}$.

The rest of the paper is organized as follows. Section 3 presents various versions of the law of large numbers for real-valued processes on Loeb product spaces. Real-valued square integrable processes on Loeb product spaces are studied in detail in Section 4. Section 5 concerns with random variables and processes taking values in a separable metric space. In Section 6, it is shown that one can construct a process with almost surely pairwise independent random variables from a general atomless Loeb product space to a Polish space with given variety of distributions, and thus our theory is always not vacuous. Section 7 characterizes the consistent law of large numbers in distribution via various multiplicative properties of random variables. In Section 8, some results in Sections 5–7 are generalized to the setting of a hyperfinite number of continuous parameter stochastic processes (a hyperprocess). In Section 9, several results on Loeb product spaces are translated to asymptotic results about triangular arrays of random variables. Section 10 contains some concluding remarks and finally some negative results on processes indexed by the unit Lebesgue interval are presented in Appendix A.

3. The law of large numbers for real-valued processes

3.1. Introduction

We focus on real-valued processes in Sections 3.2, 3.3, 3.4 and 3.5 below. Since our primary concern in this paper is the study of functions of several variables, it is important to know the probability measure with which the functions are integrated. Thus, for a real or complex valued random variable g on a probability space $(\Lambda, \mathscr{B}, \mu)$, we usually use the integral $\int_\Lambda g \, \mathrm{d}\mu$ rather than Eg to denote the expectation of g. When there is no confusion, Eg will also be used.

We establish in Section 3.2 the relevance of the law of large numbers with the conditional expectation with respect to $\mathscr{U}$. In particular, for a real-valued integrable process f on a Loeb product space, if the conditional expectation $E(f|_{\mathscr{U}})$ of f with respect to the associated product Loeb algebra $\mathscr{U}$ is essentially a function of the index variable t only, then f satisfies the law. That result is, in a certain sense, the best possible for the class of integrable processes (see Theorem 3.17 in Section 3.5). One can regard Theorem 3.1 as a primitive version of the law of large numbers which can be used to prove the Fubini theorem for Loeb measures and all other versions of the law of large numbers in this paper by using only simple measure–theoretic constructions. Section 3.3 presents the law for processes with almost surely orthogonal or uncorrelated random variables. Section 3.4 gives several versions of the law in distribution. In Section 3.5, we study in detail an important concept, the consistent law of large numbers, which formalizes and generalizes an intuitive probabilistic observation ‘‘No betting system can beat the house’’, i.e., a gambler cannot change the expectation of his return by timing his betting. [8] Note that Theorems 3.1, 3.17 and 3.21 allow us to use various properties of conditional expectations. This will prove to be very important in our study.

3.2. The conditional expectation $E(\cdot|_{\mathcal{U}})$ and the law

This section shows the importance of the conditional expectation operator with respect to $\mathcal{U}$ pertaining to the study of the law of large numbers. The main theorem of the section, Theorem 3.1, says that for a real-valued integrable process f on the Loeb product space $(T \times \Omega,\ L(\mathcal{T} \otimes \mathcal{A}),\ L(\lambda \otimes P))$, if $E(f|_{\mathcal{U}})$ is independent of the particular samples, then f satisfies the law. For any given real-valued integrable process e on the Loeb product space, let e be the residue process $f - E(f|_{\mathcal{U}})$ (or the singular part of f). It follows from a property of conditional expectation that

$$E(e|_{\mathcal{U}}) = E(f|_{\mathcal{U}}) - E(E(f|_{\mathcal{U}})|_{\mathcal{U}}) = E(f|_{\mathcal{U}}) - E(f|_{\mathcal{U}}) = 0.$$

Hence, e satisfies the law. On the other hand, as noted earlier in the introduction, one should start with a hyperfinite process on a Loeb product space for the purpose of general stochastic modeling. If the hyperfinite process considered is real-valued and integrable, then its residue process always satisfies the law. Since there is no reason to impose the additional assumption that the residue process of a general hyperfinite process is zero, one has to deal with processes satisfying the law but not measurable with respect to the $\mathcal{U}$. This also means that the law of large numbers plays an essential role in general stochastic modeling.

Note that the internal algebra $\mathcal{T} \otimes \mathcal{A}$ has a particularly simple structure. For any set $C \in \mathcal{T} \otimes \mathcal{A}$, C can be expressed as a hyperfinite union of disjoint internal rectangles. It can be shown that the Fubini property holds for $\lambda \otimes P$-integrable internal functions on $T \times \Omega$ (it is trivial when T or Ω is hyperfinite).

[8] The version of the consistency law defined here as well as other versions in Definitions 7.3 and 8.3 will prove to be crucial for the intrinsic characterizations of uncorrelatedness and independence in later sections.

Theorem 3.1. *Let f be a real-valued integrable process on the Loeb product space* $(T\times\Omega, L(\mathcal{T}\otimes\mathcal{A}), L(\lambda\otimes P))$. *If* $E(f|_{\mathcal{U}})=h$, *where h is an integrable function on* $(T, L(\mathcal{T}), L(\lambda))$, *then for* $L(P)$-*almost all* $\omega\in\Omega$, f_ω *is integrable, and* $\int_T f_\omega dL(\lambda) = \int_T h(t)\mathrm{d}L(\lambda)$. *Moreover, for any given* $A\in L(\mathcal{T})$, $\int_A f_\omega \mathrm{d}L(\lambda)=\int_A h(t)\mathrm{d}L(\lambda)$ *for* $L(P)$-*almost all* $\omega\in\Omega$.

Proof: It is easy to show that for a $L(\lambda\otimes P)$-null set D, the section $D_\omega=\{t:(t,\omega)\in D\}$ is $L(\lambda)$-null for $L(P)$-almost all $\omega\in\Omega$ (see Albeverio et al., 1986, p. 76). For the sake of completeness, we include the proof here. Choose a decreasing sequence $\{A^n\}_{n=0}^{\infty}$ of internal sets in $T\times\Omega$ such that $A^0=\bigcap_{n=1}^{\infty}A^n$, $D\subseteq A^0$ and $L(\lambda\otimes P)(A^0)=0$. Then for each $n\ge 1$,

$$L(\lambda\otimes P)(A^n)={}^\circ(\lambda\otimes P)(A^n)={}^\circ\!\int_\Omega \lambda(A^n_\omega)\mathrm{d}P=\int_\Omega {}^\circ\lambda(A^n_\omega)\mathrm{d}L(P)$$
$$=\int_\Omega L(\lambda)(A^n_\omega)\mathrm{d}L(P).$$

Hence, $\int_\Omega L(\lambda)(A^0_\omega)\mathrm{d}L(P)=0$, which implies that for $L(P)$-almost all $\omega\in\Omega$, A^0_ω is $L(\lambda)$-null. The result then follows from the fact that $D\subseteq A_0$.

Now we assume $h=0$. Let g be an S-integrable internal lifting of f on the Loeb product space and let $B_n=\{\omega:\int_T g(t,\omega)\mathrm{d}\lambda(t)>1/n\}$ for each $n\ge 1$. It is clear that $(\chi_{T\times B_n})\cdot g$ is still an S-integrable internal lifting of $(\chi_{T\times B_n})\cdot f$, where $\chi_{T\times B_n}$ is the indicator function of the set $T\times B_n$ in $T\times\Omega$. Since $E(f|_{\mathcal{U}})=0$, we have

$$\iint_{T\times\Omega}\chi_{T\times B_n} f\mathrm{d}L(\lambda\otimes P)=0,$$

and hence, $\iint_{T\times B_n} g\mathrm{d}(\lambda\otimes P)\simeq 0$ by a result on S-integrability (see, for example, Anderson, 1976, p. 20, or Albeverio et al., 1986, p. 72). On the other hand,

$$\iint_{T\times B_n} g\mathrm{d}(\lambda\otimes P)=\int_{B_n}\int_T g\mathrm{d}\lambda\mathrm{d}P>1/nP(B_n).$$

Therefore, $P(B_n)\simeq 0$. Similarly, the P-measure of the set $\{\omega:\int_T g(t,\omega)\mathrm{d}\lambda(t)< -1/n\}$ is also an infinitesimal. Hence, $\int_T g_\omega(t)\mathrm{d}\lambda\simeq 0$ for $L(P)$-almost all $\omega\in\Omega$. By the result shown in the previous paragraph, we know that for $L(P)$-almost all $\omega\in\Omega$, g_ω is an internal lifting of f_ω. We still need to show that g_w is S-integrable. The proof given here is adopted from (Albeverio et al., 1986, p. 76). It can be checked that

$$\lim_{n\to\infty}\int_\Omega {}^\circ\Big(\int_{|g_\omega|>n}|g_\omega|\mathrm{d}\lambda\Big)\mathrm{d}L(P)\le \lim_{n\to\infty}{}^\circ\!\int_\Omega\int_{|g_\omega|>n}|g_\omega|\mathrm{d}\lambda\mathrm{d}P$$
$$=\lim_{n\to\infty}{}^\circ\!\int_{|g|>n}|g|\mathrm{d}(\lambda\otimes P)=0.$$

Therefore, the monotone decreasing sequence of functions ${}^\circ\!\int_{|g_\omega|>n}|g_\omega|\mathrm{d}\lambda$ converge to 0 for $L(P)$-almost all $\omega\in\Omega$. By Proposition 3.2.10 (Albeverio et al., 1986, p. 72), we obtain the S-integrability of g_ω for $L(P)$-almost all $\omega\in\Omega$. Hence for $L(P)$-almost all $\omega\in\Omega$, f_ω is integrable and $\int_T f_\omega\mathrm{d}L(\lambda)=0$.

For the general case, let $e = f - h$. Then $E(e|_{\mathscr{U}}) = E(f|_{\mathscr{U}}) - E(h|_{\mathscr{U}}) = 0$. Hence, $\int_T e_\omega \mathrm{d}L(\lambda) = 0$ for $L(P)$-almost all $\omega \in \Omega$. Therefore, for $L(P)$-almost all $\omega \in \Omega$, $f_\omega = h + e_\omega$ is integrable and $\int_T f_\omega \mathrm{d}L(\lambda) = \int_T h(t)\mathrm{d}L(\lambda)$.

Finally, for any given $A \in L(\mathscr{T})$, it is clear that $E(\chi_A f|_{\mathscr{U}}) = \chi_A E(f|_{\mathscr{U}}) = \chi_A h$, where χ_A is the indicator function of A in T. Thus, the result shown above implies that for $L(P)$-almost all $\omega \in \Omega$, $\int_T \chi_A f_\omega \mathrm{d}L(\lambda) = \int_T \chi_A h \mathrm{d}L(\lambda)$, and hence $\int_A f_\omega \mathrm{d}L(\lambda) = \int_A h(t)\mathrm{d}L(\lambda)$. □

To appreciate the power of Theorem 3.1, we shall show in the next proposition that Theorem 3.1 together with the usual Fubini theorem for product measure spaces implies the Fubini theorem for Loeb measures. The latter Fubini theorem is also known as Keisler's Fubini theorem for Loeb measures. It was first proven for bounded functions on hyperfinite Loeb spaces in Keisler (1977) and generalized in Albeverio et al. (1986), Cutland (1983), Loeb (1985) and Stroyan and Bayod (1986).

Proposition 3.2. *We assume the following version of Theorem 3.1: if e is a real-valued integrable process on the Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ with $E(e|_{\mathscr{U}}) = 0$, then for $L(P)$-almost all $\omega \in \Omega$, e_ω is integrable and $\int_T e_\omega \mathrm{d}L(\lambda) = 0$. Then for any real-valued integrable process f on $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$, we have*

$$\int\int_{T\times\Omega} f \mathrm{d}L(\lambda \otimes P) = \int_\Omega \int_T f \mathrm{d}L(\lambda)\mathrm{d}L(P) = \int_T \int_\Omega f \mathrm{d}L(P)\mathrm{d}L(\lambda).$$

Proof: Let e be the residue process $f - E(f|_{\mathscr{U}})$ of f. Then $E(e|_{\mathscr{U}}) = 0$. By the assumed version of Theorem 3.1, we obtain $\int_T e_\omega \mathrm{d}L(\lambda) = 0$ for $L(P)$-almost all $\omega \in \Omega$, and hence $\int_\Omega \int_T e \mathrm{d}L(\lambda)\mathrm{d}L(P) = 0$. Since $E(f|_{\mathscr{U}})$ is integrable with respect to the completion of the relevant product Loeb space. The usual Fubini theorem for product measures implies that

$$\int\int_{T\times\Omega} E(f|_{\mathscr{U}})\mathrm{d}L(\lambda \otimes P) = \int_\Omega \int_T E(f|_{\mathscr{U}})\mathrm{d}L(\lambda)\mathrm{d}L(P).$$

Hence, we also have

$$\begin{aligned}
\int\int_{T\otimes\Omega} f \mathrm{d}L(\lambda \otimes P) &= \int\int_{T\times\Omega} E(f|_{\mathscr{U}})\mathrm{d}L(\lambda \otimes P) \\
&= \int_\Omega \int_T E(f|_{\mathscr{U}})\mathrm{d}L(\lambda)\mathrm{d}L(P) \\
&= \int_\Omega \int_T f \mathrm{d}L(\lambda)\mathrm{d}L(P) - \int_\Omega \int_T e \mathrm{d}L(\lambda)\mathrm{d}L(P) \\
&= \int_\Omega \int_T f \mathrm{d}L(\lambda)\mathrm{d}L(P).
\end{aligned}$$

By symmetry, the other equality follows. □

Note that integrability conditions on f_ω and on $\int_T f_\omega \mathrm{d}L(\lambda)$, which are implicitly used in the statement and proof of Proposition 3.2, follow from the standard

Fubini theorem and the corresponding integrability condition on e_ω. Since the main arguments in the proof of the Fubini theorem for Loeb measures (see Albeverio et al., 1986) are used in the proof of Theorem 3.1, it is not surprising that we can give a shorter proof of Theorem 3.1 based on that theorem. Thus, Theorem 2.1 is essentially equivalent to the Fubini theorem for Loeb measures in the sense that we can prove one by assuming the other.

Remark 3.3. *A proof of Theorem 3.1 by using the Fubini theorem for Loeb measures: Fix any $A \in L(\mathscr{T})$. Since $E(f|_{\mathscr{U}}) = h$, it follows from the definition of the conditional expectation that*

$$\int\int_{A\times B} f \mathrm{d} L(\lambda \otimes P) = \int\int_{A\times B} E(f|_{\mathscr{U}}) \mathrm{d} L(\lambda \otimes P) = \int\int_{A\times B} h \mathrm{d} L(\lambda \otimes P)$$

for any measurable set B in $L(\mathscr{A})$. By the Fubini theorem for Loeb measures

$$\int_B \int_A f_\omega(t) \mathrm{d} L(\lambda) \mathrm{d} L(P) = \int_B \int_A h(t) \mathrm{d} L(\lambda) \mathrm{d} L(P).$$

Thus, the measure ν on $(\Omega, L(\mathscr{A}))$ defined by $\nu(B) = \int_B \int_A (f_\omega^{(t)} - h(t)) \mathrm{d} L(\lambda) \mathrm{d} L(P)$ is the zero measure. By the uniqueness of the Radon–Nikodym derivative (see, for example, Royden, 1968, p. 238), we obtain for $L(P)$-almost all $\omega \in \Omega$, $\int_A (f_\omega - h) \mathrm{d} L(\lambda) = 0$, and hence, $\int_A f_\omega \mathrm{d} L(\lambda) = \int_A h \mathrm{d} L(\lambda)$. The rest is clear. □

Remark 3.4. *Let f be a real-valued integrable process such that for $L(\lambda \otimes P)$-almost all $(t, \omega) \in T \times \Omega$, $E(f|_{\mathscr{U}})(t, \omega) = h(t)$. Then we must have $h(t) = \int_\Omega f_t \mathrm{d} L(P)$ for $L(\lambda)$-almost all $t \in \Omega$. This claim is easy to prove. Since for any $A \in \mathscr{T}$, $\int_A f_\omega \mathrm{d} L(\lambda) = \int_A h \mathrm{d} L(\lambda)$ for $L(P)$-almost all $\omega \in \Omega$. By integrating the functions on both sides of the identity with respect to $L(P)$ on Ω, the Fubini theorem for Loeb measures implies that $\int_A \int_\Omega f_t \mathrm{d} L(P) \mathrm{d} L(\lambda) = \int_A h(t) \mathrm{d} L(\lambda)$. Hence, by the arbitrary choice of A and the uniqueness of the Radon–Nikodym derivative, it follows that $h(t) = \int_\Omega f_t \mathrm{d} L(P)$ for $L(\lambda)$-almost all $t \in T$.*

The next proposition presents a sort of uniform law for processes whose conditional expectations with respect to the relevant product Loeb spaces are independent of particular samples. The essential idea in the proposition will also be used in the appendix to formulate Lemma A.1. Note that a σ-algebra on a space is said to be countably generated if it is generated by some countable family in the σ-algebra (Cohn, 1980, p. 110).

Proposition 3.5. *Let f be a real-valued integrable process on the Loeb product space $(T \times \Omega, L(T \otimes A), L(\lambda \otimes P))$. Assume that $E(f|_{\mathscr{U}}) = h$ and h is an integrable function on T. Let $\mathscr{B}$ be a countably generated sub-σ-algebra of $L(\mathscr{T})$. Then for $L(P)$-almost all $\omega \in \Omega$, the following equality*

$$\int_T \phi(t) f(t,\omega) \mathrm{d} L(\lambda)(t) = \int_T \phi(t) h(t) \mathrm{d} L(\lambda)(t)$$

holds for all bounded $\mathscr{B}$-measurable function ϕ. In particular, if α is a Loeb measurable function from T to a separable metric space X, then for L(P)-almost all $\omega \in \Omega$, the identity

$$\int_T \psi(\alpha(t)) f(t,\omega) \mathrm{d}L(\lambda)(t) = \int_T \psi(\alpha(t)) h(t) \mathrm{d}L(\lambda)(t)$$

holds for all bounded Borel function ψ on X.

Proof: Since $\mathscr{B}$ is countably generated, we can choose countably many sets A_n, $n = 1,2,\dots$, such that the vector space $\boldsymbol{V}$ spanned by the indicator functions of all the A_n over the rational field is dense in the space of integrable functions over $(T,\mathscr{B},L(\lambda))$ (Cohn, 1980, p. 110). For $\phi \in \boldsymbol{V}$, since $E(f|_{\mathscr{U}}) = h$, we have $E(\phi f|_{\mathscr{U}}) = \phi h$. Hence it follows from Theorem 3.1 that

$$\int_T \phi(t) f(t,\omega) \mathrm{d}L(\lambda \otimes P) = \int_T \phi(t) h(t) \mathrm{d}L(\lambda)$$

for $L(P)$-almost all $\omega \in \Omega$. Since $\boldsymbol{V}$ is still countable, we can choose a subset B of Ω with full measure such that for each $\omega \in B$, $\int_T \phi(t) f(t, \omega) \mathrm{d}L(\lambda) = \int_T \phi h \mathrm{d}L(\lambda)$ holds for all $\phi \in \boldsymbol{V}$. Without loss of generality, we assume $A_1 = T$, and thus the constant function 1 is in $\boldsymbol{V}$. Therefore, for each $\omega \in B$, f_ω is integrable over $(T,L(\mathscr{T}),L(\lambda))$.

Now for any bounded $\mathscr{B}$-measurable function ϕ with an essential bound M, we can choose a sequence $\{\phi_n\}_{n=1}^{\infty}$ from $\boldsymbol{V}$ such that the ϕ_n are bounded by M and converge to ϕ almost surely. Thus for each fixed $\omega \in B$, the $\phi_n f_\omega$ are integrable over $(T,L(\mathscr{T}),L(\lambda))$ and converge to ϕf_w almost surely. Since the $\phi_n f_w$ and the $\phi_n h$ are dominated by the integrable functions $M|f_\omega|$ and $M|h|$ respectively, we obtain

$$\int_T \phi f_\omega \mathrm{d}L(\lambda) = \lim_{n\to\infty} \int_T \phi_n f_\omega \mathrm{d}L(\lambda) = \lim_{n\to\infty} \int_T \phi_n h \mathrm{d}L(\lambda) = \int_T \phi h \mathrm{d}L(\lambda)$$

Hence, $\int_T \phi f_\omega \mathrm{d}L(\lambda) = \int_T \phi h \mathrm{d}L(\lambda)$ for all bounded $\mathscr{B}$-measurable function ϕ. For a Loeb measurable function α from T to a separable metric space X, let $(O_n)_{n=1}^{\infty}$ be a countable open base of X and $\mathscr{B}$ the σ-algebra generated by all the $\alpha^{-1}(O_n)$. Then for all bounded Borel function ψ on X, $\psi \circ \alpha$ is $\mathscr{B}$-measurable. The rest is clear. □

3.3. The law for square integrable processes

For simplicity, hereafter the Fubini theorem shall refer to the usual Fubini theorem for product measures as well as the Fubini theorem for Loeb product measures. Theorems 3.7 and 3.8 below present versions of the law for processes with almost surely orthogonal or uncorrelated random variables. To prove the theorems, we need the following technical lemma on the integrability of some relevant functions. The proof is based on the Tonelli theorem for Loeb measures in Hurd and Loeb (1985) (p. 204). Similar integrability problems will arise quite often in the later part of this paper. They can usually be solved by using the same idea and the proofs will be omitted.

Lemma 3.6. *Let f be a real-valued process on $(T\times\Omega, L(\mathcal{T}\otimes\mathcal{A}), L(\lambda\otimes P))$. Assume that f is $L(\lambda\otimes P)$-square integrable. That is, $\int\int_{T\times\Omega} f^2(t,\omega)\mathrm{d}L(\lambda\otimes P)<\infty$. Then,*

(1) the function G on $T\times T\times\Omega$ defined by $G(t_1,t_2,\omega)=f_{t_1}(\omega)f_{t_2}(\omega)$ is $L(\lambda\otimes\lambda\otimes P)$-integrable;

(2) for $L(\lambda\otimes\lambda)$-almost all $(t_1,t_2)\in T\times T$, $f_{t_1}(\omega)f_{t_2}(\omega)$ is integrable over $(\Omega, L(\mathcal{A}), L(P))$;

(3) the function Ef_ω is square integrable over $(\Omega, L(\mathcal{A}), L(P))$.

Proof: First note that the function $G(t_1,t_2,\omega)=f_{t_1}(\omega)f_{t_2}(\omega)$ on $T\times T\times\Omega$ is $L(\mathcal{T}\otimes\mathcal{T}\otimes\mathcal{A})$-measurable. The Tonelli theorem for Loeb measures implies that, for $L(\lambda\otimes\lambda)$-almost all $(t_1,t_2)\in T\times T, G_{(t_1,t_2)}$ is $L(\mathcal{A})$-measurable, and also

$$\begin{aligned}
&\int\int\int_{T\times T\times\Omega}|G(t_1,t_2,\omega)|\mathrm{d}L(\lambda\otimes\lambda\otimes P)\\
&=\int\int_{T\times T}\int_\Omega|f_{t_1}(\omega)f_{t_2}(\omega)|\mathrm{d}L(P)\mathrm{d}L(\lambda\otimes\lambda)\\
&=\int_\Omega\int\int_{T\times T}|f(t_1,\omega)f(t_2,\omega)|\mathrm{d}L(\lambda\otimes\lambda)\mathrm{d}L(P)\\
&=\int_\Omega\int_T|f(t_1,\omega)|\mathrm{d}L(\lambda)(t_1)\int_T|f(t_2,\omega)|L(\lambda)(t_2)\mathrm{d}L(P)\\
&=\int_\Omega\left(\int_T|f(t,\omega)|\mathrm{d}L(\lambda)\right)^2\mathrm{d}L(P)\\
&=\int_\Omega\left(E|f_\omega|\right)^2\mathrm{d}L(P)\le\int_\Omega\int_T|f(t,\omega)|^2\mathrm{d}L(\lambda)\mathrm{d}L(P)\\
&=\int\int_{T\times\Omega}f^2(f,\omega)\mathrm{d}L(\lambda\otimes P)<\infty.
\end{aligned}$$

Hence, G is $L(\lambda\otimes\lambda\otimes P)$-integrable, and for $L(\lambda\otimes\lambda)$-almost all $(t_1,t_2)\in T\times T, G_{(t_1,t_2)}$ is integrable. It is also clear that $E|f_\omega|$ is square integrable over $(\Omega, L(\mathcal{A}), L(P))$. Since $|Ef_\omega|\le E|f_\omega|$ for all $\omega\in\Omega$, it is obvious that Ef_ω is square integrable over $(\Omega, L(\mathcal{A}), L(P))$. The proof is thus complete. □

Let $H(\omega_1,\omega_2,t)=f_{\omega_1}(t)f_{\omega_2}(t)$. If f is square integrable over the Loeb product space, then by symmetry, the above lemma also shows that H is $L(P\otimes P\otimes\lambda)$-integrable; Ef_t is square integrable on $(T, L(\mathcal{T}), L(\lambda))$; and $f_{\omega_1}(t)f_{\omega_2}(t)$ is $L(\lambda)$-integrable for $L(P\otimes P)$-almost all $(\omega_1, \omega_2)\in\Omega\times\Omega$. We shall now present a law of large numbers for processes with almost surely orthogonal random variables.

Theorem 3.7. *Let f be a real-valued square integrable process on the Loeb product space $(T\times\Omega, L(\mathcal{T}\otimes\mathcal{A}), L(\lambda\otimes P))$. Assume that the random variables f_t are almost surely orthogonal, i.e., for $L(\lambda\otimes\lambda)$-almost all $(t_1,t_2)\in T\times T, Ef_{t_1}, f_{t_2}=0$. Then for $L(P)$-almost all $\omega\in\Omega$, $Ef_\omega=0$.*

Proof: By Lemma 3.6, the function G on $T \times T \times \Omega$ defined by $G(t_1,t_2,\omega) = f_{t_1}(\omega) f_{t_2}(\omega)$ is $L(\lambda \otimes \lambda \otimes P)$-integrable. Then the same calculations as in that lemma together with the Fubini theorem imply that

$$\int\int\int_{T\times T\times \Omega} G(t_1,t_2,\omega) \mathrm{d}L(\lambda \otimes \lambda \otimes P)$$
$$= \int_{\Omega}\int\int_{T\times T} f(t_1,\omega) f(t_2,\omega) \mathrm{d}L(\lambda \otimes \lambda) \mathrm{d}L(P)$$
$$= \int_{\Omega}\left(\int_T f(t,\omega) \mathrm{d}L(\lambda)\right)^2 \mathrm{d}L(P).$$

Hence, $E(Ef_\omega)^2 = EG$. On the other hand,

$$EG = \int\int\int_{T\times T\times \Omega} G(t_1,t_2,\omega) \mathrm{d}L(\lambda \otimes \lambda \otimes P)$$
$$= \int\int_{T\times T}\int_{\Omega} f_{t_1} f_{t_2} \mathrm{d}L(P) \mathrm{d}L(\lambda \otimes \lambda) = 0.$$

Therefore, $E(Ef_\omega)^2 = 0$, which implies that $Ef_\omega = 0$ for $L(P)$-almost all $\omega \in \Omega$. □

Note that in the above theorem, if A is an atom of $L(\lambda)$, then there is a random variable ϕ such that $f_t = \phi$ for $L(\lambda)$-almost all $t \in A$. By the assumption of almost sure orthogonality, ϕ must be orthogonal to itself, and hence $\phi \equiv 0$. Therefore, $f_t \equiv 0$ for almost all $t \in A$. In the next theorem, we establish a version of the law for real-valued processes with almost surely uncorrelated random variables. Note that two real-valued random variables ϕ and ψ are uncorrelated, if $E(\phi\psi) = E\phi E\psi$. Atoms are also allowed for $L(\lambda)$. In particular, if A is an atom for $L(\lambda)$, then f_t needs to be uncorrelated to itself for almost all $t \in A$, i.e., $Ef_t^2 = (Ef_t)^2$. The classical Cauchy–Schwarz inequality implies that for almost all $t \in A$, f_t must be a constant function.

Theorem 3.8. *Let f be square integrable over $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Assume that 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. Then for $L(P)$-almost all $\omega \in \Omega$, $Ef_\omega = Ef$, i.e., $\int_T f_\omega dL(\lambda) = \int\int_{T\times\Omega} f dL(\lambda \otimes P)$.*

Proof: Let $e(t,\omega) = f(t,\omega) - Ef_t$. Since f is square integrable, it is clear that e is also square integrable over the Loeb product space. By the fact that the random variables f_t are almost surely uncorrelated, we know that the e_t are almost surely orthogonal. Hence, Theorem 3.7 implies that $Ee_\omega = 0$ for $L(P)$-almost all $\omega \in \Omega$. Therefore the Fubini theorem implies that

$$Ef_\omega = E(e_\omega + Ef_t) = Ee_\omega + E(Ef_t) = Ee_\omega + Ef = Ef$$

for $L(P)$-almost all $\omega \in \Omega$. □

3.4. The law in distribution

In this section, we present some sufficient conditions for a real-valued process to satisfy the law in distribution. To do this, we introduce the following type of independence, which is called diagonal independence in the sense that the diagonal of the joint distribution function is required to be the product of the relevant marginal distribution functions.

Definition 3.9. Let ϕ_1, ϕ_2 be real-valued random variables on some probability space, F_1, F_2 their distribution functions, and F their joint distribution function. If for any $x \in \mathbb{R}$, $F(x, x) = F_1(x)F_2(x)$, then ϕ_1 and ϕ_2 are said to be diagonally independent.

It is obvious that independence implies diagonal independence. But the converse is not true (see Examples 7.1 and 7.2). We can also view the above definition in another perspective. Let ψ be the maximum of real-valued random variables ϕ_1 and ϕ_2, and let $F_{\max}$ be the distribution of ψ. Then $F_{\max}(x) = F(x, x)$ (Chung, 1974, p. 141). Hence, ϕ_1 and ϕ_2 are diagonally independent if and only if $F_{\max} = F_1 F_2$. We can, of course, introduce diagonal independence for more random variables. However, only the notion of pairwise diagonal independence is relevant to our study. The following simple theorem says that almost sure pairwise diagonal independence is sufficient for the satisfiability of the law in distribution.

Theorem 3.10. *Let f be a real-valued process on $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$. If the random variables f_t are almost surely pairwise diagonally 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 diagonally independent, then for $L(P)$-almost all $\omega \in \Omega$, the distribution function F_ω of the sample function f_ω on $(T, L(\mathcal{T}), L(\lambda))$ equals the distribution function F of f viewed as a random variable on the Loeb product space.*

Proof: Fix any $x \in \mathbb{R}$. Let g be the indicator function $\chi_{f^{-1}((-\infty, x])}$ of the set $f^{-1}((-\infty, x])$ in $T \times \Omega$. Then for any fixed $t \in T$, g_t is the indicator function $\chi_{f_t^{-1}((-\infty, x])}$ of the set $f_t^{-1}((-\infty, x])$ in Ω, and for any fixed $\omega \in \Omega$, g_ω is the indicator function $\chi_{f_\omega^{-1}((-\infty, x])}$ of the set $f_\omega^{-1}((-\infty, x])$ in T. By the assumption of diagonal independence, we know that for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, $f_{t_1}^{-1}((-\infty, x])$ and $f_{t_2}^{-1}((-\infty, x])$ are independent events in Ω. Thus, the random variables g_t are almost surely uncorrelated. By Theorem 3.8, we obtain that, for $L(P)$-almost all $\omega \in \Omega$,

$$\int_T g_\omega \mathrm{d}L(\lambda) = \int\int_{T \times \Omega} g \mathrm{d}L(\lambda \otimes P),$$

which implies that $L(\lambda)(f_\omega^{-1}((-\infty, x])) = L(\lambda \otimes P)(f^{-1}((-\infty, x]))$. By working with a countable dense set in $\mathbb{R}$, we can obtain that for $L(P)$-almost all $\omega \in \Omega$, the distribution functions F_ω and F are the same. □

It is well known that if two random variables ϕ and ψ are independent, then the characteristic function of the sum $\phi+\psi$ is the product of the respective characteristic functions. However, the following example in Loéve (1977a) (p. 275) shows that the converse of the previous statement is not true. The characteristic function of a real-valued random variable α will be denoted by Γ_α.

Example 3.11. Let ϕ be a real-valued random variable with the Cauchy distribution function $G(x)=\frac{1}{2}+\frac{1}{\pi}\tan^{-1}x$, and let $\psi=\phi$. Then the characteristic functions of the random variables ϕ, ψ are the same and equal to $e^{|\varkappa|}$. The characteristic function $\Gamma_{\phi+\psi}$ of the random variable $\phi+\psi$ is equal to $e^{2|\varkappa|}$. Let F be the joint distribution function of ϕ and ψ. Then $F(1,1)=G(1)=3/4$, which is certainly not equal to $G(1)\cdot G(1)=9/16$. Thus, ϕ and ψ are even not diagonally independent and, of course, not independent. In fact, we should not expect *any* sort of independence from them, since ϕ and ψ are the same non-constant random variable.

Now take a random variable ϕ on the Loeb space $(\Omega, L(\mathscr{A}), L(P))$ with the Cauchy distribution function as above. If we define a process f on the Loeb product space $(T\times\Omega, L(\mathscr{T}\otimes\mathscr{A}), L(\lambda\otimes P))$ by letting $f(t,\omega)=\phi(\omega)$ for all $t\in T$, then for all $t_1,t_2\in T, \Gamma_{f_{t_1}+f_{t_2}}=\Gamma_{f_{t_1}}\Gamma_{f_{t_2}}$. This process f certainly does not satisfy the law of large numbers. So it seems not plausible to obtain a general version of the law of large numbers by using the multiplicative property of characteristic functions. However, it is surprising that it can be done provided that the sum is replaced by difference. For convenience, we introduce the following definition.

Definition 3.12. Two real-valued random variables ϕ, ψ are said to be Γ-uncorrelated, if $\Gamma_{\phi-\psi}=\Gamma_\phi\cdot\Gamma_{-\psi}$, i.e., the characteristic function of the random variable $\phi-\psi$ is the product of the characteristic functions of the random variables ϕ and $-\psi$.

Note that if we take α and β to be ϕ and $-\phi$ respectively, where ϕ is the random variable in Example 3.11, then α and β are Γ-uncorrelated but are neither diagonally independent nor uncorrelated. On the other hand, as shown by the following theorem, almost sure Γ-uncorrelatedness is also sufficient to ensure the validity of the law in distribution.

Theorem 3.13. *Let f be a real-valued process on $(T\times\Omega, L(\mathscr{T}\otimes\mathscr{A}), L(\lambda\otimes P))$. If the random variables f_t are almost surely Γ-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 Γ-uncorrelated, then for $L(P)$-almost all $\omega\in\Omega$, the distribution function F_ω of the sample function f_ω on T equals the distribution function F of the random variable f on $T\times\Omega$.*

Proof: Note that the Fubini theorem for complex functions on Loeb product spaces is still valid, which can be used to obtain the following: for any $u \in \mathbb{R}$,

$$\begin{aligned}
&\int_{\Omega} |Ee^{iuf_{\omega}} - Ee^{iuf}|^2 \mathrm{d}L(P)(\omega) \\
&\quad = \int_{\Omega} \int_T \int_T \left(e^{iuf_{t_1}(\omega)} - Ee^{iuf_{t_1}} \right) \left(e^{iuf_{t_2}(\omega)} - Ee^{iuf_{t_2}} \right) \\
&\qquad \times \mathrm{d}L(\lambda)(t_1) L(\lambda)(t_2) \mathrm{d}L(P)(\omega) \\
&\quad = \int\int_{T \times T} \left(\Gamma_{f_{t_1} - f_{t_2}} - \Gamma_{f_{t_1}} \cdot \Gamma_{-f_{t_2}} \right) \mathrm{d}L(\lambda \otimes \lambda).
\end{aligned}$$

It follows from the assumption of almost sure Γ-uncorrelatedness that the last integral is zero. Hence, $Ee^{iuf_{\omega}} - Ee^{iuf} = 0$ for $L(P)$-almost all $\omega \in \Omega$. This means that for $L(P)$-almost all $\omega \in \Omega$, the characteristic function $\Gamma_{f_{\omega}}$ the sample function f_{ω} on T is equal to the characteristic function of the process f viewed as a random variable on $T \times \Omega$. Since distribution functions are uniquely determined by characteristic functions (see, for example, Breiman (1968), p. 170), the theorem follows. □

We observe that the process f in Theorems 3.10 and 3.13 is not required to be integrable. However, if f is also assumed to be integrable, then the satisfiability of the law in distribution by f implies that f satisfies the law. The following corollary is thus obvious. Here we note that there are two diagonally independent random variables which are neither uncorrelated nor Γ-uncorrelated (see Example 7.1). It is also pointed out earlier that Γ-uncorrelated random variables could be neither uncorrelated nor diagonally independent. It is straightforward to check that $\sin(\pi x)$ and $\cos(\pi x)$ on the Lebesgue interval $[-1/2, 1/2]$ are uncorrelated but neither Γ-uncorrelated nor diagonally independent. However, it will be seen in Section 7.4 that the almost sure versions of diagonal independence and Γ-uncorrelatedness are, in fact, stronger than almost sure uncorrelatedness.

Corollary 3.14. Let f be a real-valued integrable process on $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Assume that the random variables f_t are either almost surely pairwise diagonally independent, or almost surely Γ-uncorrelated. Then for $L(P)$-almost all $\omega \in \Omega$,

$$\int_T f_{\omega} \mathrm{d}L(\lambda) = \int\int_{T \times \Omega} f \mathrm{d}L(\lambda \otimes P).$$

3.5. The consistency law

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’’ (see, for example, Papoulis, 1965); it means that a gambler cannot change the expectation of his return by timing his betting. For instance, in the rolling game of a fair dice, if one considers the averages of every eleventh rolling, the percentage of occurrences of a particular face approaches $1/6$, and the same average should also be obtained if every seventh rolling is counted. It is also documented by Feldman and Gilles (1985) that if the special case of a continuum of iid random variables is used to model individual risks, then one should require sample averages of any non-negligible subcollection of the random variables to be constant, which is shown by them to be inconsistent, however.

Next, we formalize and generalize the above intuitive observation to the hyperfinite setting. The following notion of consistent satisfiability of the law of large numbers will prove to be crucial in our theory. The rest of the section explores the implications of this concept.

Definition 3.15. Let f be a real-valued process on the Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. We say that f satisfies the consistent law of large numbers (or simply the consistency law) if for any internal set $A \in \mathscr{T}$ with $L(\lambda)(A) > 0$, the process f^A on the reduced Loeb product space $(A \times \Omega, L(\mathscr{T}^A \otimes \mathscr{A}), L(\lambda^A \otimes P))$ still satisfies the law. This means that for $L(P)$-almost all $\omega \in \Omega$,

$$\int_A f_\omega^A \mathrm{d} L(\lambda^A) = \int\int_{A \times \Omega} f^A \mathrm{d} L(\lambda^A \otimes P),$$

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

The above definition simply says that a process satisfies the consistency law if a set of random variables are taken away from the process and the resulting process still satisfies the law. Though we considered the reduced Loeb product spaces for all internal sub-index sets with positive Loeb measure, the following lemma shows that we can actually focus on those large index sets only. It means that we can only consider those A with $L(\lambda)(A) > 1 - \varepsilon$ for some prior given real number $\varepsilon \in (0, 1)$ (no matter how small ε is). By this fact, we can reinterpret the idea in Definition 3.15 as ‘‘a slight change in the size of the index set does not affect the satisfiability of the law’’. It will also be seen in the proof of Proposition 9.2 that this relaxation is convenient. One can also find the asymptotic version of Definition 3.15 in Proposition 9.2 (2), which seems quite reasonable from the statistical point of view.

Lemma 3.16. *Let f be a real-valued integrable process on $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. If there is a real number $\varepsilon \in (0, 1)$ such that for any internal set $A \in \mathscr{T}$ with $L(\lambda)(A) > 1 - \varepsilon$, the process f^A on the reduced Loeb product space $(A \times \Omega,$*

$L(\mathcal{T}^A \otimes \mathcal{A})$, $L(\lambda^A \otimes P))$ still satisfies the law, then f satisfies the consistency law. Here f^A, $\mathcal{T}^A$ and λ^A are the same as those in Definition 3.15.

Proof: For a given internal set $A \in \mathcal{T}$ with $L(\lambda)(A) > 0$, take a partition $A_1, \dots, A_n$ of A such that each $L(\lambda)(A_i)$ is less than ε. Fix i. Since $L(\lambda)(T - A_i)$ and $L(\lambda)(T)$ are greater than $1 - \varepsilon$, we can obtain that for $L(P)$-almost all $\omega \in \Omega$,

$$\int_{T - A_i} f_\omega \mathrm{d} L(\lambda) = \int\int_{(T - A_i) \times \Omega} f \mathrm{d} L(\lambda \otimes P)$$

and

$$\int_T f_\omega \mathrm{d} L(\lambda) = \int\int_{T \times \Omega} f \mathrm{d} L(\lambda \otimes P).$$

These two identities imply that for $L(P)$-almost all $\omega \in \Omega$,

$$\int_{A_i} f_\omega \mathrm{d} L(\lambda) = \int\int_{A_i \times \Omega} f \mathrm{d} L(\lambda \otimes P).$$

Hence, for $L(P)$-almost all $\omega \in \Omega$,

$$\int_A f_\omega \mathrm{d} L(\lambda) = \sum_{i=1}^{n} \int_{A_i} d_\omega \mathrm{d} L(\lambda) = \sum_{i=1}^{n} \int\int_{A_i \times \Omega} f \mathrm{d} L(\lambda \otimes P)$$
$$= \int\int_{A \times \Omega} f \mathrm{d} L(\lambda \otimes P).$$

It means that for $L(P)$-almost all $\omega \in \Omega$,

$$\int_A f_\omega^A \mathrm{d} L(\lambda^A) = \int\int_{A \times \Omega} f^A \mathrm{d} L(\lambda^A \otimes P),$$

and we are done. □

It is obvious that f satisfies the consistency law if and only if for any $A \in L(\mathcal{T})$,

$$\int_A f_\omega \mathrm{d} L(\lambda) = \int\int_{A \times \Omega} f \mathrm{d} L(\lambda \otimes P),$$

for $L(P)$-almost all $\omega \in \Omega$.

Next, note that for a process f, the assumption that the conditional expectation $E(f|_{\mathcal{U}})$ is equal to a function h on T seems to be a strong condition for f to satisfy the law, since some properties of conditional expectations also imply that any real-valued measurable function g on $(T \times \Omega,\ L(\mathcal{T}) \otimes L(\mathcal{A}),\ L(\lambda) \otimes L(P))$ such that $f \cdot g$ is integrable over the Loeb product space, the equality

$$\int_T f(t,\omega) g(t,\omega) \mathrm{d} L(\lambda) = \int_T h(t) g(t,\omega) \mathrm{d} L(\lambda)$$

holds for $L(P)$-almost all $\omega \in \Omega$. It will be seen below that this condition is also the weakest in a certain sense. In particular, we show in the following theorem that f satisfies the consistency law if and only if $E(f|_{\mathcal{U}})$ is essentially a function on the index set T.

Theorem 3.17. *Let f be a real-valued integrable process on the Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Then the following are equivalent:*

(1) f satisfies the consistency law;

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

Proof: For $(1) \Rightarrow (2)$, pick any measurable sets $A \in L(\mathscr{T})$ and $B \in L(\mathscr{A})$. Define h on T by letting $h(t) = \int_\Omega f_t \mathrm{d}L(P)$. By (1), we have for $L(P)$-almost all $\omega \in \Omega$,

$$\int_A f_\omega \mathrm{d}L(\lambda) = \int_A \int_\Omega f \mathrm{d}L(P) \mathrm{d}L(\lambda) = \int_A h(t) \mathrm{d}L(\lambda).$$

Whence

$$\int_B \int_A f_\omega \mathrm{d}L(\lambda) \mathrm{d}L(P) = \int_B \int_A h(t) \mathrm{d}L(\lambda) \mathrm{d}L(P).$$

Thus,

$$\int\int_{A \times B} f \mathrm{d}L(\lambda \otimes P) = \int\int_{A \times B} E(f|_{\mathscr{U}}) \mathrm{d}L(\lambda \otimes P) = \int\int_{A \times B} h \mathrm{d}L(\lambda \otimes P).$$

This means that the singed measures defined on $(T \times \Omega, L(\mathscr{T}) \otimes L(\mathscr{A}))$ by integrating $E(f|_{\mathscr{U}})$ and h respectively on sets in $L(\mathscr{T}) \otimes L(\mathscr{A})$ agree on the rectangles. Since $L(\mathscr{T}) \otimes L(\mathscr{A})$ is generated by all the rectangles $A \times B$ and the collection of rectangles is also closed under finite intersections. It means that the collection is a π-system. By applying Dynkin's $\pi - \lambda$ theorem (see Cohn, 1980, p. 44 and Durrett, 1991, p. 404), we obtain the fact that the two singed measures are equal to each other on $(T \times \Omega, L(\mathscr{T}) \otimes L(\mathscr{A}))$. Thus both $E(f|_{\mathscr{U}})$ and h are Radon–Nikodym derivatives of the same measure. By the uniqueness of the Radon–Nikodym derivatives, we have $E(f|_{\mathscr{U}}) = h$.
$(2) \Rightarrow (1)$ follows from Theorem 3.1. □

From the above proof, it is easy to conclude that for any given $A \in T$, $\int_A f_\omega \mathrm{d}L(\lambda) = 0$ for $L(P)$-almost all $\omega \in \Omega$ if and only if for any given $B \in \mathscr{A}$, $\int_B f_t \mathrm{d}L(P) = 0$ for $L(\lambda)$-almost all $t \in T$, which is equivalent to $E(f|_{\mathscr{U}}) = 0$. The following example shows that there are processes which satisfy the law but not the consistency law.

Example 3.18. Let $(T, L(\mathscr{T}), L(\lambda))$ and $(\Omega, L(\mathscr{A}), L(P))$ be atomless Loeb spaces. Choose random variables ϕ and ψ from T and Ω to $\{-1,1\}$ respectively such that both ϕ and ψ are equally distributed on -1 and 1. Define a process f on $T \times \Omega$ such that $f(t, \omega) = \phi(t)\psi(\omega)$. Then it is obvious that $E(f|_{\mathscr{U}}) = f$ and

$$\int_T f_\omega \mathrm{d}L(\lambda) = \int_T \int_\Omega f \mathrm{d}L(P) \mathrm{d}L(\lambda) = 0.$$

On the other hand, $h(t) = \int_\Omega f_t \mathrm{d}L(P) = 0$ for any $t \in T$, and hence $E(f|_{\mathscr{U}}) \neq h$. By Theorem 3.17, we know that f satisfies the law but not the consistency law.

Note that if we have a reasonable sufficient condition for the validity of the law, then a process satisfying the condition should satisfy the consistency law. This claim is illustrated by the following simple observation.

Proposition 3.19. *Let $\mathscr{P}$ denote some symmetric binary relation between pairs of integrable functions on $(\Omega, L(\mathscr{A}), L(\mathscr{P}))$. Suppose we have the following version of the law of large numbers:*

> *for any real-valued integrable process f on an arbitrary Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ if for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, $\mathscr{P}(f_{t_1}, f_{t_2})$ holds, then for L(P)-almost all $\omega \in \Omega$, $\int_T f_\omega \mathrm{d}L(\lambda) = \int\int_{T\times\Omega} f dL(\lambda \otimes P)$, where $(T, \mathscr{T}, \lambda)$ is any internal probability space*

.

Then f satisfies the consistency law and $E(f|_{\mathscr{U}}) = h$, where h is the function on T defined by $h(t) = \int_\Omega f_t dL(P)$.

Proof: Pick any internal set $A \in \mathscr{T}$ with $L(\lambda)(A) > 0$. Let f^A, $\mathscr{T}^A$, and λ^A be the same as in Definition 3.15. Since $L(\lambda^A \otimes \lambda^A)$ is also a rescaling of $L(\lambda \otimes \lambda)$ on $A \times A$, it is certainly true that for $L(\lambda^A \otimes \lambda^A)$-almost all $(t_1, t_2) \in A \times A$, $\mathscr{P}(f^A_{t_1}, f^A_{t_2})$ holds, and hence by the assumed law,

$$\int_A f^A_\omega \mathrm{d}L(\lambda^A) = \int\int_{A\times\Omega} f^A \mathrm{d}L(\lambda^A \otimes P)$$

for $L(P)$-almost all $\omega \in \Omega$. Therefore f satisfies the consistency law. By Theorem 3.17, $E(f|_{\mathscr{U}}) = h$. □

Remark 3.20. *In the above proposition, if we assume that $\mathscr{P}(g, 0)$ holds for any g integrable over $(\Omega, L(\mathscr{A}), L(P))$, then it is even not necessary to introduce subprocesses. To illustrate, suppose $\mathscr{P}$ satisfies the additional condition and we have the following version of the law of large numbers:*

> *for any real-valued integrable process f on the Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ if for $L(\lambda \otimes \lambda)$-almost all $(t_1,t_2) \in T \times T$, $\mathscr{P}(f_{t_1}, f_{t_2})$ holds, then for L(P)-almost all $\omega \in \Omega$, $\int_T f_\omega \mathrm{d}L(\lambda) = \int\int_{T\times\Omega} f dL(\lambda \otimes P)$*

.

Then we can claim that f satisfies the consistency law. To prove this claim, pick any measurable set A in $L(\mathscr{T})$. By the assumption on the binary relation $\mathscr{P}$, it is obvious that for any given $t_1, t_2 \in T$, if $\mathscr{P}(f_{t_1}, f_{t_2})$ holds, then so does $\mathscr{P}(\chi_A(t_1) f_{t_1}, \chi_A(t_2) f_{t_2})$. Thus, for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, $\mathscr{P}(\chi_A(t_1) f_{t_1}, \chi_A(t_2) f_{t_2})$ holds. By the assumed law, we know that for $L(P)$-almost all $\omega \in \Omega$, $\int_T \chi_A f_\omega \mathrm{d} L(\lambda) = \int\int_{T\times\Omega} \chi_A f \mathrm{d} L(\lambda \otimes P)$, hence, $\int_A f_\omega \mathrm{d} L(\lambda) = \int\int_{A\times\Omega} f \mathrm{d} L(\lambda \otimes P)$. Therefore f satisfies the consistency law.

We only considered binary relations above. It is clear that we can also formulate results concerning the law of large numbers involving other kind of relations. The same proof can also be used to show the satisfiability of the consistency law.

Next we observe that the almost sure versions of pairwise diagonal independence, Γ-uncorrelatedness, and uncorrelatedness are all stronger than the assumption involving conditional expectations as in Theorem 3.1 (they are not necessarily strictly stronger).

Theorem 3.21. *Let f be a real-valued integrable process on a Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Define a function h on T by letting $h(t) = \int_\Omega f_t dL(P)$. If one of the following is satisfied,*

(1) the random variables f_t are almost surely pairwise diagonally independent;

(2) the random variables f_t are almost surely Γ-uncorrelated;

(3) f is square integrable and the random variables f_t are almost surely uncorrelated;

then $E(f|_{\mathscr{U}})$ is essentially a function on T. In addition, if the random variables f_t are almost surely orthogonal, then $E(f|_{\mathscr{U}}) = 0$.

Proof: If item (1) or (2), or (3) is satisfied, we can prove directly that f satisfies the consistency law by using Theorem 3.8 or Corollary 3.14 as in the proof of the results in Proposition 3.19 and Remark 3.20, and then appeal to Theorem 3.17 to finish the proof; we can also define binary relations on the space of integrable functions by using pairwise diagonal independence, or Γ-uncorrelatedness, or uncorrelatedness, and then use Proposition 3.19 or Remark 3.20 to finish the proof.

If the random variables f_t are almost surely orthogonal, then for any $A \in \mathscr{T}$, the $\chi_A(t) \cdot f_t$ are still almost surely orthogonal, and hence by Theorem 3.7, $\int_T \chi_A(t) f_\omega(t) \mathrm{d} L(\lambda) = 0$ for $L(P)$-almost all $\omega \in \Omega$. Thus, $\int\int_{A\times B} f \mathrm{d} L(\lambda \otimes P)$ for any $A \in \mathscr{T}$ and $B \in \mathscr{A}$. By applying Dynkin's $\pi - \lambda$ theorem as in the proof of Theorem 3.17, we obtain $E(f|_{\mathscr{U}}) = 0$. □

Remark 3.22. *For $1 \le p < \infty$, let $\mathscr{L}^p(L(\lambda \otimes P))$ the space real-valued processes on a Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ such that $|f|^p$ is $L(\lambda \otimes P)$-integrable. Let $\mathscr{L}^\infty(L(\lambda \otimes P))$ be the space essentially bounded real-valued processes on $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. For $1 \le p \le \infty$, we use $V_{\mathscr{U}}^p$ to denote the space of those processes in $\mathscr{L}^p(L(\lambda \otimes P))$ such that $E(f|_{\mathscr{U}}) = 0$. We also regard the space $\mathscr{L}^p(L(\lambda))$ as a subspace of $\mathscr{L}^p(L(\lambda \otimes P))$. Since the conditional expectation operator is a linear contractive mapping on $\mathscr{L}^p(L(\lambda \otimes P))$, Theorem 3.17 shows that the linear space of the processes in $\mathscr{L}^p(L(\lambda \otimes P))$ which* *satisfy the consistency law is the direct sum of* $V_{\mathscr{U}}^p$ *and* $\mathscr{L}^p(L(\lambda))$*. It will be shown in Proposition 6.7 that* $V_{\mathscr{U}}^p$ *is, in fact, a large space. In addition, we note that Conditions (1), (2) and (3) in Theorem 3.21 are all microscopic conditions which guarantee the validity of the consistency law. On the other hand, the existence of a process which satisfies both an independence condition and the consistency law is trivial. One can simply transfer an independent sequence of random variables to obtain a hyperfinite process by standardization. The validity of the consistency law for this hyperfinite process can be obtained by applying the classic law to subsequences. The work of Green (1994) contains the construction of a particular process satisfying the consistency law based on a different measure space. As noted earlier, such kind of results which involve the existence of particular processes satisfying both the microscopic independence condition and the macroscopic stability are not really law of large numbers.*

4. Conditional expectations for square integrable processes

4.1. Introduction

In Sections 4.2 and 4.3 we provide a more detailed study for square integrable processes. For a square integrable process f on a Loeb product space $(T \times \Omega,$ $L(\mathscr{T} \otimes \mathscr{A}),$ $L(\lambda \otimes P))$, Lemma 3.6 shows that the joint moment $\int_\Omega f_{t_1}(\omega) f_{t_2}(\omega) \mathrm{d} L(P)$ exists for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$. Define a function R on $T \times T$ by letting $R(t_1,t_2) = \int_\Omega f(t_1,\omega) f(t_2,\omega) \mathrm{d} L(P)$. This function is called the autocorrelation function of the process f (see Papoulis, 1965, p. 282). Since the conditional expectation $E(\cdot|_{\mathscr{U}})$ has been shown to be crucial in our study, it will be desirable to obtain a more concrete description of the process $E(\cdot|_{\mathscr{U}})$ in terms of the original process f. In Section 4.2, we provide a biorthogonal representation for $E(f|_{\mathscr{U}})$ in terms of the autocorrelation function of f. Section 4.3 contains various characterizations for processes with almost surely orthogonal or uncorrelated random variables. It is also noted that almost sure uncorrelatedness of random variables f_t is the weakest condition for f to satisfy the consistent law of large numbers. Finally, by combining some of these results together in Corollary 4.8, a structural result for general square integrable processes (called a hyperfinite factor model) is presented.

4.2. A biorthogonal representation of $E(f|_{\mathscr{U}})$

Let f be a square integrable process on $(T \times \Omega,$ $L(\mathscr{T} \otimes \mathscr{A}),$ $L(\lambda \otimes P))$ and R its autocorrelation function. The proof of Lemma 3.6 shows that R is $L(\lambda \otimes \lambda)$-integrable. In fact, we can show that R is $L(\lambda \otimes \lambda)$-square integrable by the same method. The following lemma says that R is essentially the same as the autocorrelation function of the process $E(f|_{\mathscr{U}})$.

Lemma 4.1. *The autocorrelation function R of the process f is $L(\lambda \otimes \lambda)$-almost surely equal to the $L(\mathscr{T}) \otimes L(\mathscr{T})$-measurable autocorrelation function of the process $E(f|_{\mathscr{U}})$.*

Proof: For notational simplicity, denote $E(f|_{\mathscr{U}})$ by g and $f-g$ by e. Then $E(e|_{\mathscr{U}})=0$. By Theorem 3.1, $E(e|_{\mathscr{U}})=0$ implies that $Ee_{\omega}=0$ for $L(P)$-almost all $\omega \in \Omega$. By symmetry, one can obtain $Ee_t=0$ for almost all $t \in T$. Since both f and g are square integrable, the Fubini theorem implies that there is a measurable subset C_1 of T with $L(\lambda)(C_1)=1$ such that for any $t \in C_1$, f_t, g_t and e_t are $L(P)$-square integrable. Thus, for any fixed $t' \in C_1$, $E(f_{t'} e|_{\mathscr{U}})=0$, which also implies that $E(f_{t'} e_t)=0$ for almost all $t \in T$. Hence, for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, $E(f_{t_1} e_{t_2})=0$. Similarly, for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times$ T, $E(e_{t_1} g_{t_2})=0$, and hence,

$$E(f_{t_1} f_{t_2}) = E(f_{t_1} g_{t_2}) + E(f_{t_1} e_{t_2}) = E(g_{t_1} g_{t_2}) + E(e_{t_1} g_{t_2}) = E(g_{t_1} g_{t_2}).$$

Therefore, the desired result follows. □

Now we move to a series expansion for $E(f|_{\mathscr{U}})$. In the classical theory of stochastic processes where parameter sets are intervals, a given process has various kinds of orthogonal decompositions (for example, Fourier expansion). Since $E(f|_{\mathscr{U}})$ is measurable with respect to the product measure space in the usual sense, it can be written as an infinite series in many ways (see, for example, Kadison and Ringrose, 1983); but a general series expansion will not provide any additional information on $E(f|_{\mathscr{U}})$ other than its square integrability with respect to the product measure. We shall focus on a special series expansion. A proper orthogonal decomposition theorem [9] is contained in Loéve (1977b) (see p. 144). The theorem provides a biorthogonal expansion for processes continuous in quadratic mean on an interval (or simply require the autocorrelation function to be continuous; see Papoulis, 1965, p. 457). Let f be a real-valued square integrable process with autocorrelation function R. By serving as a kernel, the autocorrelation function defines an integral operator K on the space $L^2(L(\lambda))$ of functions square integrable on $(T, L(\mathscr{T}))$. That is, $K(h)(t_1)=\int_T R(t_1, t_2) h(t_2) \mathrm{d} L(\lambda)(t_2)$ for $h \in L^2(L(\lambda))$. One can also define the autocorrelation function S of sample functions f_{ω} by letting $S(\omega_1, \omega_2)=\int_T f(t,\omega_1) f(t,\omega_2) \mathrm{d} L(\lambda)$. Let L be the integral operator with S serving as the kernel function. Now K and L are Hilbert–Schmidt operators and thus compact operators (see Conway, 1985, p. 43, or Kadison and Ringrose, 1983, p. 170). These operators are also self-adjoint and semi-definite. Let $\gamma_1, \gamma_2, \ldots$ be the non-increasing sequence of all the positive eigenvalues of K with each eigenvalue being repeated up to its multiplicity. Let

[9] It is also called the Karhunen–Loéve expansion theorem, which is simply the continuous version of the expansion in classical principal components analysis; see Basilevsky (1994), Chap. 3.

$\psi_1, \psi_2, \ldots$ 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 $E(f|_{\mathscr{U}})$ in terms of the eigensystems of K and L. Note that if there are only m positive eigenvalues, then we should replace the infinite sum by a finite sum of m terms.

Theorem 4.2. *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*

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

where $\lambda_n = \gamma_n^{1/2}$, $\varphi_n(\omega) = \frac{1}{\lambda_n} \int_T f_\omega(t) \psi_n(t) \mathrm{d}L(\lambda)$, 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.

Proof: First, we note that for a function ϕ in $\mathscr{L}^2(L(\lambda))$, if ϕ is orthogonal to every function in the range of the operator K, then ϕ must be in the kernel of the operator F defined by $F(h) = \int_T f(t,\omega) h(t) \mathrm{d}L(\lambda)$ for $h \in L^2(L(\lambda))$, i.e., $F(\phi)(\omega) = \int_T f(t,\omega)\phi(t)\mathrm{d}L(\lambda) = 0$. We simply observe that

$$\begin{aligned}
&\int_{t_2 \in T} \phi(t_2) K(\phi)(t_2) \mathrm{d}L(\lambda)(t_2) \\
&\quad = \int_{t_2 \in T} \phi(t_2) \int_{t_1 \in T} \phi(t_1) \int_\Omega f(t_1,\omega) f(t_2,\omega) \\
&\qquad \times \mathrm{d}L(P) \mathrm{d}L(\lambda)(t_1) \mathrm{d}L(\lambda)(t_2) \\
&\quad = \int_\Omega \int_{t_1 \in T} f(t_1,\omega) \phi(t_1) \mathrm{d}L(\lambda)(t_1) \int_{t_2 \in T} f(t_2,\omega) \phi(t_2) \\
&\qquad \times \mathrm{d}L(\lambda)(t_2) \mathrm{d}L(P) \\
&\quad = \int_\Omega \left(\int_T f(t,\omega) \phi(t) \mathrm{d}L(\lambda) \right)^2 \mathrm{d}L(P) = \int_\Omega (F(\phi))^2 \mathrm{d}L(P).
\end{aligned}$$

Hence, $F(\phi) = 0$.

Now fix bounded Loeb measurable functions α on T and β on Ω. Let ψ be the projection of α on the range space of K and let $\phi = \alpha - \psi$. Then ϕ is orthogonal to the range space of K, and hence $F(\phi) = 0$, $F(\alpha) = F(\psi)$. Since the functions ψ_n form a complete orthonormal basis for the range space of K, it is also clear that

$$\psi(\cdot) = \sum_{n=1}^{\infty} \int_T \alpha(t) \psi_n(t) \mathrm{d}L(\lambda) \psi_n(\cdot).$$

Next, we compute the following integrals.

$$\int\int_{T\times\Omega} \alpha(t)\beta(\omega)\sum_{n=1}^{\infty}\int_T f(t',\omega)\psi_n(t')\mathrm{d}L(\lambda)(t')\psi_n(t)\mathrm{d}L(\lambda\otimes P)(t,\omega)$$

$$= \int_\Omega \beta(\omega)\sum_{n=1}^{\infty}\int_T f(t',\omega)\psi_n(t')\mathrm{d}L(\lambda)(t')\int_T \alpha(t)\psi_n(t)$$

$$\times \mathrm{d}L(\lambda)(t)\mathrm{d}L(P)$$

$$= \int_\Omega \beta(\omega)\int_{t'\in T} f(t',\omega)\sum_{n=1}^{\infty}\int_{t\in T}\alpha(t)\psi_n(t)\mathrm{d}L(\lambda)(t)\psi_n(t')$$

$$\times \mathrm{d}L(\lambda)(t')\mathrm{d}L(P)$$

$$= \int_\Omega \beta(\omega)\int_{t'\in T} f(t',\omega)\psi(t')\mathrm{d}L(\lambda)(t')\mathrm{d}L(P)(\omega)$$

$$= \int_\Omega \beta(\omega)F(\psi)(\omega)\mathrm{d}L(P)$$

$$= \int_\Omega \beta(\omega)F(\alpha)(\omega)\mathrm{d}L(P)$$

$$= \int_\Omega \beta(\omega)\int_T f(t,\omega)\alpha(t)\mathrm{d}L(\lambda)\mathrm{d}L(P)$$

$$= \int\int_{T\times\Omega} \alpha(t)\beta(\omega)f(t,\omega)\mathrm{d}L(\lambda\otimes P).$$

Since $\sum_{n=1}^{\infty}\int_T(t',\omega)\psi_n(t')\mathrm{d}L(\lambda)(t')\psi_n(t)$ is $L(\mathscr{T})\otimes L(\mathscr{A})$-measurable, we thus have

$$E(f|_{\mathscr{U}})(t,\omega) = \sum_{n=1}^{\infty}\int_T f(t',\omega)\psi_n(t')\mathrm{d}L(\lambda)(t')\psi_n(t).$$

by the arbitrary choices of the α and β.

It is easy to check that

$$\int_\Omega\int_T f(t_1,\omega)\psi_m(t_1)\mathrm{d}L(\lambda)(t_1)\int_T f(t_2,\omega)\psi_n(t_2)\mathrm{d}L(\lambda)(t_2)\mathrm{d}L(P)(\omega)$$

$$= \gamma_n\delta_n^m,$$

where δ_n^m is equal to 0 if $m\neq n$ and 1 if $m=n$. Hence, the functions φ_n are orthonormal, and $E(f|_{\mathscr{U}})(t,\omega)=\sum_{n=1}^{\infty}\lambda_n\varphi_n(\omega)\psi_n(t)$. By symmetry, Lemma 4.1 also implies that S is $L(P\otimes P)$ essentially equal to the autocorrelation function of the sample functions of the process $E(f|_{\mathscr{U}})$. Thus, $S(\omega_1,\omega_2)=\sum_{n=1}^{\infty}\lambda_n^2\varphi_n(\omega_1)\varphi_n(\omega_2)$, which implies that the functions φ_n form a complete eigensystem for the Hilbert–Schmidt operator L. □

Note that the biorthogonal expansion of $E(f|_{\mathscr{U}})$ is almost unique in the sense that if $E(f|_{\mathscr{U}})=\sum_{n=1}^{\infty}\mu_n\alpha_n(\omega)\beta_n(t)$ with the conditions that both sequences of functions α_n, $n\geq 1$ and β_n, $n\geq 1$ are orthonormal, and also the numbers μ_n, $n\geq 1$ are non-negative and decreasing in terms of n, then $\mu_n=\lambda_n$, and α_n and β_n are the respective eigenfunctions of L and K corresponding to the eigenvalue

λ_n^2. In particular, if for some $m \geq 1$, λ_m^2 is an eigenvalue with multiplicity 1, then either $\alpha_m = \varphi_m$ and $\beta_m = \psi_m$, or $\alpha_m = -\varphi_m$ and $\beta_m = -\psi_m$. The following estimation on the eigenvalues is clear.

$$\sum_{n=1}^{\infty} \lambda_n^2 = \int\int_{T \times \Omega} \left(E(f|_{\mathscr{U}}) \right)^2 \mathrm{d} L(\lambda \otimes P) \leq \int\int_{T \times \Omega} f^2 \mathrm{d} L(\lambda \otimes P) < \infty.$$

In practice, one is usually able to compute the autocorrelation function R of f by statistical methods. Once R is computed, the eigenfunctions and eigenvalues of the compact self-adjoint semi-definite operator K can be computed (see, for example, Heuser, 1982, p. 281). Thus, one can compute $E(f|_{\mathscr{U}})$ if f is given. It is important to know that the functions ψ_n and φ_n in the biorthogonal representation of $E(f|_{\mathscr{U}})$ can be determined from f directly. Thus $E(f|_{\mathscr{U}})$ can be approximated before we know what it is.

Note that f can be written as $f = \sum_{n=1}^{\infty} \lambda_n \psi_n(t) \varphi_n(\omega) + e$, where e is the residue process $f - E(f|_{\mathscr{U}})$ of f. The following corollary shows that the joint moment of f_t or f_ω with a square integrable function on the relevant product Loeb space only involves the countably many functions in the principal part, i.e., the relevant joint moments involving the residual process are zero. In particular, the results hold when the function on the relevant product Loeb space depends only on one variable.

Corollary 4.3. *Let f be a real-valued square integrable process on a Loeb product space* $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ *and* $e = f - E(f|_{\mathscr{U}})$. *For a* $\mathscr{U}$*-measurable, square integrable function a on* $T \times \Omega$, *we have for* $L(P)$*-almost all* $\omega \in \Omega$, $\int_T a(t,\omega) e(t,\omega) \mathrm{d} L(\lambda) = 0$,

$$\int_T a(t,\omega) f(t,\omega) \mathrm{d} L(\lambda) = \sum_{n=1}^{\infty} \lambda_n \int_T a(t,\omega) \psi_n(t) \mathrm{d} L(\lambda) \varphi_n(\omega),$$

and for $L(\lambda)$*-almost all* $t \in T$, $\int_\Omega a(t,\omega) e(t,\omega) dL(P) = 0$,

$$\int_\Omega a(t,\omega) f(t,\omega) \mathrm{d} L(P) = \sum_{n=1}^{\infty} \lambda_n \int_\Omega a(t,\omega) \varphi_n(\omega) \mathrm{d} L(P) \psi_n(t),$$

where the λ_n, φ_n *and* ψ_n *are the same as in Theorem 4.2.*

Proof: Let $g = a \cdot e$. Then a property of conditional expectation implies that $E(g|\mathscr{U}) = a \cdot E(e|_{\mathscr{U}}) = 0$. It follows from Theorem 3.1 that for $L(P)$-almost all $\omega \in \Omega$, $\int_T g(t,\omega) \mathrm{d} L(\lambda) = 0$. By Theorem 4.2,

$$\begin{aligned} &\int_T a(t,\omega) f(t,\omega) \mathrm{d} L(\lambda) \\ &\quad = \int_T a(t,\omega) E(f|_{\mathscr{U}})(t,\omega) \mathrm{d} L(\lambda) \\ &\quad = \sum_{n=1}^{\infty} \lambda_n \int_T a(t,\omega) \psi_n(t) \mathrm{d} L(\lambda) \varphi_n(\omega) \end{aligned}$$

for $L(P)$-almost all $\omega \in \Omega$. By symmetry, we can obtain the second identity. □

The next proposition provides a characterization for those centered processes satisfying the law in terms of the eigenfunctions of the operator K. A real-valued integrable process f is said to be centered if $Ef_t = 0$ for almost all t.

Proposition 4.4. Let f be a square integrable centered process on a Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Then f satisfies the law if and only if the mean of any eigenfunction corresponding to a positive eigenvalue of the operator K is zero.

Proof: For the sufficiency part, we simply note that the condition implies that $E\psi_n = 0$ for all n, and hence it follows from Corollary 4.3 that f satisfies the law.

For the necessity part, let ψ be an eigenfunction corresponding to a positive eigenvalue γ of K. Then

$$\int_{t_2 \in T} R(t_1, t_2)\psi(t_2)\mathrm{d}L(\lambda)(t_2) = \gamma\psi(t_1),$$

which implies that

$$\begin{aligned}
&\gamma \cdot \int_T \psi(t)\mathrm{d}L(\lambda) \\
&\quad = \int_{t_1 \in T}\int_{t_2 \in T}\int_{\omega \in \Omega} f(t_1, \omega) f(t_2, \omega)\psi(t_2)\mathrm{d}L(P)(\omega) \\
&\qquad \times \mathrm{d}L(\lambda)(t_2)\mathrm{d}L(\lambda)(t_1) \\
&\quad = \int_{\omega \in \Omega}\int_{t_1 \in T} f(t_1, \omega)\mathrm{d}L(\lambda)(t_1)\int_{t_2 \in T} f(t_2, \omega)\psi(t_2) \\
&\qquad \times \mathrm{d}L(\lambda)(t_2)\mathrm{d}L(P).
\end{aligned}$$

Since the law is satisfied, $\int_{t_1 \in T} f(t_1, \omega)\mathrm{d}L(\lambda)(t_1) = 0$ for $L(P)$-almost all $\omega \in \Omega$, and hence, $\gamma \cdot \int_T \psi(t)\mathrm{d}L(\lambda) = 0$. Therefore, the mean $E\psi$ is the zero. □

4.3. Characterizations of almost sure orthogonality and uncorrelatedness

This section contains some characterizations for those square integrable processes f with almost surely orthogonal (or almost surely uncorrelated) random variables. Note that some of the characterizations are included not because of their importance, but for the sake of easy references in some proofs. We shall, first, consider almost sure orthogonality. By Theorem 3.21, if the random variables f_t are almost surely orthogonal, then $E(f|_{\mathscr{U}}) = 0$. It will be seen in Theorem 4.5 that the two conditions are actually equivalent.

The equivalence of (2) and (3) in Theorem 4.5 states that the almost sure orthogonality of the random variables in a process is equivalent that of the sample functions in the process. Such a duality shows the perfect symmetry between the parameter variable t and the sampling variable ω for the situation under consideration.

Note that there are essential differences between orthogonal processes and almost surely orthogonal processes. For example, let T be a hyperfinite set, $\Omega=\{-2,-1,1,2\}^T$, and λ and P be the internal counting probability measures on T and Ω respectively. Define an internal process f from $(T\times\Omega, \mathscr{T}\otimes\mathscr{A}, \lambda\otimes P)$ to $\{-1,1\}$ by letting $f(t,\omega)$ be the sign of ω_t. Then the random variables f_t are orthogonal to each other. The previous duality result also yields the almost sure orthogonality of the sample functions f_ω. On the other hand, there are only $2^{|T|}$ many different internal functions from T to $\{-1,1\}$; but there are $4^{|T|}$ many sample functions f_ω. This means that many of the sample functions f_ω are equal to each other. Therefore, the sample functions f_ω cannot be orthogonal to each other.

The equivalence of (2) and (6) also deserves some attention. Note that it is easy to find random variables that are orthogonal, but their expectations do not vanish. However, for a square integrable process f, if the random variables f_t are almost surely orthogonal, then the expectations of the random variables f_t are almost surely zero.

Theorem 4.5. *For a square integrable real-valued process f on $(T\times\Omega, L(\mathscr{T}\otimes\mathscr{A}), L(\lambda\otimes P))$, the following are equivalent:*

(1) $E(f|_{\mathscr{U}})=0$;

(2) The random variables f_t are almost surely orthogonal in $\mathscr{L}^2(L(P))$;

(3) The sample functions f_ω are almost surely orthogonal in $\mathscr{L}^2(L(\lambda))$;

(4) The compact, self-adjoint and semi-definite integral operator L on the Hilbert space $L^2(L(P))$ defined by the autocorrelation function S of the sample functions f_ω has no positive eigenvalue;

(5) The compact, self-adjoint and semi-definite integral operator K on the Hilbert space $L^2(L(\lambda))$ defined by the autocorrelation function R of the process f has no positive eigenvalue;

(6) The means of the random variables f_t are almost surely 0, and the eigenfunctions of L corresponding to any possible positive eigenvalue must be the constant functions;

(7) The means of the sample functions f_ω are almost surely 0, and the eigenfunctions of K corresponding to any possible positive eigenvalue must be constant functions.

Proof: First, consider (1) ⇒ (2). By Lemma 4.1, the autocorrelation function of f is $L(\lambda\otimes\lambda)$-almost surely equal to that of $E(f|_{\mathscr{U}})$, and hence, is zero. This means that (2) holds. (2) ⇒ (1) follows from Theorem 3.21.

By symmetry, (1) and (3) are equivalent, and thus (2) and (3) are also equivalent. It might be interesting to see a direct proof of the equivalence of (2) and (3). By the Fubini theorem,

$$\iint_{\Omega\times\Omega}\left(\int_T f_{\omega_1}(t) f_{\omega_2}(t)\mathrm{d}L(\lambda)\right)^2 \mathrm{d}L(P\otimes P)$$
$$=\iint_{\Omega\times\Omega}\int_{t_1\in T} f_{\omega_1}(t_1) f_{\omega_2}(t_1)\mathrm{d}L(\lambda)(t_1)\int_{t_2\in T} f_{\omega_1}(t_2) f_{\omega_2}(t_2)$$
$$\times \mathrm{d}L(\lambda)(t_2)\mathrm{d}L(P\otimes P)$$
$$=\iint_{T\times T}\int_{\omega_1\in\Omega} f_{t_1}(\omega_1) f_{t_2}(\omega_1)\mathrm{d}L(P)(\omega_1)\int_{\omega_2\in\Omega} f_{t_1}(\omega_2) f_{t_2}(\omega_2)$$
$$\times \mathrm{d}L(P)(\omega_2)\mathrm{d}L(\lambda\otimes\lambda)$$
$$=\iint_{T\times T}\left(\int_\Omega f_{t_1}(\omega) f_{t_2}(\omega)\mathrm{d}L(P)\right)^2 \mathrm{d}L(\lambda\otimes\lambda).$$

Hence, the required equivalence is clear.

(2) $\Rightarrow$ (5) and (3) $\Rightarrow$ (4) clearly follow from Lemma 4.1. (4) $\Rightarrow$ (1) and (5) $\Rightarrow$ (1) follow from Theorem 4.2. Thus, the first five statements are equivalent.

Now, we consider (1) $\Rightarrow$ (6). Since $E(f|_{\mathscr{U}})=0$, as noted in the proof of Lemma 4.1, Theorem 3.1 also implies that $Ef_t=0$ for almost all t by symmetry. Next, since (1) $\Rightarrow$ (4), L thus has no positive eigenvalue at all. So (6) follows.

Assume (6) holds. If there is a positive eigenvalue λ_1^2 for L, then we can take φ_1 to be the constant function 1 and all other φ_n to be zero in Theorem 4.2. Hence $E(f|_{\mathscr{U}})=\lambda_1\psi_1$, where ψ_1 is a normalized eigenfunction of K. By Theorem 3.1, we must have $\lambda_1\psi_1(t)=Ef_t$ for $L(\lambda)$-almost all $t\in T$. By the assumption that $Ef_t=0$ for $L(\lambda)$-almost all $t\in T$, we have $\lambda_1\psi_1=0$, which contradicts the conditions on λ_1 and ψ_1. Therefore, L has no positive eigenvalue at all. Thus (4) holds, and hence (1) holds.

By symmetry, we can also obtain the equivalence of (1) and (7). □

Next, we move to the characterizations of square integrable processes f with almost surely uncorrelated random variables. Let e be the centered process $(f-Ef_t)$ constructed from f. Then the random variables f_t are almost surely uncorrelated if and only if the random variables e_t are almost surely orthogonal. Thus, Theorem 4.5 also gives us many equivalent conditions in terms of e. In the following proposition, we present some additional equivalent conditions in terms of f directly. As in Theorem 4.5, the equivalence of (2) and (3) in the following proposition also 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 (7), then we shall say that those processes themselves are almost surely uncorrelated. Thus, the equivalence of (2) and (7) says 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 4.6. *Let f be a real-valued square integrable process on the Loeb product space $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$. Then the following are equivalent:*

(1) $E(f|_{\mathcal{U}})$ is essentially a function on T;

(2) The random variables f_t are almost surely uncorrelated;

(3) The autocorrelation function S of the sample functions f_ω is the constant function with value $E(Ef_t)^2$;

(4) Let L be the compact self-adjoint semi-definite operator on the Hilbert space $L^2(L(P))$ defined by using the autocorrelation function S of sample functions of f as the kernel function; then the eigenfunctions of L corresponding to any possible positive eigenvalue must be the constant functions;

(5) The covariance function and the mean function of the sample functions f_ω are constant functions;

(6) For any square integrable function ϕ on $(\Omega, L(\mathcal{A}), L(P))$, f_t and ϕ are uncorrelated for $L(\lambda)$-almost all $t \in T$;

(7) Given another square integrable process g on $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$, we have for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, $\int_\Omega f_{t_1} \cdot g_{t_2} \mathrm{d}L(P) = \int_\Omega f_{t_1} \mathrm{d}L(P) \int_\Omega g_{t_2} \mathrm{d}L(P)$;

(8) f satisfies the consistency law.

Proof: Let $e = f - Ef_t$. Then (1) is equivalent to $E(e|_{\mathcal{U}}) = 0$, and (2) is equivalent to the fact that the random variables e_t are almost surely orthogonal. Hence (1) is equivalent to (2).

(1) ⇒ (3) follows from Lemma 4.1 by symmetry.

Now, assume (3). Then $L(\phi) = \int_\Omega c\phi(\omega) dL(P)$, where c is the constant coming from the autocorrelation function of the sample functions f_ω. Thus the eigenfunctions of L corresponding to any possible positive eigenvalue must be the constant functions. Hence, (3) ⇒ (4).

Assume (4). By Theorem 4.2, $E(f|_{\mathcal{U}})$ is essentially a function on T, and hence (1) holds. Thus, (1), (2), (3) and (4) are equivalent.

Since $E((f_{\omega_1} - Ef_{\omega_1})(f_{\omega_2} - Ef\omega_2)) = E(f_{\omega_1} f_{\omega_2}) - Ef_{\omega_1} Ef_{\omega_2}$, we have (5) ⇒ (3).

Now, assume (1). Then by Theorem 3.1, Ef_ω defines a constant function on Ω. By (3), $E(f_{\omega_1} f_{\omega_2})$ defines a constant function on $\Omega \times \Omega$. Hence, the covariance function is constant, and thus (1) ⇒ (5).

To show (1) ⇒ (6), let $h(t) = Ef_t$; then $E(f|_{\mathcal{U}}) = h$. Since $E(f\phi|_{\mathcal{U}}) = h\phi$, we have for any $A \in L(\mathcal{T})$,

$$\int_A \int_\Omega f\phi \mathrm{d}L(P) \mathrm{d}L(\lambda) = \int_A \int_\Omega h\phi \mathrm{d}L(P) \mathrm{d}L(\lambda) = \int_A h \mathrm{d}L(\lambda) \int_\Omega \phi \mathrm{d}L(P).$$

Hence, for $L(\lambda)$-almost all $t \in T$, $\int_\Omega f_t \phi \mathrm{d}L(P) = h(t) \int_\Omega \phi \mathrm{d}L(P)$. Whence (6) holds.

If we assume (6), then for any $t_2 \in T$ with g_{t_2} being square integrable, we know that f_{t_1} and g_{t_2} are uncorrelated for $L(\lambda)$-almost all $t_1 \in T$. By the Fubini theorem, we obtain (7).

By taking $g = f$, (7) ⇒ (2) is clear.

The equivalence of (1) and (8) is already shown in Theorem 3.17. □

It follows from Theorems 4.5 and 4.6 that the random variables f_t are almost surely uncorrelated if and only if there is a function h on T and a process e whose sample functions e_ω are almost surely orthogonal such that f is the sum of e and h. It will be interesting to know when the sample functions f_ω are also almost surely uncorrelated. Note that if the random variables f_t are almost surely uncorrelated, then the covariance function of the sample functions is essentially the non-negative constant $\int_T (Ef_t)^2 \mathrm{d} L(\lambda) - (Ef)^2$. It will be seen in the following proposition that we also have a sort of duality between uncorrelatedness and constant means for the random variables and the sample functions in a real-valued square integrable process. The almost sure uncorrelatedness of the random variables implies that the sample functions have constant means almost surely; and if in addition, the random variables themselves have constant means almost surely, then the sample functions are almost surely uncorrelated. By symmetry, the random variables in the process are uncorrelated with constant means almost surely if and only if so are the sample functions in the process.

Proposition 4.7. *Let f be a square integrable process on $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Assume that the random variables f_t are almost surely uncorrelated. Then the sample functions f_ω are almost surely uncorrelated if and only if the random variables f_t have a common mean almost surely, i.e., the mean function Ef_t is essentially a constant function on T*

Proof: If the sample functions f_ω are almost surely uncorrelated, then, by symmetry, we obtain from Theorem 3.8 that $Ef_t = Ef$ for $L(\lambda)$-almost all $t \in T$.

On the other hand, if the mean function Ef_t is essentially a constant number c on T, then the centered process $e = f - c$ has almost surely orthogonal random variables. Theorem 3.7 implies that $Ee_\omega = Ef_\omega - c$ is zero for almost all $\omega \in \Omega$, and hence, Ef_ω is equal to the constant c almost surely. By Theorem 4.5, the sample functions e_ω are also almost surely orthogonal, which implies that the sample functions f_ω are almost surely uncorrelated. □

To conclude this section, we collect in the following corollary part of Theorems 4.2, 4.5, and Corollary 4.3 to have a more or less complete description about the structure of a general square integrable process.

Corollary 4.8. *Let f be a real-valued square integrable centered process on the Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Then f has the following expression*

$$f(t,\omega) = \sum_{n=1}^{\infty} \lambda_n \psi_n(t) \varphi_n(\omega) + e(t,\omega),$$

with properties:

(1) λ_n, $1 \le n < \infty$ is a decreasing sequence of positive numbers; the collection $\{\psi_n: 1 \le n < \infty\}$ is orthonormal; and $\{\varphi_n: 1 \le n < \infty\}$ is a collection of orthonormal and centered random variables.

(2) $E(f|_{\mathscr{U}})(t,\omega) = \sum_{n=1}^{\infty} \lambda_n \psi_n(t) \varphi_n(\omega)$ *and* $E(e|_{\mathscr{U}}) = 0$.

(3) The random variables e_t are almost surely orthogonal, which is to say that $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, $\int_\Omega e_{t_1}(\omega) e_{t_2}(\omega) \mathrm{d} L(P)(\omega) = 0$.

(4) If p is a square integrable real-valued function on $(T, L(\mathscr{T}), L(\lambda))$, then for $L(P)$-almost all $\omega \in \Omega$, $\int_T p(t) e_\omega(t) L(\lambda) = 0$ and $\int_T p(t) f(t,\omega) \mathrm{d} L(\lambda)(t) = \sum_{n=1}^{\infty} \lambda_n (\int_T p(t) \psi_n(t) \mathrm{d} L(\lambda)(t)) \varphi_n(\omega)$.

(5) If α is a square integrable random variable on $(\Omega, L(\mathscr{A}), L(P))$, then for $L(\lambda)$-almost all $t \in T$, it is orthogonal, to e_t, and $\int_\Omega \alpha(\omega) f(t,\omega) \mathrm{d} L(P)(\omega) = \sum_{n=1}^{\infty} \lambda_n (\int_\Omega \alpha(\omega) \varphi_n(\omega) \mathrm{d} L(P)(\omega)) \psi_n(t)$.

Note that the principal part $E(f|_{\mathscr{U}})$ in the above expression already captures the relevant correlational structure of f by Lemma 4.1. The residual part e also has low intercorrelation. This is a requirement for the error term in the classical factor models (see, for example, Basilevsky, 1994). Thus, it is natural to adopt the terminologies of factor analysis to call the above structural result a hyperfinite factor model with factors φ_n and factor loadings ψ_n. Since the factors are endogenously derived from the autocorrelation function, the model is also called a endogenous hyperfinite factor model. As noted in Basilevsky (1994), the main objective of the factor analysis is to find a small set of latent random variables to explain the systematic behavior of the directly observed random variables. In our case, if the hyperfinite number of random variables f_t are directly observed, then the countably many latent random variables φ_n can be used to explain the correlational structure of f. By the fact that a countable collection is significantly smaller than a collection with a hyperfinite number of elements, this reduction procedure is indeed effective. Since the residual part has almost surely orthogonal random variables (i.e., the uncertainty or risk is on the individual level) and satisfies the law of large numbers in the strong sense as described in (4) and (5), so that it will disappear in most useful circumstances, the phenomenon of ‘aggregation removes individual uncertainty’ thus appears in a very general setting.

5. Random variables in metric spaces and the law

5.1. Introduction

In Sections 5.2, 5.3 and 5.4, we consider metric space valued random variables and processes. Section 5.2 includes a version of the law for such processes. Since Loeb counting spaces have proven to be convenient in many places (see, for example, Anderson, 1976; Keisler, 1977 and Keisler, 1984), we present in Section

5.3 several results relating processes on general Loeb product spaces to processes involving Loeb counting spaces through measure preserving mappings. These results are then used in Sections 6 and 8 to obtain other structural results. Section 5.4 contains a version of the law for a hyperfinite number of discrete parameter stochastic processes.

5.2. The law for metric space valued processes

Let f be a process from a Loeb product space $(T \times \Omega,\ L(\mathscr{T} \otimes \mathscr{A}),\ L(\lambda \otimes P))$ to a separable metric space X, and $\{O_n\}_{n=1}^{\infty}$ a countable open base of X. As mentioned before, all processes on a Loeb product space are assumed to be measurable with respect to the relevant Loeb product algebra. Thus for each $n \geq 1$, the measurability of $f^{-1}(O_n)$ in $L(\mathscr{T} \otimes \mathscr{A})$ together with the Fubini theorem implies that for $L(\lambda)$-almost all $t \in T$, $f_t^{-1}(O_n)$ is in $L(\mathscr{A})$. Thus, for $L(\lambda)$-almost all $t \in T$, $f_t^{-1}(O_n)$ is in $L(\mathscr{A})$. Thus, for $L(\lambda)$-almost all $t \in T$, f_t^{-1} (O_n) is in $L(\mathscr{A})$ for every $n \geq 1$. It implies that the functions f_t are Loeb measurable for $L(\lambda)$-almost all $t \in T$. Similarly, for $L(P)$-almost all $w \in \Omega$, f_ω is Loeb measurable.

Definition 5.1. Let f be a process from $(T \times \Omega,\ L(\mathscr{T} \otimes \mathscr{A}),\ L(\lambda \otimes P))$ to a separable metric space X. The random variables f_t are said to be almost surely pairwise independent, if for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, the pair f_{t_1} and f_{t_2} are independent random variables.

Using the above notion of almost sure pairwise independence, the following theorem provides an *exact* version of the law of large numbers for metric space valued processes in terms of empirical distributions.

Theorem 5.2. *Let f be a process from $(T \times \Omega,\ L(\mathscr{T} \otimes \mathscr{A}),\ L(\lambda \otimes P))$ to a separable metric space X. Assume that the random variables f_t are almost surely pairwise independent. Then f satisfies the law in distribution. This means that for $L(P)$-almost all $\omega \in \Omega$, the distribution μ_ω on X induced by the sample function f_ω on T is equal to the distribution μ on X induced by f viewed as a random variable on $T \times \Omega$.*

Proof: Fix a countable open base $\{Q_m\}_{m=1}^{\infty}$ for X. Let $\{O_n\}_{n=1}^{\infty}$ be a list of all the finite intersections of the sets Q_m. By the independence assumption, we know that for each $n \geq 1$, the random variables $(\chi_{f^{-1}(O_n)})_t$ in the process $\chi_{f^{-1}(O_n)}$ are almost surely uncorrelated. By Theorem 3.8, there is an $L(P)$-null set E_n such that for any $\omega \notin E_n$, $L(\lambda)(f_\omega^{-1}(O_n)) = L(\lambda \times P)(f^{-1}(O_n))$. Let E be the union of all the E_n. Then E is still an $L(P)$-null set. For any $\omega \notin E$,

$$\mu_\omega(O_n) = L(\lambda)\big(f_\omega^{-1}(O_n)\big) = L(\lambda \times P)\big(f^{-1}(O_n)\big) = \mu(O_n)$$

for all $n \geq 1$. Since the class of all the sets O_n generates the Borel algebra, and is also closed under formation of finite intersections, i.e., a π-system, it follows from Dynkin's $\pi - \lambda$ theorem [10] that $\mu_\omega = \mu$ for any $\omega \notin E$. $\square$

As before, if A is an atom of $L(\lambda)$, then there is a random variable ϕ with distribution ν such that $f_t = \phi$ for $L(\lambda)$-almost all $t \in A$. The almost sure pairwise independence assumption on the process f implies that the random variable ϕ is independent to itself. Suppose ϕ is not constant. Then there is a Borel set D in X such that $0 < \nu(D) = L(P)(\phi^{-1}(D)) < 1$. The regularity of ν as shown in Billingsley (1968) (p. 7) yields the fact that the set D can be taken as an open set O. Thus

$$L(P)\left(\phi^{-1}(O)\right) = L(P)\left(\phi^{-1}(O) \cap \phi^{-1}(O)\right)$$
$$\neq L(P)\left(\phi^{-1}(O)\right) \cdot L(P)\left(\phi^{-1}(O)\right),$$

which is in contradiction with the independence of ϕ to itself. Thus, for $L(\lambda)$-almost all $t \in A$, f_t is a constant function.

5.3. Representations on hyperfinite Loeb counting spaces

Among Loeb spaces, the Loeb counting spaces over hyperfinite sets are the ones with the simplest structure. Many constructions are based on the Loeb counting spaces (see Albeverio et al., 1986; Anderson, 1976; Hoover and Perkins, 1983; Keisler, 1984 and Stroyan and Bayod, 1986). They also have some convenient special properties (see, Keisler, 1984, 1988, and Proposition 5.7 below).

In this section, we shall first show how to represent an internal class of functions to an internal class of functions on an internal counting space. Let T be an internal parameter space and let $\{(\Omega_t, A_t, P_t): t \in T\}$ be an collection of internal probability spaces. Assume that each space $(\Omega_t, \mathscr{A}_t, P_t)$ is atomless, that is, for any $\varepsilon \in \mathbb{R}^+$ and for any $B \in \mathscr{A}_t$, if $P_t(B) > \varepsilon$, then, there is a $C \in \mathscr{A}_t$ such that $P_t(C) < \varepsilon/2$. This version of atomlessness is equivalent to the usual atomlessness of the standard Loeb measure $L(P_t)$.

Proposition 5.3. *Let X be a separable metric space with a totally bounded metric d and $\{F_t: t \in T\}$ an internal class of functions such that for each $t \in T$, F_t is a function from Ω_t to ${}^*X$. Then, there is an internal counting probability space $(\Omega, \mathscr{A}, P)$, and two internal classes $\{\pi_t: t \in T\}$, $\{G_t: t \in T\}$ of functions such that for each $t \in T$, π_t is a measure preserving mapping from the Loeb space $(\Omega_t, L(\mathscr{A}_t)$,*

[10] It implies that two probability measures which agree on a π-system must be the same; see Cohn (1980), p. 44 or Durrett (1991), p. 404.

$L(P_t)$) to the Loeb counting space (Ω, $L(\mathscr{A})$, $L(P)$), G_t is a function from Ω to ${}^*X$, and

$$L(P_t)\left(\left\{\omega_t \in \Omega_t : {}^*\mathrm{d}\left(F_t(\omega_t), G_t(\pi_t(\omega_t))\right) \simeq 0\right\}\right) = 1.$$

Proof: Fix an arbitrary natural number $m > 2$. Choose a finite sequence $x_1, \dots, x_\ell$ of points in X such that the open balls $B(x_i, 1/m)$, $i = 1, \dots, \ell$ form a cover of X. Define an internal set valued function on $T \times \{1, \dots, \ell\}$ by letting

$$C_t^i = F_t^{-1}\left({}^*B(x_i, 1/m) - \cup_{j=1}^{i-1}{}^*B(x_j, 1/m)\right).$$

Then for each $t \in T$, the sets C_t^i form an internal partition of Ω_t. Pick a natural number $n > m\ell$. Since each $(\Omega_t, \mathscr{A}_t, P_t)$ is atomless, we can choose an internal set valued function on $T \times \{1, \dots, l\}$ such that $D_t^i \subseteq C_t^i$, $P(D_t^i)$ is infinitely close to an integral multiple of $1/n$, and $P_t(C_t^i - D_t^i) < 1/n$. Let $D_t^0 = \cup_{i=1}^{\ell}(C_t^i - D_t^i)$. Then $P(D_t^0) < 1/m$. Partition the sets D_t^i to obtain an internal partition $B_t^1, \dots, B_t^n$ of Ω_t such that $|P(B_t^j) - 1/n| \simeq 0$ for $1 \le j \le n$ and $t \in T$. By the construction of the set B_t^j, if B_t^j is contained in some D_t^i for $1 \le i \le \ell$, then the internal variation $\mathrm{var}(F_t, B_t^j)$ of F_t on B_t^j is less than $1/m$. Thus the internal cardinality of the set $\{j: \mathrm{var}(F_t, B_t^j) > 1/m\}$ divided by n is less than $1/m$.

By the spillover principle in Hurd and Loeb (1985) (p. 101), we can find m, $n \in {}^*\mathbb{N}_\infty$ with $n > m$ such that there is an internal function of internal partitions $B_t^1, \dots, B_t^n$ of Ω_t with the following properties:

1. $|P(B_t^j) - 1/n| < 1/n^2$ for $1 \le j \le n$ and $t \in T$;
2. $\frac{1}{n}|\{j: \mathrm{var}(F_t, B_t^j) > 1/m\}| < 1/m$.

Let $\Omega = \{1, 2, \dots, n\}$ and $(\Omega, \mathscr{A}, P)$ be the internal counting probability space on Ω. Define π_t: $\Omega_t \to \Omega$ by letting $\pi_t(\omega_t) = j$ if $\omega_t \in B_t^j$. For any $C \in \mathscr{A}$, (1) implies that

$$|P_t\left(\pi_t^{-1}(C)\right) - P(C)| = |\sum_{j \in C} P_t\left(B_t^j\right) - \sum_{j \in C} \frac{1}{n}| \le \sum_{j \in C} \frac{1}{n^2} \le \frac{1}{n}.$$

Thus for any $D \in L(\mathscr{A})$, $L(\mathrm{P}_t)(\pi_t^{-1}(D)) = L(P)(D)$, and hence π_t is measure preserving from $(\Omega_t, L(\mathscr{A}_t), L(P_t))$ to $(\Omega, L(\mathscr{A}), L(P))$. Choose an internal function c_t from Ω to Ω_t such that $c_t(j) \in B_t^j$. Define a mapping G_t from Ω to ${}^*X$ by $G_t(j) = F_t(c_t(j))$ for $j = 1, \dots, n$. If $\mathrm{var}(F_t, B_t^j) \le 1/m$, then for $\omega_t \in B_t^j$,

$$\begin{aligned} {}^*d\left(F_t(\omega_t), G_t(\pi_t(\omega_t))\right) &= {}^*d\left(F_t(\omega_t), G_t(j)\right) \\ &= {}^*\mathrm{d}\left(F_t(\omega_t), F_t(c_t(j))\right) \le 1/m. \end{aligned}$$

Hence, (2) deduces that

$$\begin{aligned} &L(P_t)\left(\left\{\omega_t \in \Omega_t : \mathrm{d}\left(F_t(\omega_t), G_t(\pi_t(\omega_t))\right) \simeq 0\right\}\right) \\ &\quad \ge L(P)\left(\left\{j : var\left(F_t, B_t^j\right) \le 1/m\right\}\right) = 1, \end{aligned}$$

and we are done. □

The following corollary says that a sequence of random variables on (possibly different) atomless Loeb spaces can be expressed as the compositions of random variables on a common hyperfinite Loeb counting space together with some Loeb measure preserving mappings.

Corollary 5.4. *Let $\{(\Omega_n, L(\mathscr{A}_n), L(P_n))\}_{n=1}^{\infty}$ be a sequence of atomless Loeb probability spaces and $\{\varphi_n\}_{n=1}^{\infty}$ a sequence of random variables, where for each $n \geq 1$, φ_n is from $(\Omega_n, L(\mathscr{A}_n), L(P_n))$ to a separable metric space X. Then there is a hyperfinite Loeb counting space $(\Omega, L(\mathscr{A}), L(P))$, a sequence of mappings $\{\pi_n\}_{n=1}^{\infty}$, and a sequence of random variables $\{\psi_n\}_{n=1}^{\infty}$ from $(\Omega, L(\mathscr{A}), L(P))$ to X such that $\varphi_n = \psi_n(\pi_n)$ for each $n \geq 1$, where π_n is measure preserving from $(\Omega_n, L(\mathscr{A}_n), L(P_n))$ to $(\Omega, L(\mathscr{A}), L(P))$ for each $n \geq 1$.*

Proof: For each $n \geq 1$, let F_n be in internal lifting of φ_n. Extend the sequences $\{(\Omega_n, \mathscr{A}_n, P_n)\}_{n=1}^{\infty}$ and $\{F_n\}_{n=1}^{\infty}$ to internal sequences of probability spaces and random variables respectively. Choose $k \in {}^*\mathbb{N}_{\infty}$ such that for each $n \leq k$, $(\Omega_n, \mathscr{A}_n, P_n)$ is atomless. Let $T = \{1,2,\dots,k\}$. The rest is clear from the previous proposition. □

We can also represent a process with an atomless sample space to a process with a hyperfinite Loeb counting sample space.

Corollary 5.5. Let f be a process from $(T \times \Omega_1, L(\mathscr{T} \otimes \mathscr{A}_1), L(\lambda \otimes P_1))$ to a separable metric space X. Assume that $L(P_1)$ is atomless. Then there is an internal counting probability space $(\Omega_2, \mathscr{A}_2, P_2)$, an internal function π from $T \times \Omega_1$ to Ω_2, and a process g from $(T \times \Omega_2, L(\mathscr{T} \otimes \mathscr{A}_2), L(\lambda \otimes P_2))$ to X such that for $L(\lambda)$-almost all $t \in T$, $f_t = g_t(\pi_t)$, where π_t is measure preserving from $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ to $(\Omega_2, L(\mathscr{A}_2), L(P_2))$ for each $t \in T$.

Proof: Let d be a totally bounded metric on X and F an internal lifting of f from $T \times \Omega_1$ to ${}^*X$. Proposition 5.3 implies that there is an internal counting probability space $(\Omega_2, \mathscr{A}_2, P_2)$, an internal function π from $T \times \Omega_1$ to Ω_2, and an internal process G from $T \times \Omega_2$ to X such that

$$L(P_1)\left(\left\{\omega_1 \in \Omega_1 : {}^*\mathrm{d}\left(F_t(\omega_1), G_t(\pi_t(\omega_1))\right) \simeq 0\right\}\right) = 1,$$

where π_t is measure preserving from $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ to $(\Omega_2, L(\mathscr{A}_2), L(P_2))$ for each $t \in T$. By the construction of G, it is clear that $G(t, \omega_2)$ is near standard for $L(\lambda \otimes P_2)$-almost all $(t, \omega_2) \in T \times \Omega_2$. Let $g = {}^\circ G$. Then, for $L(\lambda)$-almost all $t \in T$, $f_t = g_t(\pi_t)$. □

It was shown by Keisler (1984, 1988) that for any two given random variables from a hyperfinite Loeb counting space to a complete separable metric space, if they have the same distribution, then there is an internal permutation of the internal counting space sending one random variable to the other. The property is called the homogeneity of hyperfinite Loeb counting spaces. Note that the property also holds when the target space is not complete. The following proposition provides a characterization for those random variables on general atomless Loeb spaces with the same induced distributions on the target space.

Proposition 5.6. *Let φ and ψ be random variables from Loeb spaces $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ and $(\Omega_2, L(A_2), L(P_2))$ to a separable metric space X respectively. Assume that $L(P_1)$ and $L(P_2)$ are atomless. Then the following are equivalent:*

(1) φ and ψ have the same distribution;

(2) One can find a Loeb counting space $(\Omega, L(\mathscr{A}), L(P))$, two measure preserving mappings π_1 and π_2 from $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ and $(\Omega_2, L(\mathscr{A}_2), L(P_2))$ respectively to $(\Omega, L(\mathscr{A}), L(P))$, and a random variable α from $(\Omega, L(\mathscr{A}), L(P))$ to X such that $\phi = \alpha o \pi_1$ and $\psi = \alpha o \pi_2$.

Proof: (2) $\Rightarrow$ (1) is clear. For (1) $\Rightarrow$ (2), note that Corollary 5.4 implies that there is a Loeb counting space $(\Omega, L(\mathscr{A}), L(P))$, two measure preserving mappings q_1 and q_2 from $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ and $(\Omega_2, L(\mathscr{A}_2), L(P_2))$ respectively to $(\Omega, L(\mathscr{A}), L(P))$, and random variables β_1, β_2 from $(\Omega, L(\mathscr{A}), L(P))$ to X such that $\varphi = \beta_1 \circ q_1$ and $\psi = \beta_2 \circ q_2$. Since the random variables β_1, β_2 on the hyperfinite Loeb counting space $(\Omega, L(\mathscr{A}), L(P))$ have the same distribution, by the homogeneity property, there is an internal permutation p of Ω such that $\beta_1 \circ p = \beta_2$. Let $\pi_1 = q_1$, $\pi_2 = p \circ q_2$, and $\alpha = \beta_1$. Then $\varphi = \alpha \circ \pi_1$ and $\psi = \alpha \circ \pi_2$. □

Next, we present a result on the parametric homogeneity of hyperfinite Loeb counting spaces. This property will be used in Section 6 to characterize processes with the same hyper distributions.

Proposition 5.7. *Let f and g be processes from $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ to a separable metric space X, where $(\Omega, \mathscr{A}, P)$ is an internal counting probability space. Assume that for $L(\lambda)$-almost all $t \in T$, f_t and g_t have the same distribution. Then there is an internal function π from $T \times \Omega$ to Ω such that for each $t \in T$, π_t is a permutation of Ω, and for $L(\lambda)$-almost all $t \in T$, $g_t = f_t \circ \pi_t$.*

Proof: Let d be a totally bounded metric on X. Fix a natural number $m \geq 1$. Choose a Borel measurable partition $A^1, \ldots, A^n$ of X such that the diameter of each A^i is less than $1/m$. Let $B^i = f^{-1}(A^i)$ and $C^i = g^{-1}(A^i)$. Then both the sets B^i and the sets C^i form Loeb measurable partitions of $T \times \Omega$. It is clear that for each $t \in T$, the sections $B^i_t = f_t^{-1}(A^i)$ and $C^i_t = g_t^{-1}(A^i)$. Thus, for $L(\lambda)$-almost all $t \in T$, $L(P)(B^i_t) = L(P)(C^i_t)$ for all $i = 1, \ldots, n$. For each $1 \leq i \leq n$, take internal sets D^i and E^i in $T \times \Omega$ such that the symmetric differences between B^i and D^i and between C^i and E^i are $L(\lambda \otimes P)$-null. By modifying the sets D^i and

E^i, we can assume that the sets D^i as well as the sets E^i form internal partitions of $T \times \Omega$. It is clear that for $L(\lambda)$-almost all $t \in T$ $P(D^i_t) \simeq P(E^i_t)$ for all $i = 1, \ldots, n$. For each $h \geq 1$ in ${}^*\mathbb{N}$, let

$$T_h = \left\{ t: \max_{1 \leq i \leq n} |P\left(D^i_t\right) - P\left(E^i_t\right)| < 1/h \right\}.$$

Then $\lambda(T_h) > 1 - 1/h$ for any positive integer h. By spillover, it holds for some $h \in {}^*\mathbb{N}_\infty$. Fix such an h. For each $t \in T_h$ and each $i = 1, \ldots, n$, let $c^i_t = \min\{P(D^i_t), P(E^i_t)\}$. Since P is the counting probability measure on Ω, we can choose internal subsets K^i_t of D^i_t and L^i_t of E^i_t such that $P(K^i_t) = P(L^i_t) = c^i_t$. For each $t \in T_h$, define $K^0_t = \Omega - \cup^n_{i=1} K^i_t$ and $L^0_t = \Omega - \cup^n_{i=1} L^i_t$. Then we have $P(K^0_t) = P(L^0_t) \simeq 0$. Hence, there is an internal function π^m from $T \times \Omega$ to Ω such that π^m_t is an internal permutation of Ω for each $t \in T$ and $\pi^m_t(L^i_t) = K^i_t$ for each $t \in T_h$. Therefore, for almost all $t \in T_h$, $d(f_t(\pi^m_t(\omega)), g_t(\omega)) < 1/m$ for $L(P)$-almost all $\omega \in \Omega$. Let F and G be internal liftings of f and g respectively. Then

$$(\lambda \otimes P)\left(\left\{(t,\omega): {}^*\mathrm{d}\left(F\left(t,\pi^m_t(\omega)\right), G(t,\omega)\right) < \frac{1}{m}\right\}\right) > 1 - \frac{1}{m}.$$

By $\aleph_1$-saturation, there is an $m \in {}^*\mathbb{N}_\infty$ and an internal function π^m from $T \times \Omega$ to Ω such that the functions π^m_t are internal permutations on Ω and the above inequality holds. Fix such an m. Let $\pi = \pi^m$. Then π has the required property. □

5.4. The law for discrete parameter stochastic processes

Since a discrete parameter stochastic process as discussed in Doob (1953) can be viewed as a random variable taking values in the relevant product space, it is easy to obtain a version of the law of large numbers for a hyperfinite number of such processes. For simplicity, we only state the results for the case that the parameter space is the set $\mathbb{Z}^+$ of positive integers. Here, a hyperfinite number of discrete parameter stochastic processes on a Loeb product space $(T \times \Omega,\ L(\mathcal{T} \otimes \mathcal{A}),\ L(\lambda \otimes P))$ simply means a hyperprocess f on $T \times \Omega \times \mathbb{Z}^+$ such that for each $n \in \mathbb{Z}^+$, f_n is a process on $(T \times \Omega,\ L(\mathcal{T} \otimes \mathcal{A}),\ L(\lambda \otimes P))$. When we regard the processes f_n as random variables on the Loeb product space, the hyperprocess f can also be viewed as a discrete parameter stochastic process with $\mathbb{Z}^+$ as the parameter set. For each $t \in T$, f_t is also a discrete parameter stochastic process with sample space Ω and parameter space $\mathbb{Z}^+$. The f_t will be called the stochastic processes in the hyperprocess f. For each $\omega \in \Omega$, the sample function f_ω on $T \times \mathbb{Z}^+$ is again a discrete parameter stochastic process with sample space T and parameter space $\mathbb{Z}^+$. The sample functions f_ω will also be called the empirical processes in the hyperprocess f.

Note that two discrete parameter stochastic processes $\varphi = \{\varphi_n\}^\infty_{n=1}$ and $\psi = \{\psi_n\}^\infty_{n=1}$ on some possibly different sample spaces are said to have the same finite dimensional distributions, if for any positive integer indexes $i_1, \dots, i_m$, the random variables $(\varphi_{i_1}, \dots, \varphi_{i_m})$ and $(\psi_{i_1}, \dots, \psi_{i_n})$ have the same distribution. When φ and ψ are on the same sample space, we say they are independent, if the random variables $(\varphi_{i_1}, \dots, \varphi_{i_m})$ and $(\psi_{j_1}, \dots, \psi_{j_n})$ are independent for any positive integer indexes $i_1, \dots, i_m$ and $j_1, \dots, j_n$.

Theorem 5.8. *Let f be a hyperprocess from $(T \times \Omega \times \mathbb{Z}^+)$ to a separable metric space X. Assume that the discrete parameter stochastic processes f_t are almost surely pairwise independent, i.e., for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, the stochastic processes f_{t_1} and f_{t_2} are independent. Then for $L(P)$-almost all $\omega \in \Omega$, the discrete parameter empirical process f_ω on $T \times \mathbb{Z}^+$ and the hyperprocess f viewed as a discrete parameter stochastic process on $(T \times \Omega) \times \mathbb{Z}^+$ have the same finite dimensional distributions.*

Proof: Define an X^∞-valued process g on $T \times \Omega$ by letting $g(t, \omega) = \{f_n(t, \omega)\}_{n=1}^\infty$. Then, it is clear that the almost sure pairwise independence of the X-valued discrete parameter stochastic processes f_t is equivalent to the almost sure pairwise independence of the X^∞-valued random variables g_t. Hence, Theorem 5.2 implies that for $L(P)$-almost all $w \in \Omega$, the distribution μ_ω on X^∞ induced by the random variable g_ω on T is equal to the distribution μ on X^∞ induced by the random variable g on $T \times \Omega$. It is also obvious that if $\mu_\omega = \mu$, then the discrete parameter stochastic process f_ω on $T \times \mathbb{Z}^+$ and the discrete parameter stochastic process f on $(T \times \Omega) \times \mathbb{Z}^+$ have the same finite dimensional distributions. Hence the result follows. □

Remark 5.9. As indicated by the proof of the above theorem, any result for a hyperfinite number of random variables taking values in a general metric space can be restated for a hyperfinite number of discrete parameter stochastic processes. Such a result will not be written down explicitly afterwards (except for the result in Corollary 7.17).

6. Some universal phenomena

6.1. Introduction

Before we can claim that the results presented in Sections 3–5 for processes with almost surely uncorrelated or almost surely pairwise independent random variables are not vacuous, the important question that we have to address is: Do such processes always exist? Of course, one can transfer a standard sequence of independent random variables to obtain a hyperfinite sequence of $*$-independent internal random variables (see Anderson, 1991; Keisler, 1977; Nelson, 1987 and

Stroyan and Bayod, 1986). However, as noted in Section 2, such processes are not very interesting from our point view.[11] Section 6.2 contains the following universality of atomless Loeb product spaces: for any two given atomless Loeb spaces and for any given hyper distribution on the space of probability measures on a Polish space, one can construct a process with almost surely pairwise independent random variables from the relevant Loeb product space to the Polish space such that the hyper distribution of the process is the given one, where the hyper distribution of a process measures the variety of the distributions of the random variables in the process. This result guarantees the abundance of processes with almost surely pairwise independent random variables.

As noted earlier, the product Loeb space $(T \times \Omega, L(\mathcal{T}) \otimes L(\mathcal{A}), L(\lambda) \otimes L(P))$ is contained in the Loeb product space $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$. 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 Albeverio et al., 1986, p. 74 or Cutland, 1983, p. 557) Proposition 6.6 in Section 6.3, in fact, says much more: the stated inclusion is proper if and only if neither $L(\lambda)$ nor $L(P)$ is purely atomic.

In Section 6.4, some constructions in Section 6.2 are used to show the non-separability of the space $V^p_{\mathcal{U}}$ of all $\mathcal{L}^p$-integrable centered processes which satisfy the consistent law of large numbers.

6.2. Universality of atomless loeb product spaces

As remarked in Section 5.2, for a process f from a Loeb product space $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$ to a separable metric space X, the functions f_t are Loeb measurable for $L(\lambda)$-almost all t. Define a mapping Ψ from T to the space $\mathcal{M}(X)$ of Borel probability measures on X by letting $\Psi(t)$ be the distribution on X induced by f_t, i.e. $\Psi(t) = L(\lambda)(f_t)^{-1}$. The following lemma shows that Ψ is Loeb measurable.

Lemma 6.1. Let f and Ψ be defined as above. Then Ψ is measurable from $(T, L(\mathcal{T}), L(\lambda))$ to the space $\mathcal{M}(X)$ with the topology of weak convergence of measures.

Proof: Choose a totally bounded metric d on X. Then, as noted in Parthasarathy (1967) (p. 43), the space $U(X)$ of uniformly continuous functions on (X, d) is separable with the supremum norm. Take a sequence $\{\varphi_n\}_{n=1}^{\infty}$ in the unit ball of

[11] On the other hand, the transferred law of large numbers as well as relevant results for $*$-independent random variables does have important applications in the hyperfinite model theory of Keisler; see Keisler (1977).

$U(X)$ such that the vector space spanned by the sequence is dense in $U(X)$. Define a metric ρ on $\mathscr{M}(X)$ by letting

$$\rho(\mu,\nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int_X \varphi_n \mathrm{d}\mu - \int_X \varphi_n \mathrm{d}\nu \right|.$$

Then ρ induces the weak topology on $\mathscr{M}(X)$ (see Parthasarathy, 1967, p. 43). The Fubini theorem can be used to obtain that for each $n \geq 1$, $\int_T \varphi_n \circ f_t \mathrm{d}L(P)$ is Loeb measurable on T. Thus, $\rho(\Psi(t), \nu)$ is Loeb measurable on T for any $\nu \in \mathscr{M}(X)$. This implies that $\Psi^{-1}(B)$ is $L(\mathscr{T})$-measurable for any open ball B in $(\mathscr{M}(X), \rho)$. Hence Ψ is Loeb measurable on T. □

Since Ψ is Loeb measurable from $(T, L(\mathscr{T}))$ to $\mathscr{M}(X)$, we can define the distribution of Ψ on the space $\mathscr{M}(X)$. Such a distribution is called the hyper distribution of the process f, which is a Borel probability measure on the space of Borel probability measures on the separable metric space X. Note that one can also define the hyper distribution of a process on general probability spaces provided that the Fubini type result holds. The hyper distribution of a process measures the variety of the distributions of the random variables in the process. Unlike the finite dimensional distributions of a process, it does not provide any information about the correlations between the random variables. However, the notion is helpful in the formulation of the following type of universality result on atomless Loeb product spaces. It states that one can construct processes on the Loeb space of the internal product space of *any* two given atomless internal probability spaces with almost surely pairwise independent random variables and with any given hyper distribution. Here we require X to be complete, since a version of the Prohorov theorem (see Billingsley, 1968, p. 37, Theorem 6.2) will be needed in the proof.

Theorem 6.2. *Let X be a Polish space and μ any Borel probability measure on the space $\mathscr{M}(X)$ of Borel probability measures on X, where $\mathscr{M}(X)$ is endowed with the topology of weak convergence of measures. If both $L(\lambda)$ and $L(P)$ are atomless, then there is a process f from $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ to X such that the random variables f_t are almost surely pairwise independent and the hyper distribution of f on $\mathscr{M}(X)$ is the given μ.*

Proof: For any given positive integer h, since both $L(\lambda)$ and $L(P)$ are atomless, we can choose a system of internal sets $\{B_{i_1 \dots i_h}: 1 \leq i_1, \dots, i_h \leq h\}$ and another system of internal sets $\{A_1, \dots, A_h\}$ such that the first forms an internal partition of Ω with $P(B_{i_1 \dots i_h}) \simeq 1/h^h$, and the second forms an internal partition of T with $\lambda(A_i) \simeq 1/h$. By the usual spillover principle, we can find some $n \in {}^*\mathbb{N}_\infty$ such that there is an internal partition $\{B_{i_1 \dots i_n}: 1 \leq i_1, \dots, i_n \leq n\}$ of Ω with $|P(B_{i_1, \dots, i_n}) - 1/n^n| < 1/n^{2n}$, and an internal partition $\{A_1, \dots, A_n\}$ of T with $|\lambda(A_i) - 1/n| < 1/n^2$. We fix such a hyperinteger n.

As in Lemma 6.1, choose a sequence $\{\varphi_p\}_{p=1}^{\infty}$ of bounded continuous functions on X and define a metric ρ on $\mathscr{M}(X)$ accordingly. Since $(\mathscr{M}(X), \rho)$ is again a separable metric space, we can choose a sequence $\{\Phi_p\}_{p=1}^{\infty}$ of bounded continuous functions on $\mathscr{M}(X)$ and define a metric σ on the space $\mathscr{M}^{\mathscr{M}}(X)$ of hyper distributions as in the case of $\mathscr{M}(X)$. Fix any $\varepsilon \in \mathbb{R}^+$. Since hyper distributions with finite support are dense in $\mathscr{M}^{\mathscr{M}}(X)$ (see Parthasarathy, 1967, p. 44), there is a hyper distribution μ_ε having finite support such that $\sigma(\mu_\varepsilon, \mu) < \varepsilon$. Let $\mu_\varepsilon = \alpha_1 \delta(\nu_1) + \ldots + \alpha_m \delta(\nu_m)$, where $\alpha_1, \ldots, \alpha_m \geq 0$, $\alpha_1 + \ldots, \alpha_m = 1$, $\nu_1, \ldots, \nu_m$ are probability measures on X, and the hyper distributions $\delta(\nu_j)$ are the corresponding Dirac measures concentrated on ν_j. Let $S = \{1, \ldots, n\}$ and endow the internal counting probability measure τ on S. Since $L(\tau)$ is atomless, we can find for each $1 \leq j \leq m$, a Loeb measurable mapping h_j from S to X such that the distribution $L(\tau)h_j^{-1}$ on X induced by h_j is ν_j. Take an internal partition $\{S_1, \ldots, S_m\}$ of S such that $L(\tau)(S_j) = \alpha_j$ for each $1 \leq j \leq m$. Let H_j be an internal lifting of h_j. Since for each $p \geq 1$ and $1 \leq j \leq m$, ${}^*\varphi_p \circ H_j$ is an internal lifting of $\varphi_p \circ h_j$, by Loeb's theorem (Loeb, 1975), we have

$$\int_S {}^*\varphi_p(H_j) \mathrm{d}\tau \simeq \int_S \varphi_p(h_j) \mathrm{d}L(\tau) = \int_X \varphi_p \mathrm{d}\nu_j.$$

Hence, ${}^*\rho(\tau H_j^{-1}, \nu_j) \simeq 0$ for all $j = 1, \ldots, m$.

Next, we define an X-valued process g^ε on $S \times S$ as follows. First, consider an internal process G^ε on $S \times S$ by letting $G^\varepsilon(r, s) = H_j(s)$ for $r \in S_j$ and $s \in S$. Let g^ε be the standard part of G^ε. It is clear that for each $r \in S_j$, $(g^\varepsilon)_r = h_j$, and hence $L(P)((g^\varepsilon)_r)^{-1} = \nu_j$. Since $L(\tau)(S_j) = \alpha_j$, it is then easy to see that the hyper distribution of the process g^ε is μ_ε. For each $r \in S_j$, we have $(G^\varepsilon)_r = H_j$ and $(g^\varepsilon)_r = h_j$, and hence ${}^*\rho(\tau((G^\varepsilon)_r)^{-1}, \nu_j) \simeq 0$. For each $p \geq 1$, since Φ_p is continuous, ${}^*\Phi_p(\tau((G^\varepsilon)_r)^{-1}) \simeq \Phi_p(\nu_j)$ holds for $r \in S_j$. Therefore,

$$\int_S {}^*\Phi_p\left(\tau\left((G^\varepsilon)_r\right)^{-1}\right) \mathrm{d}\tau(r) \simeq \sum_{j=1}^{m} \int_{S_j} \Phi_p(\nu_j) \mathrm{d}L(\tau)(r) = \int_{\mathscr{M}(X)} \Phi_p \mathrm{d}\mu_\varepsilon.$$

Now, consider the sequence $\{G^{1/q}\}_{q=1}^{\infty}$. Since the sequence $\{\mu_{1/q}\}_{q=1}^{\infty}$ of hyper distributions converges weakly to μ, it follows that for each $p \geq 1$,

$$\lim_{q \to \infty} \int_{\mathscr{M}(X)} \Phi_p \mathrm{d}\mu_{1/q} = \int_{\mathscr{M}(X)} \Phi_p \mathrm{d}\mu.$$

Hence, for any pair (p, k) of positive integers, there is $N_{pk} \in \mathbb{N}$ such that for any $q \geq N_{pk}$,

$$\left| \int_S {}^*\Phi_p\left(\tau\left((G^{1/q})_r\right)^{-1}\right) d\tau(r) - \int_{\mathscr{M}(X)} \Phi_p \mathrm{d}\mu \right| < \frac{1}{k}. \tag{6-1}$$

Since $\mathscr{M}(X)$ is also a Polish space and $\{\mu_{1/q}\}_{q=1}^{\infty}$ is a tight sequence, by Prohorov's theorem, there is for each $\ell \geq 1$, a compact set C_ℓ in $\mathscr{M}(X)$ such that $\mu_{1/q}(C_\ell) > 1 - 1/\ell$ for all $q \geq 1$. The same theorem can be applied to the compact set C_ℓ of measures in $\mathscr{M}(X)$ to obtain a compact set K_ℓ in X such that for any $\nu \in C_\ell$, $\nu(K_\ell) > 1 - 1/\ell$. Let

$$S_\ell^q = \left\{ r \in S : L(\tau)\left(g^{1/q}\right)_r^{-1} \in C_\ell \right\}.$$

Since the hyper distribution of $g^{1/q}$ is $\mu_{1/q}$, $L(\tau)(S_\ell^q) > 1 - 1/\ell$. Hence for all q, $\ell \geq 1$,

$$\begin{aligned} L(\tau \otimes \tau)\left(\left(g^{1/q}\right)^{-1}(K_\ell)\right) &= \int_S L(\tau)\left(\left(g^{1/q}\right)_r^{-1}(K_\ell)\right) \mathrm{d}L(\tau)(r) \\ &\geq \int_{S_\ell^q} L(\tau)\left(g^{1/q}\right)_r^{-1}(K_\ell) \mathrm{d}L(\tau)(r) \\ &\geq \int_{S_\ell^q} (1 - 1/\ell) \mathrm{d}L(\tau)(r) > (1 - 1/\ell)^2. \end{aligned} \tag{6-2}$$

By the $\aleph_1$-saturation principle, the sequence $\{G^{1/q}\}_{q=1}^\infty$ can be extended to an internal sequence $\{G^{1/q}\}_{q \in {}^*\mathbb{N}}$ of processes on $S \times S$. By the Permanence Principle in Hurd and Loeb (1985) (p. 100), for any pair (p, k) of positive integers, there is an $M_{pk} \in {}^*\mathbb{N}_\infty$ such that for any $N_{pk} \leq q \leq M_{pk}$, the inequality in Eq. (6-1) still holds. By spillover, we can claim that there is an $u_1 \in {}^*\mathscr{N}_\infty$ such that $u_1 \leq M_{pk}$ for all p, $k \geq 1$. Hence, for every positive integer p and for every $q \in {}^*\mathbb{N}_\infty$ with $q \leq u_1$, we have

$$\int_S {}^*\Phi_p\left(\tau\left(G^{1/q}\right)_r^{-1}\right) \mathrm{d}\tau(r) \approx \int_{\mathscr{M}(X)} \Phi_p \mathrm{d}\mu. \tag{6-3}$$

For a given positive integer j, let $B(K_\ell, 1/j)$ be the set of all points $x \in X$ with $d(x, K_\ell) < 1/j$. Then by Eq. (6-2), we have for all positive integers j, ℓ, q,

$$\tau \otimes \tau\left(\left(G^{1/q}\right)^{-1}\left({}^*B(K_\ell, 1/j)\right)\right) > (1 - 1/\ell)^2. \tag{6-4}$$

By another appeal to the Permanence Principle, we can find, for each pair (j, ℓ) of positive integers, a hyperinteger $R_{j\ell} \in {}^*\mathbb{N}_\infty$ such that for any q with $1 \leq q \leq R_{j\ell}$, Eq. (6-4) still holds. Choose u_2 in ${}^*\mathbb{N}_\infty$ such that $u_2 \leq R_{j\ell}$ for all j, ℓ. Hence, Eq. (6-4) holds for any positive integers j, ℓ and for any q with $1 \leq q \leq u_2$. By a similar proof, there is an u_3 in ${}^*\mathbb{N}_\infty$ such that for any positive integer ℓ and for any hyperinteger q with $1 \leq q \leq u_2$,

$$\tau \otimes \tau\left(\left(G^{1/q}\right)^{-1}\left({}^*B\left(K_\ell, \frac{1}{u_3}\right)\right)\right) > 1 - \frac{1}{\ell}. \tag{6-5}$$

Choose $u = \min\{u_1, u_2\}$; inequality Eq. (6-5) implies that

$$L(\tau \otimes \tau)\left(\left(G^{1/u}\right)^{-1}(ns\,{}^*X)\right) = 1,$$

where $\mathrm{ns}\,{}^*X$ is the set of all near-standard points in ${}^*X$. Hence, there is a measurable function g from $(S \times S, L(\mathscr{S} \otimes \mathscr{S}))$ to $(X, \mathscr{B}(X))$ such that $G^{1/u}$ is an internal lifting of g. The Fubini theorem implies that for $L(\tau)$-almost all $r \in S$, $(G^{1/u})_r$ is an internal lifting of g_r, and hence ${}^*\rho(\tau(G^{1/u})_r^{-1}, L(\tau) g_r^{-1}) \simeq 0$ by the same argument used in the second paragraph. Thus, for $L(\tau)$-almost all $r \in S$, ${}^*\Phi_p(\tau(G^{1/u})_r^{-1})$ is infinitely close to $\Phi_p(L(\tau)g_r^{-1})$, and hence by Eq. (6-3)

$$\int_{\mathscr{M}(X)} \Phi_p \mathrm{d}\mu \simeq \int_S {}^*\Phi_p\left(\tau\left(G^{1/u}\right)_r^{-1}\right)\mathrm{d}\tau(r) \simeq \int_S \Phi_p\left(L(\tau)\, g_r^{-1}\right)\mathrm{d}L(\tau)(r).$$

Therefore,

$$\int_S \Phi_p\left(L(\tau)\, g_r^{-1}\right)\mathrm{d}L(\tau)(r) = \int_{\mathscr{M}(X)} \Phi_p \mathrm{d}\mu.$$

By Theorem 6.1 of Parthasarathy (1967) (p. 40), the hyper distribution of g is μ.

For each $1 \le r \le n$, define a mapping π_r from Ω to S by letting $\pi_r(\omega) = i_r$ if $\omega \in B_{i_1 \dots i_r \dots i_n}$. Then for any internal set $C \subseteq S$,

$$|P\left(\pi_r^{-1}(C)\right) - \tau(C)|$$
$$\le \sum_{i_1=1}^{n} \cdots \sum_{i_r \in C} \cdots \sum_{i_n=1}^{n} |P\left(B_{i_1 \dots i_r \dots i_n}\right) - \frac{1}{n^n}| \le \frac{n^n \tau(C)}{n^{2n}} \le \frac{1}{n^n}.$$

Thus, π_r is Loeb measure preserving from $(\Omega, L(\mathscr{A}), L(P))$ to $(S, L(\mathscr{S}), L(\tau))$. Similarly, we can define a mapping π^0 from T to S by letting $\pi^0(t) = r$ if $t \in A_r$. Then π^0 is also Loeb measure preserving. Let Π be the internal mapping from $T \times \Omega$ to $S \times S$ such that for $(t, \omega) \in T \times \Omega, \Pi(t, \omega) = (\pi^0(t), \pi_{\pi^0(t)}(\omega))$, i.e., $\Pi(t, \omega) = (r, \pi_r(\omega))$ if $t \in A_r$. Then II is Loeb measure preserving. To see this, let C be an internal subset of $S \times S$ and C_r be the set $\{s \in S: (r,s) \in C\}$. We have

$$|\lambda \otimes P\left(\Pi^{-1}(C)\right) - \tau \otimes \tau(C)|$$
$$= \left| \sum_{r=1}^{n} \lambda \otimes P\left(\pi^{-1}(\{r\} \times C_r)\right) - \sum_{r=1}^{n} \frac{\tau(C_r)}{n} \right|$$
$$= \left| \sum_{r=1}^{n} \left[\lambda(A_r) P\left(\pi_r^{-1}(c_r)\right) - \frac{\tau(C_r)}{n} \right] \right|$$
$$\le \sum_{r=1}^{n} \left| \lambda(A_r) - \frac{1}{n} \right| P\left(\pi_r^{-1}(C_r)\right) + \sum_{r=1}^{n} \frac{1}{n} |P\left(\pi_r^{-1}(C_r)\right) - \tau(C_r)|$$
$$\le n \cdot \frac{1}{n^2} + n \cdot \frac{1}{n} \cdot \frac{1}{n^n}.$$

Hence, Π is Loeb measure preserving. Define an internal process F on $T \times \Omega$ by letting $F(t, \omega) = G(\Pi(t, \omega))$. Let $f = g \circ \prod$. By the fact that the inverse image of a null set under Π is still null, we obtain that F is an internal lifting of f.

By the definition of f, we know that for any $t \in A_r$, $L(P) f_t^{-1} = L(\tau) g_r^{-1}$. Let Ψ_1: $T \to \mathscr{M}(X)$ and Ψ_2: $S \to \mathscr{M}(X)$ be defined as follows: $\Psi_1(t) = L(P) f_t^{-1}$ and $\Psi_2(r) = L(\tau) g_r^{-1}$. Then $\Psi_1 = \Psi_2 \circ \pi^0$. Since the distribution of Ψ_2 is μ and π^0 is measure preserving, the distribution of Ψ_1 must also be μ. Therefore, the hyper distribution of f is μ.

By modifying the values of g on a null set if necessary, we can assume that g_r is Loeb measurable for each $r \in S$. Thus, all the f_t are Loeb measurable. We shall now show that for any given $t_1 \in A_m$, $t_2 \in A_r$ for some $1 \leq m \neq r \leq n$, f_{t_1} and f_{t_2} are independent. Note that $f_{t_1} = g_m \circ \pi_m$ and $f_{t_2} = g_r \circ \pi_r$. For any internal sets $C_1, C_2 \subseteq S$, it can be checked that

$$P\left(\pi_m^{-1}(C_1) \cap \pi_r^{-1}(C_2)\right) \simeq P\left(\pi_m^{-1}(C_1)\right) \cdot P\left(\pi_r^{-1}(C_2)\right).$$

Thus, for any Loeb measurable sets $D_1, D_2 \subseteq S$

$$L(P)\left(\pi_m^{-1}(D_1) \cap \pi_r^{-1}(D_2)\right) = L(P)\left(\pi_m^{-1}(D_1)\right) \cdot L(P)\left(\pi_r^{-1}(D_2)\right).$$

Therefore, for any Borel sets $B_1, B_2 \subseteq X$,

$$\begin{aligned} L(P)\left(f_{t_1}^{-1}(B_1) \cap f_{t_2}^{-1}(B_2)\right) &= L(P)\left(\pi_m^{-1}\left(g_m^{-1}(B_1)\right) \cap \pi_r^{-1}\left(g_r^{-1}(B_2)\right)\right) \\ &= L(P)\left(\pi_m^{-1}\left(g_m^{-1}(B_1)\right)\right) \cdot L(P)\left(\pi_r^{-1}\left(g_r^{-1}(B_2)\right)\right) \\ &= L(P)\left(f_{t_1}^{-1}(B_1)\right) \cdot L(P)\left(f_{t_2}^{-1}(B_2)\right). \end{aligned}$$

Whence f_{t_1} and f_{t_2} are independent. Therefore, the set of all the pairs (t_1, t_2) such that f_{t_1} and f_{t_2} are not independent is contained in the set $\cup_{r=1}^{n} A_r \times A_r$. Since $\lambda(A_r) \leq (\frac{1}{n} + \frac{1}{n^2})$, it is obvious that

$$(\lambda \otimes \lambda)\left(\cup_{r=1}^{n} A_r \times A_r\right) \leq n\left(\frac{1}{n} + \frac{1}{n^2}\right)^2 \simeq 0,$$

and hence, the proof is complete. □

If we restrict our attention to processes with a hyperfinite Loeb counting parameter space, then we can characterize those having the same hyper distributions.

Proposition 6.3. *Let f and g be two processes from $(T \times \Omega_1, L(\mathcal{T} \otimes \mathcal{A}_1), L(\lambda \otimes P_1))$ and $(T \times \Omega_2, L(\mathcal{T} \otimes \mathcal{A}_2), L(\lambda \otimes P_2))$ respectively to some separable metric space X. Assume that $(T, \mathcal{T}, \lambda)$ is the internal counting probability space on T, and P_1, P_2 are atomless. Then f and g have the same hyper distribution if and only if there exists an internal permutation π of T, a hyperfinite Loeb counting space $(\Omega, L(\mathcal{A}), L(P))$, and a process h on $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$ with the following properties: for almost all (t, ω),*

$$f(t,\omega) = h(t, \alpha(t,\omega)) \text{ and } g(t,\omega) = h(\pi(t), \beta(t,\omega)),$$

where α and β are internal mappings from $T \times \Omega_1$ and $T \times \Omega_2$ respectively to Ω, and for each fixed $t \in T$, α_t and β_t are Loeb measure preserving from $(\Omega_1, L(\mathcal{A}_1), L(P_1))$ and $(\Omega_2, L(\mathcal{A}_2), L(P_2))$ respectively to $(\Omega, L(\mathcal{A}), L(P))$.

Proof: The 'if' part is clear. We only consider the 'only if' part. As in the proof of Corollary 5.5, Proposition 5.3 implies that there exists a Loeb counting space $(\Omega, L(\mathscr{A}), L(P))$, and there are processes h^1 and h^2 from $(T\times\Omega, L(\mathscr{T}\otimes\mathscr{A}), L(\lambda\otimes P))$ respectively to X, and internal mappings α^1 and β^1 from $T\times\Omega_1$ and $T\times\Omega_2$ respectively to Ω such that for $L(\lambda)$-almost all $t\in T$, $f_t=h_t^1\circ\alpha_t^1$ and $g_t=h_t^2\circ\beta_t^1$, and also for each fixed $t\in T$, α_t^1 and β_t^1 are Loeb measure preserving from $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ and $(\Omega_2, L(\mathscr{A}_2), L(P_2))$ respectively to $(\Omega, L(\mathscr{A}), L(P))$.

Define Ψ_1 and Ψ_2 to be the mappings from T to $\mathscr{M}(X)$ by letting $\Psi_1(t)=L(P)(h_t^1)^{-1}$ and $\Psi_2(t)=L(P)(h_t^2)^{-1}$. Then we know that Ψ_1 and Ψ_2 are Loeb measurable and have the same distribution. By the homogeneity property of hyperfinite Loeb counting spaces, there is an internal permutation π of T such that $\Psi_2=\Psi_1\circ\pi$. Thus, for $L(\lambda)$-almost all $t\in T$, h_t^2 and $h_{\pi(t)}^1$ have the same distribution on X. By the parametric homogeneity of hyperfinite Loeb counting spaces in Proposition 5.7, there is an internal process ϕ from $T\times\Omega$ to Ω such that for each $t\in T$, ϕ_t is a permutation of Ω, and for $L(\lambda)$-almost all $t\in T$, $h_t^2=h_{\pi(t)}^1\circ\phi_t$.

Let $h=h^1$, $\alpha=\alpha_1$, and define β by letting $\beta_t=\phi_t\circ\beta_t^1$. Then, for $L(\lambda)$-almost all $t\in T$, $f_t=h_t\circ\alpha_t$ and $g_t=h_{\pi(t)}\circ\beta_t$. The rest is clear. □

6.3. Measurability and the law

The following proposition shows that a process satisfying the consistency law is always not measurable with respect to $\mathscr{U}$ except for the trivial case that the random variables in the process are constants.

Proposition 6.4. *Let f be a real-valued integrable process on a Loeb product space $(T\times\Omega, L(\mathscr{T}\otimes\mathscr{A}), L(\lambda\otimes P))$. If f satisfies the consistency law, then f is not measurable with respect to the completion of $L(\mathscr{T})\otimes L(\mathscr{A})$ unless for $L(\lambda)$-almost all $t\in T$, f_t is a constant function.*

Proof: Suppose that f is measurable with respect to the completion of $L(\mathscr{T})\otimes L(\mathscr{A})$. Then, by Theorem 3.17, $f=E(f|_{\mathscr{U}})=h$ for some integrable function on T. Thus for $L(\lambda)$-almost all $t\in T$, f_t is the function taking the constant value $h(t)=Ef_t$, and we are done. □

When a process on a Loeb product space takes values in a general separable metric space, we cannot use the properties of conditional expectations as above. However, we can still formulate the following measurability result, which says that for a metric space valued process on a Loeb product space, if the process has almost surely pairwise independent random variables, then it is not measurable with respect to $\mathscr{U}$ except the trivial case that the random variables in the process are constants.

Proposition 6.5. *Let f be a process from $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ to a separable metric space X. Assume that the random variables f_t are almost surely pairwise independent. If f is measurable with respect to the completion of $L(\mathscr{T}) \otimes L(\mathscr{A})$, then for $L(\lambda)$-almost all $t \in T$, f_t is a constant mapping. It also means that the hyper distribution of the process is supported on the set of Dirac measures on X.*

Proof: Choose a countable open base $\{O_n\}_{n=1}^{\infty}$ in X, and let ϕ_n be the indicator function χ_{O_n} of O_n. For each $n \geq 1$, consider the real-valued process $\phi_n \circ f$. The almost sure pairwise independence condition on f implies that the random variables $(\phi_n \circ f)_t$ are almost surely uncorrelated. By Theorem 3.21, $E(\phi_n \circ f|_{\mathscr{U}}) = h_n$, where $h_n(t) = E(\phi_n \circ f)_t$. Since f is assumed to be measurable with respect to the completion of $L(\mathscr{T}) \otimes L(\mathscr{A})$, for each $n \geq 1$, $\phi_n \circ f$ must be so. We can find $A_n \in L(\mathscr{T})$ with $L(\lambda)(A_n) = 1$ such that $\phi_n \circ f(t, \cdot) = h_n(t)$ for all $t \in A_n$. Let A be the intersection of all the A_n. Then $L(\lambda)(A) = 1$, and for any $t \in A$, $\phi_n \circ f_t(\cdot) = h_n(t)$. Now, fix an arbitrary $t \in A$. We can choose a set $B \in L(\mathscr{A})$ such that $L(P)(B) = 1$ and $\phi_n \circ f_t(\omega) = h_n(t)$ for any $\omega \in B$ and for any $n \geq 1$. Hence for any $\omega_1, \omega_2 \in B$, we must have $f_t(\omega_1) = f_t(\omega_2)$, since the functions ϕ_n separate points. Therefore, the random variables f_t are constant mappings almost surely, and the proof is thus complete. □

The next proposition shows that the measurability with respect to the Loeb space of the internal product space but not with respect to the product of the two Loeb spaces is a universal phenomenon whenever non-purely atomic Loeb measures are concerned.

Proposition 6.6. *The completion of the product Loeb algebra $L(\mathscr{T}) \otimes L(\mathscr{A})$ is strictly contained in the Loeb product algebra $L(\mathscr{T} \otimes \mathscr{A})$ if and only if both $L(\lambda)$ and $L(P)$ are not purely atomic.*

Proof: We first consider the case that $L(\lambda)$ is purely atomic and let $\{C_n\}_{n=1}^{\infty}$ be all the atoms of $L(\lambda)$. It can be checked that for any $D \in L(\mathscr{T} \otimes \mathscr{A})$, there is an $L(\lambda)$-null set $E \in L(\mathscr{T})$ such that $D \cup (E \times \Omega) = \cup_{i=1}^{\infty} A_i \times B_i$, where A_i is one of the atoms C_n and B_i is in $L(\mathscr{A})$. Hence, D is measurable with respect to the $L(\lambda) \otimes L(P)$-completion of the product Loeb algebra $L(\mathscr{T}) \otimes L(\mathscr{A})$. It implies that the completion of the product Loeb algebra is equal to the Loeb product algebra. The same identity holds, when $L(P)$ is purely atomic.

Next, assume that both $L(\lambda)$ and $L(P)$ are not purely atomic. We can find internal sets T_0 and Ω_0 such that the respective restrictions of $L(\lambda)$ to T_0 and $L(P)$ to Ω_0 constitute the respective non-atomic parts of $L(\lambda)$ and $L(P)$. Let λ_0 and P_0 be the internal probability measures normalized from the restrictions of λ to T_0 and P to Ω_0 respectively. By Theorem 6.2, we can find a process g from the Loeb product space $(T_0 \times \Omega_0, L(\mathscr{T}_0 \otimes \mathscr{A}_0), L(\lambda_0 \otimes P_0))$ to $\{0, 1\}$ such that the random variables g_t are almost surely pairwise independent and identically distributed with equal distribution on 0 and 1, where $\mathscr{T}_0$ and $\mathscr{A}_0$ are the respective restrictions of $\mathscr{T}$ to T_0 and $\mathscr{A}$ to Ω_0. By Proposition 6.5, g is not measurable with respect to the completion of $L(\mathscr{T}_0) \otimes L(\mathscr{A}_0)$, which is the restriction of the completion of $L(\mathscr{T}) \otimes L(\mathscr{A})$ to $T_0 \times \Omega_0$. It is clear that g is the indicator function of some subset C of $T_0 \times \Omega_0$. Hence, C is in $L(\mathscr{T} \otimes \mathscr{A})$ but not in the $L(\lambda) \otimes L(P)$-completion of $L(\mathscr{T}) \otimes L(\mathscr{A})$, and we are done. □

6.4. The non-separability of $\boldsymbol{V}_{\mathscr{U}}^p$

As noted in Remark 3.22, all $\mathscr{L}^p$-integrable centered processes satisfying the consistency law are in the space $\boldsymbol{V}_{\mathscr{U}}^p$, where $\boldsymbol{V}_{\mathscr{U}}^p$ is the collection of those processes whose conditional expectations with respect to the relevant product Loeb space are zero. Recall that a real-valued process f is said to be centered if the means of the random variables f_t are almost surely zero. Now we shall show that the vector space $\boldsymbol{V}_{\mathscr{U}}^p$ is a big space for $p \geq 1$ in the sense that it is non-separable in the $\mathscr{L}^p$-norm.

Proposition 6.7. *If both $L(\lambda)$ and $L(P)$ are not purely atomic, then for any $1 \leq p \leq \infty$, $\boldsymbol{V}_{\mathscr{U}}^p$ is a non-separable subspace of the Banach space $\mathscr{L}^p(L(\lambda \otimes P))$ with the $\mathscr{L}^p$-norm.*

Proof: We only consider the case when both $L(\lambda)$ and $L(P)$ are atomless. By the same idea used in the first paragraph of the proof of Theorem 6.2, we can find some $H \in {}^*\mathbb{N}_\infty$ and an internal partition $\{B_{i_1 \ldots i_H} : i_1, \ldots, i_H = -1, 1\}$ of Ω with line $P(B_{i_1 \ldots i_H}) - 1/2^H| < 1/2^{2H}$ for each $i_1, \ldots, i_H = -1, 1$. Choose an $N \in {}^*\mathbb{N}_\infty$ such that $2N^2 < H$. Let $\{A_1, \ldots, A_N\}$ be an internal partition of T with $|\lambda(A_i) - 1/N| < 1/N^2$. For any $1 \leq n \leq H$, define ϕ_n on Ω by letting $\phi_n(\omega) = i_n$ if $\omega \in B_{i_1 \ldots i_n \ldots i_H}$. Then it is easy to check that $\int_\Omega |\phi_n - \phi_m| \mathrm{d}P \simeq 1$ for any $1 \leq n < m \leq H$. For any $0 \leq k \leq N$, let $f^k(t, \omega) = \phi_{kN+n}(\omega)$ if $t \in A_n$. It can be checked that for a fixed $0 \leq k \leq N$, the centered process f^k has almost surely pairwise independent random variables, and hence, $E(f^k|_{\mathscr{U}}) = 0$. Now for $0 \leq k < m \leq N$,

$$\int\int_{T \times \Omega} |f^k - f^m| \mathrm{d}\lambda \otimes P = \sum_{n=1}^{N} \int_{A_n} \int_\Omega |\phi_{kN+n} - \phi_{mN+n}| \mathrm{d}P \mathrm{d}\lambda \simeq 1.$$

Hence, $\int\int_{T \times \Omega} |f^k - f^m| \mathrm{d}L(\lambda \otimes P) = 1$. Note that for any $x \in \mathscr{L}^\infty(L(\lambda \otimes P))$, the $\mathscr{L}^1$-norm of x is always less than or equal to the $\mathscr{L}^p$-norm of x for any $1 \leq p \leq \infty$. Since all the f^k are bounded, we thus know that the pairwise distance of the processes f^k in the $\mathscr{L}^p$-norm is greater than or equal to 1 for any $1 \leq p \leq \infty$. Since the external cardinality of a hyperfinite set is at least the continuum, we thus know that the Banach spaces $\boldsymbol{V}_{\mathscr{U}}^p$s are non-separable. □

7. Unification of multiplicative properties

7.1. Introduction

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 (see, for example, Loéve, 1977a). 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 aim of the following sections is to attempt to answer the question. In particular, we shall provide some intrinsic characterizations of almost sure pairwise independence. These characterizations allow us to unify various multiplicative properties in the probabilistic literature. It also shows that many seemingly different descriptions of individual uncertainty are, in fact, all equivalent.

Section 7.2 contains two examples showing the differences among several notions which are weaker than independence. In Section 7.3, we formalize the probabilistic observation ''No betting system can beat the house'' in the setting of empirical distributions. The main theorem of the section, Theorem 7.6, says that almost sure pairwise independence is necessary and sufficient for the satisfiability of the consistency law in distribution. Based on separating classes of Borel functions, we also present in the theorem two other general characterizations of almost sure pairwise independence. Section 7.4 concerns with real-valued processes. We derive from Theorem 7.6 some corollaries which unify multiplicative properties involving the maximum of random variables, characteristic functions, the method of moments, generating functions, and a condition based on the real exponential family. Though these multiplicative properties are all strictly weaker than independence for a fixed finite number of random variables, it is surprising that their almost sure versions are all equivalent to almost sure pairwise independence. The duality result in Proposition 7.16 shows a perfect symmetry in the sense that the random variables in a process are almost surely pairwise independent as well as identically distributed almost surely if and only if so are the sample functions. In Section 7.5, we present some additional characterizations of almost sure pairwise independence for metric space valued processes. Some of them are simple observations and are listed for the purpose of providing easy references at other places.

7.2. Two examples

Definition 3.9 presents a different notion of independence, the so-called diagonal independence, for real-valued random variables. For general metric spaces, we have no way to define a sort of diagonal independence. However, another kind of independence, called semi-independence, can still be defined. Two random variables φ and ψ from a probability space $(\Sigma, \mathscr{F}, \mu)$ to a metric space X are said to be semi-independent [12] if for any open set O in X, the events $\varphi^{-1}(O)$ and $\psi^{-1}(O)$ are independent, i.e.,

$$\mu(\varphi^{-1}(O) \cap \psi^{-1}(O)) = \mu(\varphi^{-1}(O))\mu(\psi^{-1}(O)).$$

It is shown in Sun (1995) that semi-independence preserves some sort of multiplicative properties. In fact, a primary reason to introduce the notion of semi-independence in Sun (1995) is to study its relevance with multiplicative properties of random variables and the related central limit theorem.

The following example shows that diagonal independence is weaker than semi-independence, and it can also be neither Γ-uncorrelated nor uncorrelated.

Example 7.1. Define a probability measure μ on the set $\{-1, 0, 1\}^2$ such that

$$\mu(\{(-1,1)\}) = 2/9, \mu(\{(0,1)\}) = 0, \mu(\{(1,1)\}) = 1/9,$$

$$\mu(\{(-1,0)\}) = 0, \mu(\{(0,0)\}) = 1/3, \mu(\{(1,0)\}) = 0,$$

$$\mu(\{(-1,-1)\}) = 1/9, \mu(\{(0,-1)\}) = 0, \mu(\{(1,-1)\}) = 2/9.$$

Let ϕ, ψ be any random variables taking values from $\{-1, 0, 1\}$ such that the joint distribution of (ϕ, ψ) is μ. Then it is easy to check that ϕ and ψ are diagonally independent but neither semi-independent, nor Γ-uncorrelated, nor uncorrelated.

The next example adopted from Sun (1995) shows that there are semi-independent random variables which are neither independent nor Γ-uncorrelated.

Example 7.2. Define a probability measure μ on the set $\{1, 2, 3\}^2$ such that

$$\mu(\{(1,1)\}) = \mu(\{(2,2)\}) = \mu(\{(3,3)\}) = 1/9$$

$$\mu(\{(1,3)\}) = \mu(\{(2,1)\}) = \mu(\{(3,2)\}) = 2/9$$

$$\mu(\{(1,2)\}) = \mu(\{(2,3)\}) = \mu(\{(3,1)\}) = 0$$

Let ϕ, ψ be any random variables taking values from $\{1, 2, 3\}$ such that the joint distribution of (ϕ, ψ) is μ. Then it is easy to check that ϕ and ψ are semi-independent but not independent.

[12] It is easy to see that the proof of Theorem 5.2 only uses the fact that the random variables f_t are semi-independent.

Now since ϕ and ψ have the same distribution, we obtain $\Gamma_{\phi} \cdot \Gamma_{-\psi} = |\Gamma_{\phi}|^2$, which is always non-negative. On the other hand, it is easy to see that

$$\Gamma_{\phi-\psi}(u) = \frac{1}{3} + \frac{2}{9}(2e^{iu} + e^{-2iu}).$$

Thus, $\Gamma_{\phi} \cdot \Gamma_{-\psi} \neq \Gamma_{\phi-\psi}$, and hence ϕ and ψ are not Γ-uncorrelated.

7.3. The consistency law in distribution

As in Definition 3.15, we can introduce the following version of the consistent law of large numbers. Again, it is a sort of idealization of the phenomenon ''No betting system can beat the house''. This concept is used below to characterize those processes satisfying the law in distribution.

Definition 7.3. Let f be a process from a Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ 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 $A \in \mathscr{T}$ with $L(\lambda)(A) > 0$, the process f^A from the reduced Loeb product space $(A \times \Omega, L(\mathscr{T}^A \otimes \mathscr{A}), L(\lambda^A \otimes P))$ 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 $(A, L(\mathscr{T}^A), L(\lambda^A))$ to X is equal to the distribution μ^A induced by the random variable f^A from $(A \times \Omega, L(\mathscr{T}^A \otimes \mathscr{A}), L(\lambda^A \otimes P))$ to X. Here, f^A is the restriction of f to $A \times \Omega$, $\mathscr{T}^A$ is the collection of all internal subsets of A in $\mathscr{T}$, and λ^A is the internal probability measure on $(A, \mathscr{T}^A)$ rescaled from λ.

As noted in Lemma 3.16, we can only consider in the above definition those A with $L(\lambda)(A) > 1 - \varepsilon$ for any prior given number $\varepsilon \in (0, 1)$ (no matter how small ε is). This relaxation will prove to be convenient in the proof of Proposition 9.4. The following example shows that there are processes which satisfy the law in distribution but not the consistency law in distribution.

Example 7.4. Let f be the process on a product Loeb space $(T \times \Omega, L(\mathscr{T}) \otimes L(\mathscr{A}), L(\lambda) \otimes L(P))$ as defined in Example 3.18. For any $\omega \in \Omega$, f_{ω} is equally distributed on -1 and 1. Similarly, as a random variable on the Loeb product space, f is also equally distributed on -1 and 1. Hence, f satisfies the law in distribution. Suppose f satisfies the consistency law in distribution. Since f is also integrable, f must satisfy the consistency law. This contradicts the result in Example 3.18. Therefore, f does not satisfy the consistency law in distribution.

The main theorem of this section 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 7.5. Let $\mathscr{D}$ be a collection of Borel probability measures on a separable metric space X and let $\mathscr{E}$ be a class of real or complex valued Borel functions on X such that for any given pair $(\phi, \mu) \in \mathscr{E} \times \mathscr{D}$, ϕ is μ-integrable. The class $\mathscr{E}$ is said to be separating for $\mathscr{D}$, if there is a sequence $\{\phi_n\}_{n=1}^{\infty}$ of functions $\mathscr{E}$ which distinguishes the members of $\mathscr{D}$. That is, μ and ν in $\mathscr{D}$ are equal if $\int_X \phi_n \mathrm{d}\mu = \int_X \phi_n \mathrm{d}\nu$ for all $n \geq 1$.

We shall now move to the main theorem of this section. Since almost sure pairwise independence is shown to be necessary and sufficient for the satisfiability of the consistency law in distribution, the condition is thus intrinsic pertaining to the study of the law of large numbers. General characterizations for almost sure pairwise independence are also presented in terms of suitable separating classes. As before, the equivalence of (3) and (4) presents a sort of duality between the random variables and the sample functions.

Theorem 7.6. *Let $\mathscr{D}$ be a collection of distributions on a separable metric space X, $\mathscr{E}$ a separating class for $\mathscr{D}$ of real or complex valued Borel functions, and f a process from a Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ to X such that the distributions of all the sample functions f_ω^A as well as the distribution of the process f^A viewed as a random variable are all in $\mathscr{D}$, where $A \in \mathscr{T}$ with $L(\lambda)(A) > 0$, and f^A is the restriction of f to $A \times \Omega$ with the rescaled Loeb product measure. Assume that $\varphi(f)$ is a square integrable process for each $\varphi \in \mathscr{E}$. 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 $\varphi \in \mathscr{E}$, the $(\varphi(f))_t$ are almost surely uncorrelated, i.e., for $L(\lambda \otimes \lambda)$-almost all $(t_1,t_2) \in T \times T, E((\varphi(f))_{t_1}\overline{(\varphi(f))_{t_2}}) = E(\varphi(f))_{t_1} \times E\overline{(\varphi(f))_{t_2}}$;

(4) for each $\varphi \in \mathscr{E}$, $E(\varphi(f_{\omega_1})\overline{\varphi(f_{\omega_2})}) = E|E\varphi(f_t)|^2$ for $L(P \otimes P)$-almost all $(\omega_1, \omega_2) \in \Omega \times \Omega$.

Proof: For $(1) \Rightarrow (2)$, choose a countable open base $\mathscr{B}$ for X such that $\mathscr{B}$ is closed under the formation of finite intersections. For a fixed open set $O \in \mathscr{B}$, let ψ denote the indicator function of O. Then we observe that $L(\lambda \otimes P)(\psi(f))^{-1} = (L(\lambda \otimes P)f^{-1})\psi^{-1}$, and for any $\omega \in \Omega$, $L(\lambda)\ ((\psi(f))_\omega)^{-1} = (L(\lambda)(f_\omega)^{-1})\psi^{-1}$. If we start with an internal set A with $L(\lambda)(A) > 0$, we have similar identities for the subprocesses f^A and $(\psi(f))^A$. Hence, the process $\psi(f)$ also satisfies the consistency law in distribution. Thus, $\psi(f)$ satisfies the consistency law. By Theorem 4.6, $\psi(f)$ has almost surely uncorrelated random variables.

Next pick any O_1, $O_2 \in \mathscr{B}$. Let χ_{O_1} and χ_{O_2} be the respective indicator functions of O_1 and O_2. The previous paragraph shows that the random variables

$(\chi_{O_1}(f))_t$ are almost surely uncorrelated. By Theorem 4.6 (7), the processes $\chi_{O_1}(f)$ and $\chi_{O_2}(f)$ themselves are almost surely uncorrelated. Whence

$$\begin{aligned}&\int_\Omega \left(\chi_{O_1}(f)\right)_{t_1} \cdot \left(\chi_{O_2}(f)\right)_{t_2} \mathrm{d}L(P)\\ &\quad= \int_\Omega \left(\chi_{O_1}(f)\right)_{t_1} \mathrm{d}L(P) \int_\Omega \left(\chi_{O_2}(f)\right)_{t_2} \mathrm{d}L(P)\end{aligned}$$

for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$. Since $\mathscr{B}$ is countable, we can choose an $L(\lambda \otimes \lambda)$-null subset A_0 of $T \times T$ such that for all $(t_1,t_2) \notin A_0$,

$$L(P)\left(f_{t_1}^{-1}(O_1) \cap f_{t_2}^{-1}(O_2)\right) = L(P)\left(f_{t_1}^{-1}(O_1)\right) \cdot L(P)\left(f_{t_2}^{-1}(O_2)\right)$$

for all $O_1, O_2 \in \mathscr{B}$. Since $\mathscr{B}$ is closed under the formation of finite intersections, i.e., a π-system, Dynkin's $\pi - \lambda$ theorem (see Cohn, 1980, p. 44 and Durrett, 1991, p. 404) implies that for all $(t_1, t_2) \notin A_0$,

$$L(P)\left(f_{t_1}^{-1}(B_1) \cap f_{t_2}^{-1}(B_2)\right) = L(P)\left(f_{t_1}^{-1}(B_1)\right) \cdot L(P)\left(f_{t_2}^{-1}(B_2)\right)$$

holds for all Borel subsets B_1, B_2 in X. This means that f_{t_1} and f_{t_2} are independent for all $(t_1, t_2) \notin A_0$, and hence (2) follows.

For (2) $\Rightarrow$ (3), simply note that for each $\varphi \in \mathscr{E}$, $\varphi(f)$ is a square integrable process with almost surely pairwise independent random variables. Hence, the random variables $(\varphi(f))_t$ are almost surely uncorrelated, i.e., (3) holds.

We shall now show (3) $\Rightarrow$ (1). For each fixed $\varphi \in \mathscr{E}$, define a centered process g by letting $g(t, \omega) = \varphi(f)(t, \omega) - E(\varphi(f))_t$. Since it is assumed in (3) that the real or complex valued random variables $(\varphi(f))_t$ are almost surely uncorrelated, we know that the random variables g_t are almost surely orthogonal. As noted earlier, Theorem 3.7 is still valid for complex valued processes (see the proof of Theorem 3.13). Hence for $L(P)$-almost all $\omega \in \Omega$, $Eg_\omega = 0$, and thus $E\varphi(f_\omega) = E\varphi(f)$. Let μ_ω be the distribution on X induced by the sample function f_ω on T and μ the distribution induced by f viewed as a random variable on $T \times \Omega$. Then $\int_X \varphi \mathrm{d}\mu_\omega = \int_X \varphi \mathrm{d}\mu$ for $L(P)$-almost all $\omega \in \Omega$. Since $\mathscr{E}$ has a countable separating subclass $\mathscr{E}_0$ for $\mathscr{D}$, we can find an $L(P)$-null subset B of Ω such that for any fixed $\omega \notin B$, $\int_X \varphi \mathrm{d}\mu_\omega = \int_X \varphi \mathrm{d}\mu$ for all $\varphi \in \mathscr{E}_0$. Since μ_ω and μ are in $\mathscr{D}$, the fact that $\mathscr{E}_0$ is separating for $\mathscr{D}$ implies $\mu_\omega = \mu$ for all $\omega \notin B$. Therefore, f satisfies the law in distribution. The same argument can be used to show that f satisfies the consistency law in distribution. Hence, we obtain (3) $\Rightarrow$ (1).

It can be checked that the results in Theorem 4.6 are still valid for complex processes. Thus the equivalence of (3) and (4) follows. In fact, similar identities as in the proof of (2) $\Leftrightarrow$ (3) in Theorem 4.5 can be established for complex valued processes. These identities can, then, be used to prove (3) $\Leftrightarrow$ (4) here directly. □

7.4. Some specific multiplicative properties for real-valued processes

We have seen that independence is in general strictly stronger than diagonal independence and Γ-uncorrelatedness, and the latter two do not imply each other.

In Theorems 3.10 and 3.13, we present versions of the law in distribution by using the almost sure versions of pairwise diagonal independence and Γ-uncorrelatedness. Theorem 7.6 allows us to claim in the next two corollaries that the almost sure versions of both conditions are actually equivalent to almost sure pairwise independence and thus, stronger than almost sure uncorrelatedness. Following these two corollaries, multiplicative properties involving separating classes for some collections of special distributions on $\mathbb{R}$ are also considered.

We start with almost sure pairwise diagonal independence. Besides its equivalence with almost sure pairwise independence, it is also equivalent to the fact that the diagonals of the pairwise joint distribution functions of the sample functions are essentially independent of the choice of particular samples.

Corollary 7.7. *Let f be a real-valued process on a Loeb product space $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$. Then the following are equivalent:*

(1) the random variables f_t are almost surely pairwise independent;

(2) the random variables f_t are almost surely pairwise diagonally independent;

(3) let F_t be the distribution function of f_t and $F_{\omega_1\omega_2}$ the joint distribution function of the sample functions f_{ω_1} and f_{ω_2}; then the diagonal of the joint distribution function is in the following special form: $F_{\omega_1\omega_2}(x,x) = \int_T (F_t(x))^2 \mathrm{d}L(\lambda)$ for $L(P \otimes P)$-almost all $(\omega_1, \omega_2) \in \Omega \times \Omega$.

Proof: For the proof, we note that the class $\mathcal{E}_0$ of the indicator functions of the intervals $(-\infty, x]$ for all rational numbers x is a countable separating class for all distributions on $\mathbb{R}$. Hence, the class $\mathcal{E}$ of the indicator functions of the intervals $(-\infty, x]$ for all real numbers x is a separating class for all distributions on $\mathbb{R}$. If (2) is satisfied, then for any $\varphi \in \mathcal{E}$, the random variables $(\varphi(f))_t$ are almost surely uncorrelated, and hence, (1) follows from Theorem 7.6. Since $(1) \Rightarrow (2)$ is obvious, (1) and (2) are equivalent.

By rewriting (3), we obtain for $L(P \otimes P)$-almost all $(\omega_1, \omega_2) \in \Omega \times \Omega$, $E(\varphi(f_{\omega_1})\varphi(f_{\omega_2})) = E|E\varphi(f_t)|^2$ for all $\varphi \in \mathcal{E}$. Hence, $(3) \Rightarrow (1)$ is clear.

If we assume (1), then Theorem 7.6 implies that for each $\varphi \in \mathcal{E}_0$, $E(\varphi(f_{\omega_1})\varphi(f_{\omega_2})) = E|E\varphi(f_t)|^2$ holds for $L(P \otimes P)$-almost all $(\omega_1, \omega_2) \in \Omega \times \Omega$. Since $\mathcal{E}_0$ is countable, we can obtain that except for an $L(P \otimes P)$-null set, the previous identity holds for all $\varphi \in \mathcal{E}_0$. By taking limits, we can derive (3). □

Now we move to almost sure Γ-uncorrelatedness. We simply note that the complex exponentials e^{iux} form a separating class for distributions on $\mathbb{R}$. Thus, as in the previous corollary, it is easy to prove this corollary by applying Theorem 7.6. Here, we omit the proof.

Corollary 7.8. *Let f be a real-valued process on $(T \times \Omega, L(\mathcal{T} \otimes \mathcal{A}), L(\lambda \otimes P))$. Then the following are equivalent:*

(1) the random variables f_t are almost surely pairwise independent;

(2) the random variables f_t are almost surely Γ-uncorrelated.

In the next three corollaries, we consider separating classes for collections of special distributions on $\mathbb{R}$. The proofs will be omitted. We shall first work with the class of functions $\{x^k: k \in \mathbb{N}\}$. It is quite easy to show that this class is a separating class for distributions with a compact support. However, as noted in Durrett (1991) (p. 89), it is not a separating class for all distributions on $\mathbb{R}$. Let $\mathscr{M}^m$ be the class of distribution functions $F(x)$ such that

$$\overline{\lim}_{k\to\infty} \frac{|m_k|^{1/k}}{k} < \infty,$$

where m_k is the k-th moment of F, i.e., $m_k = \int_{\mathbb{R}} x^k \mathrm{d}F(x)$. Then $\{x^k: k \in \mathbb{N}\}$ is a separating class for $\mathscr{M}^m$ (see Breiman, 1968, p. 182). The relevant theory is called the 'method of moments'. The following corollary follows immediately from Theorem 7.6.

Corollary 7.9. *Let f be a real-valued process on $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ such that the distribution of f viewed as a random variable on $T \times \Omega$ is in $\mathscr{M}^m$.* [13] *Then the following are equivalent:*

(1) The random variables f_t are almost surely pairwise independent;

(2) For all $k \in \mathbb{N}$, the random variables $(f_t)^k$ are almost surely uncorrelated; this means that the joint moments of f_{t_1} and f_{t_2} are almost surely the product of the corresponding moments of f_{t_1} and f_{t_2}.

Let $\mathscr{M}^n$ be the collection of distribution functions of random variables taking values in natural numbers. Note that for a random variable ϕ taking values in $\mathbb{N}$, one can define the generating function $\mathscr{G}_\phi$ for ϕ by letting $\mathscr{G}_\phi(z) = \sum_{k=0}^{\infty} b_k z^k$ (see Chung, 1974 and Durrett, 1991), where b_k is the probability of ϕ taking the value k. The following corollary follows from the fact that the class of functions $\{z^x: z \in (-1,1)\}$ is separating for $\mathscr{M}^n$.

Corollary 7.10. *Let f be a natural number valued process on a Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Then the following are equivalent:*

(1) The random variables f_t are almost surely pairwise independent;

(2) The generating functions of $f_{t_1} + f_{t_2}$ are almost surely the product of the respective generating functions of f_{t_1} and f_{t_2}, i.e., $\mathscr{G}_{f_{t_1}+f_{t_2}} = \mathscr{G}_{f_{t_1}} \cdot \mathscr{G}_{f_{t_2}}$ for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$.

[13] Let $m_k(f)$ and $m_k(f_\omega)$ denote the k-th moments of f and f_ω respectively. Since $L(\lambda \otimes P)f^{-1} \in \mathscr{M}^m$, there is a positive number M such that $\sup_{k\geq 1}[m_k(f)]^{1/k}/k \leq M$. The following inequality can be established: $L(P)(\{\omega \in \Omega: \sup_{k\geq 1}[m_k(f_\omega)]^{1/k}/k \leq \alpha M\}) \geq (\alpha-2)/(\alpha-1)$ for any $\alpha > 1$. Thus for almost all ω, $L(\lambda)f_\omega^{-1} \in \mathscr{M}^m$. It is then easy to see that the distributions of f^A and f_ω^A are also in $\mathscr{M}^m$ for any $L(\lambda)(A) > 0$. Hence, the condition of Theorem 7.6 is satisfied.

Finally, we consider the collection $\mathscr{M}^+$ of the distribution functions of non-negative random variables. For a non-negative random variable ϕ, let ε_ϕ be the function on $[0, \infty)$ such that $E_\phi(v) = Ee^{-v\phi}$ for each $v \in [0,\infty)$. Since the real exponential family $\{e^{iv\chi}: v \geq 0\}$ is a separating class for $\mathscr{M}^+$ (see Breiman, 1968, p. 183), it is easy to derive the following corollary from Theorem 7.6.

Corollary 7.11. *Let f be a non-negative process on $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. Then the following are equivalent*:

(1) The random variables f_t are almost surely pairwise independent;

(2) For $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \in T \times T$, $\varepsilon_{f_{t_1} + f_{t_2}} = \varepsilon_{f_{t_1}} \cdot \varepsilon_{f_{t_2}}$.

7.5. Additional characterizations of almost sure pairwise independence

Theorem 7.6 can be used to provide further characterizations of almost sure pairwise independence for metric space valued processes. We shall first consider semi-independence. Example 7.2 shows that for real-valued random variables, semi-independence is in general strictly weaker than independence and also has no relevance with Γ-uncorrelatedness. Furthermore, for the random variables ϕ and ψ in Example 7.2, it can also be checked directly that $E\phi^k\psi^k = E\phi^k E\psi^k$, $\mathscr{G}_{\phi+\psi} = \mathscr{G}_\phi \mathscr{G}_\psi$, and $\varepsilon_{\phi+\psi} = \varepsilon_\phi \varepsilon_\psi$. Thus, as already noted in Sun (1995), those multiplicative properties are all weaker than independence; their almost sure versions are, however, already proven to be equivalent to almost sure pairwise independence in Section 7.4. The following corollary, which is an obvious consequence of Theorem 7.6 and the fact that the collection $\{\chi_O : O$ is an open set in $X\}$ is separating for a separable metric space X, shows that the almost sure versions of independence and semi-independence are equivalent.

Corollary 7.12. *Let f be a process from $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ to a separable metric space X. Then the random variables f_t are almost surely pairwise semi-independent if and only if they are almost surely pairwise independent*.

Other than the semi-independence condition, we list in the following proposition some more necessary and sufficient conditions for a metric space valued process to satisfy the consistency law in distribution. The implication (1) ⇒ (2) states that if a process satisfies the consistency law in distribution, then the consistency law in distribution also holds for any parameter invariant transformations of the process (see Papoulis, 1965, Section 9.5 for some examples). Note that a Borel space is a separable metric space homeomorphic to a Borel set of a Polish space. (3) repeats a version of (2) for the real case. Since there are many equivalent conditions for the real-valued processes $\varphi(f)$ to satisfy the consistency law (see Sections 3 and 4), one can include plenty of other equivalent conditions for (1) below, which will not be repeated here.

Proposition 7.13. *Let f be a process from $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ to a separable metric space X. The following are equivalent:*

(1) f satisfies the consistency law in distribution;

(2) Given a Borel space Y whose cardinality is at least that of X, the process $\Phi(f)$ satisfies the consistency law in distribution for any Borel measurable mapping Φ from X to Y;

(3) For any bounded real-valued Borel measurable function φ on X, $\varphi(f)$ satisfies the consistency law;

(4) For any bounded real-valued continuous function φ on X, the random variables $(\varphi \circ f)_t$ are almost surely uncorrelated.

Proof: For $(1) \Rightarrow (2)$, we simply observe that $L(\lambda \otimes P)(\Phi(f))^{-1} = (L(\lambda \otimes P)f^{-1})\Phi^{-1}$, and for any $\omega \in \Omega$, $L(\lambda)(\Phi(f))_{\omega}^{-1} = (L(\lambda)f_{\omega}^{-1})\Phi^{-1}$. If we start with an internal set A with $L(\lambda)(A) > 0$, we have similar identities for the subprocesses f^A and $(\Phi(f))^A$. The rest is clear. Note that for this implication, it is not necessary to assume the target space Y to be Borel.

To check $(2) \Rightarrow (1)$, we observe that for a given Borel space Y whose cardinality is at least that of X, the classical Borel isomorphism theorem in Parthasarathy (1967) (p. 14) implies that there is a Borel isomorphism Φ from X to a subset Z of Y. Let Ψ be the inverse mapping of Φ. By the assumption in (2), $\Phi(f)$ satisfies the consistency law in distribution. Since $\Psi(\Phi(f)) = f$, we obtain that the process f satisfies the consistency law in distribution, i.e., (1) follows.

The class of bounded real-valued Borel measurable functions $\varphi(f)$ on X is separating for all distributions on X. If $\varphi(f)$ satisfies the consistency law, then the random variables $\varphi(f_t)$ are almost surely uncorrelated by Theorem 4.6. Thus $(1) \Leftrightarrow (3)$ holds by Theorem 7.6.

If we choose a sequence of bounded continuous functions as in the proof of Lemma 6.1, then the sequence is a separating class for all distributions on X. Hence, $(1) \Leftrightarrow (4)$ follows from Theorem 7.6 easily. □

From Proposition 7.13 (3), we observe that if for any $A \in L(\mathscr{T})$ and for any bounded real-valued Borel function φ on X,

$$\int_A \varphi(f_\omega)(t) = \int_A \int_\Omega \varphi(f) \mathrm{d}L(P) \mathrm{d}L(\lambda)$$

holds for $L(P)$-almost all $\omega \in \Omega$, then f satisfies the consistency law in distribution, and hence the random variables f_t are almost surely pairwise independent. As noted in Corollary 7.12, instead of using any bounded real-valued Borel functions, we can only consider the class of indicator functions of open sets to derive the same result. When real-valued random variables are concerned, the class of indicator functions of the infinite intervals $(-\infty, x]$ is also separating. It is thus interesting to have a general condition on a class of Borel sets such that the relevant class of indicator functions is separating. The following proposition shows from another aspect that why the two mentioned classes of indicator functions can be used to present some versions of the law of large numbers in Sections 3 and 5 respectively.

Proposition 7.14. *Let $\mathscr{H}$ be a class of Borel subsets of a separable metric space X. Assume that $\mathscr{H}$ is closed under formation of finite intersections (a π-system) and also generates the Borel σ-algebra of X. Let f be a process from $(T\times\Omega, L(\mathscr{T}\otimes\mathscr{A}), L(\lambda\otimes P))$ to X. Then the following are equivalent*:

(1) The random variables f_t are almost surely pairwise independent;

(2) For any $H\in\mathscr{H}$, the events $f_t^{-1}(H)$ are almost surely pairwise independent;

(3) For any $A\in\mathscr{A}$ and $H\in\mathscr{H}$,

$$\int_T \chi_A\cdot\chi_H(f_\omega)\,\mathrm{d}L(\lambda)=\int_T\int_\Omega \chi_A\cdot\chi_H(f_\omega)\,\mathrm{d}L(\lambda)\,\mathrm{d}L(P)$$

for $L(P)$-almost all $\omega\in\Omega$.

Proof: (1) ⇒ (2) and (2) ⇒ (3) are clear.

We only need to show that (3) ⇒ (1). For a countable subclass $\mathscr{H}'$ of $\mathscr{H}$, let $\sigma(\mathscr{H}')$ be the σ-algebra generated by $\mathscr{H}'$. It is easy to see that

$$\bigcup\{\sigma(\mathscr{H}'):\mathscr{H}' \text{ is countable subclass of } \mathscr{H}\}$$

is the Borel σ-algebra. We can choose subclasses $\mathscr{H}'_n$ of $\mathscr{H}$, $n=1,2,\dots$, such that $\bigcup_{n=1}^{\infty}\sigma(H'_n)$ contains a countable open base of X. Let $\mathscr{H}'$ be the union of all the $\mathscr{H}'_n$ and $\mathscr{H}_0$ be the class of sets obtained by taking finite intersections of sets in $\mathscr{H}'$. Then $\mathscr{H}_0$ is a countable π-system which generates the Borel σ-algebra of X. It is clear that $\mathscr{H}_0\subseteq\mathscr{H}$. Hence, (3) implies that for $L(P)$-almost all $\omega\in\Omega$,

$$\int_T \chi_H(f_\omega)\,\mathrm{d}L(\lambda)=\int_T\int_\Omega \chi_H(f_\omega)\,\mathrm{d}L(\lambda)\,\mathrm{d}L(P)$$

for any $H\in\mathscr{H}_0$. By a result on the uniqueness of the extension of measures in Durrett (1991) (p. 402), we obtain for $L(P)$-almost all $\omega\in\Omega, L(P)f_\omega^{-1}=L(P)f^{-1}$. Similarly, we can show that f satisfies the consistency law in distribution. Hence the random variables f_t are almost surely pairwise independent. □

Except for independence, none of the probabilistic notions considered by us, such as uncorrelatedness, diagonal independence, Γ-uncorrelatedness, semi-independence, etc., are preserved under parameter dependent transformations. Thus, the equivalence of almost sure pairwise independence with the almost sure versions of these notions is very helpful. We list below some results for transformations of processes which are parameter dependent. Note that we do not have to require the processes obtained from the parameter dependent transformations to satisfy the consistency law (or the consistency law in distribution), since that can be deduced. As before, by the results in Sections 3 and 4, we can obtain many other equivalent conditions easily for the relevant real-valued processes in the following proposition.

Proposition 7.15. *Let f be a process from $(T \times \Omega,\ L(\mathcal{T} \otimes \mathcal{A}),\ L(\lambda \otimes P))$ to a separable metric space X. The following are equivalent:*

(1) The random variables f_t are almost surely pairwise independent;

(2) Given a Borel space Y whose cardinality is at least that of X, let $\{\Phi_t : t \in T\}$ any given class of Borel mappings from X to Y such that the process g defined by $g(t,\omega) = \Phi_t(f(t,\omega))$ is measurable on the Loeb product space; then the process g satisfies the law in distribution;

(3) For any given class $\{\varphi_t : t \in T\}$ of real-valued Borel measurable functions on X such that the process g defined by $g(t,\omega) = \varphi_t(f(t,\omega))$ is integrable on the Loeb product space, the process g satisfies the law;

(4) For any given class $\{\varphi_t : t \in T\}$ of real-valued uniformly bounded Borel measurable functions on X such that the process g defined by $g(t,\omega) = \varphi_t(f(t,\omega))$ is measurable on the Loeb product space, $\int_T \varphi_t(f_\omega(t))\mathrm{d}L(\lambda) = \int_T \int_\Omega \varphi_t(f)\mathrm{d}L(P)\mathrm{d}L(\lambda)$ for $L(P)$-almost all $\omega \in \Omega$;

(5) For any given class $\{\varphi_t : t \in T\}$ of uniformly bounded real-valued Borel measurable functions on X such that the process g defined by $g(t,\omega) = \varphi_t(f(t,\omega))$ is a centered measurable process on the Loeb product space, we have $\int_T \varphi_t(f_\omega(t))\mathrm{d}L(\lambda) = 0$ for $L(P)$-almost all $\omega \in \Omega$;

(6) Let φ be a bounded real-valued function on $T \times X$ such that for each $t \in T$, φ_t is continuous on X and for each $x \in X$, φ_x is Loeb measurable on $(T, L(\mathcal{T}), L(\lambda))$; then $\int_T \varphi_t(f_\omega(t))\mathrm{d}L(\lambda) = \int_T \int_\Omega \varphi_t(f)\mathrm{d}L(P)\mathrm{d}L(\lambda)$, for $L(P)$-almost all $\omega \in \Omega$.

Proof: The equivalence of (1), (2), (3), and (4) as well as (4) ⇒ (5) is clear.

Now we assume (5). Let $\{\varphi_t : t \in T\}$ be a given class of real-valued uniformly bounded Borel measurable functions on X such that the process g defined by $g(t,\omega) = \varphi_t(f(t,\omega))$ is measurable on the Loeb product space. Define a new class $\{\psi_t : t \in T\}$ of functions on X by letting $\psi_t = \varphi_t - \int_\Omega g_t \mathrm{d}L(P)$. Let h be the centered process defined by $h(t,\omega) = \psi_t(f(t,\omega))$. By (5),

$$\int_T \varphi_t(f_\omega(t))\mathrm{d}L(\lambda) - \int_T \int_\Omega \varphi_t(f)\mathrm{d}L(P)\mathrm{d}L(\lambda) = \int_T \psi_t(f_\omega(t))\mathrm{d}L(\lambda) = 0,$$

for $L(P)$-almost all $\omega \in \Omega$. Hence, (4) holds.

To show (4) ⇒ (6), we note that φ has a uniform lifting ψ in the sense ψ is an internal mapping from $T \times {}^*X$ to ${}^*\mathbb{R}$ such that ${}^\circ\psi(t,x) = \varphi(t, {}^\circ x)$ for $L(\lambda)$-almost all $t \in T$ and all near-standard $x \in {}^*X$ (see Albeverio et al., 1986, p. 136). Let F be an internal lifting of f. Then the mapping G defined by $G(t,\omega) = \psi(t, F(t,\omega))$ is an internal lifting of g, and hence, g is a bounded measurable function on the Loeb product space. By (4), $\int_T \varphi_t(f_\omega(t))\mathrm{d}L(\lambda) = \int_T \int_\Omega \varphi_t(f)\mathrm{d}L(P)\mathrm{d}L(\lambda)$ for $L(P)$-almost all $\omega \in \Omega$.

Now assume (6). Let φ be a bounded real-valued continuous function on X. For any $A \in T$, define a class $\{\varphi_t : t \in T\}$ of Borel functions by letting $\varphi_t = \varphi$ if $t \in A$ and $\varphi_t \equiv 0$ if $t \notin A$. By (6),

$$\int_T \varphi_t(f_\omega(t))\mathrm{d}L(\lambda) = \int_T \int_\Omega \varphi_t(f)\mathrm{d}L(P)\mathrm{d}L(\lambda)$$

for $L(P)$-almost all $\omega \in \Omega$. Hence,

$$\int_A \varphi(f_\omega(t))\mathrm{d}L(\lambda) = \int_A \int_\Omega \varphi(f)\mathrm{d}L(P)\mathrm{d}L(\lambda)$$

for $L(P)$-almost all $\omega \in \Omega$. Thus $\varphi(f)$ satisfies the consistency law. It means that Proposition 7.13 (4) holds. By Theorem 7.6 and Proposition 7.13, (1) follows. □

If we are given a process with almost surely pairwise independent random variables, then the sample functions are almost surely identically distributed. It will be interesting to see when the sample functions are also almost surely pairwise independent. The following proposition, which is an analog of Proposition 4.7, offers a perfect answer.

Proposition 7.16. *Let f be a process from $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$ to a separable metric space X. Then the random variables f_t are almost surely pairwise independent and almost surely identically distributed if and only if so are the sample functions.*

Proof: Assume that the random variables f_t are almost surely pairwise independent and almost surely identically distributed. Let φ be a bounded real-valued continuous function on X. Then the random variables $(\varphi(f))_t$ are uncorrelated with constant means almost surely, and so are the sample functions $(\varphi(f))_\omega$ by Proposition 4.7. By symmetry, we can obtain a version of the equivalence of Proposition 7.13 (4) and almost sure pairwise independence in terms of sample functions. Then the almost sure uncorrelatedness of the functions $(\varphi(f))_\omega$ together with the arbitrary choice of φ imply that the sample functions f_ω are almost surely pairwise independent. By Theorem 7.6, the sample functions f_ω are also almost surely identically distributed. By symmetry, we can also obtain the converse. □

It is indicated in Section 5.4 that a discrete parameter stochastic process can be regarded as a random variable taking values in some product metric space. As noted in Remark 5.9, we can obtain a version of the duality result in Proposition 7.16 for a hyperfinite number of discrete parameter stochastic processes.

Corollary 7.17. *Let f be a hyperprocess from $(T \times \Omega \times \mathbb{Z}^+)$ to a separable metric space X. Then the discrete parameter stochastic processes f_t on $\mathbb{Z}^+$ are almost surely pairwise independent and essentially have the same finite dimensional distributions with some fixed discrete parameter stochastic process if and only if so are the discrete parameter empirical processes f_ω on $T \times \mathbb{Z}^+$.*

8. Continuous parameter stochastic processes and the law

8.1. Introduction

In Sections 8.2, 8.3, 8.4 and 8.5, we study a hyperfinite number of continuous parameter stochastic processes, i.e., a hyperprocess. Section 8.2 presents several versions of the law for hyperprocesses in finite dimensional distribution. Section 8.3 contains the universality of atomless Loeb product spaces in the setting of hyperprocesses. We study in Section 8.4 the representation and homogeneity properties for hyperprocesses relevant to the Loeb counting spaces. Section 8.5 concerns the law for the special iid case. The version of such kind result in the usual continuum setting is often assumed in the economic literature (though it is, in general, not correct). The modeling of some macroscopic physical phenomena by hyperprocesses are also discussed in this section.

8.2. Finite dimensional distributions and the law

As noted in Keisler (1988), the following lemma of Hoover and Keisler (1984) (p. 172) is often helpful in extending results concerning random variables to results concerning stochastic processes. Note that the target space is not necessarily complete.

Lemma 8.1. *Let $x(\omega, i)$ be a stochastic process from a product probability space $(\Lambda \times [0, 1], \mathscr{B} \otimes \mathscr{I}, \nu \otimes \mu)$ to a separable metric space X. Then there is a sequence $\{i_n\}_{n=0}^{\infty}$ of elements in [0,1] and two Borel functions $\psi: [0,1] \times X^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$, $\phi: \{0,1\}^{\mathbb{N}} \to X$ such that for all $i \in [0,1]$, $x(\omega, i) = \phi(\psi(i,\{x(\omega, i_n)\}_{n=0}^{\infty}))$ for ν-almost all $\omega \in \Lambda$.*

We shall now study the law of large numbers for a hyperfinite number of continuous parameter stochastic processes. For simplicity, we only state the results for the case that the continuous parameter space is the unit Lebesgue interval $(I, \mathscr{I}, \mu)$, where $I = [0,1]$, $\mathscr{I}$ is the collection of Lebesgue measurable sets in I, and μ is the Lebesgue measure on I. By a continuous parameter stochastic process on a probability space, we shall mean a measurable mapping on the product measure space of a probability space with the Lebesgue interval (see, for example, Keisler, 1988, p. 112). In this section, we shall consider a hyperfinite number of continuous parameter stochastic processes on a Loeb product space $(T \times \Omega, L(\mathscr{T} \otimes \mathscr{A}), L(\lambda \otimes P))$. It simply means a measurable mapping on the mixed product space $(T \times \Omega \times I, L(\mathscr{T} \otimes \mathscr{A}) \otimes \mathscr{I}, L(\lambda \otimes P) \otimes \mu)$ (called a hyperprocess). We shall interpret the set T as an index space, Ω a sample space, and I the time space or the parameter space. For this section, we simply assume that $\mathscr{T}$ is the collection of all internal subsets of a hyperfinite set T.

Now, let f be a hyperprocess from $(T\times\Omega\times I, L(\mathscr{T}\otimes\mathscr{A})\otimes\mathscr{I}, L(\lambda\otimes P)\otimes\mu)$ to a separable metric space X. As indicated by Stroyan and Bayod (1986) (p. 164), one can obtain a Fubini type result for real or complex valued hyperprocesses on mixed product spaces. One can do that simply by working with some internal liftings of the hyperprocesses. Then, as shown in the paragraph above Theorem 5.2, we can obtain that for $L(\lambda)$-almost all $t\in T$, f_t is $L(\mathscr{A})\times\mathscr{I}$-measurable. The stochastic processes f_t with sample space Ω and parameter space I will be called the (continuous parameter) stochastic processes in the hyperprocess f. Similarly, the sample functions f_ω are stochastic processes with sample space T and parameter space I, which will be called the empirical processes in the hyperprocess f. The hyperprocess f can also be viewed as a continuous parameter stochastic process with sample space $T\times\Omega$ and parameter space I. For a given tuple $i_1,\dots,i_n$ in I, $\alpha^{i_1\cdots i_n}=(f_{i_1}\dots f_{i_n})$ is a process on the Loeb product space. The collection of the hyper distributions of the processes $\alpha^{i_1\cdots i_n}$ for all possible tuples is called the finite dimensional hyper distributions of f.

Note that two continuous parameter stochastic processes φ and ψ on some sample spaces (possibly different) with parameter space I are said to have the same finite dimensional distributions, if for any $i_1,\dots,i_n\in I$, the random variables $(\varphi_{i_1},\dots,\varphi_{i_n})$ and $(\psi_{i_1},\dots,\psi_{i_n})$ have the same distribution. When φ and ψ are on the same sample space, we say that they are independent, if the random variables $(\varphi_{i_1^1},\dots,\varphi_{i_m^1})$ and $(\psi_{i_1^2},\dots,\varphi_{i_n^2})$ are independent for any $i_1^1,\dots,i_m^1$ and $i_1^2,\dots,i_n^2$ in I.

Note that for stochastic processes whose paths are coming from some function space with a separable metric, for example, the continuous function spaces or the Skorokhod spaces, they can be regarded as random variables in the function space. If the stochastic processes in a hyperprocess have the properties just described, then one can usually regard the hyperprocess as a function space valued process on a Loeb product space, and the law for such hyperprocesses follows from Theorem 5.2. This is the approach used for the discrete case. In this section, we shall consider the case of general continuous parameter stochastic processes. The following theorem presents a law of large numbers for general continuous parameter hyperprocesses.

Theorem 8.2. *Let f be a hyperprocess from a mixed product space $(T\times\Omega\times I, L(\mathscr{T}\otimes\mathscr{A})\otimes\mathscr{I}, L(\lambda\otimes P)\otimes\mu)$ to a separable metric space X. Assume that the stochastic processes f_t in the hyperprocess f are almost surely pairwise independent, i.e., for $L(\lambda\otimes\lambda)$-almost all $(t_1, t_2)\in T\times T$, the stochastic processes f_{t_1} and f_{t_2} are independent. Then for $L(P)$-almost all $\omega\in\Omega$, the empirical process f_ω on $T\times I$ and the hyperprocess f viewed as a stochastic process on $(T\times\Omega)\times I$ have the same finite dimensional distributions.*

Proof: We first regard f as a stochastic process with I as the parameter space and $T\times\Omega$ as the sample space. Then Lemma 8.1 implies that there is a sequence

$\{i_n\}_{n=0}^{\infty}$ of elements in [0, 1] and two Borel functions ψ:$[0,1]\times X^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$, ϕ:$\{0,1\}^{\mathbb{N}} \to X$ such that for all $i \in [0,1]$, $f(t,\omega,i) = \phi(\psi(i,\{f(t,\ \omega,i_n)\}_{n=0}^{\infty}))$ for $L(\lambda \otimes P)$-almost all $(t,\omega) \in T \times \Omega$. Define a process α from $T \times \Omega$ into $X^{\mathbb{N}}$ by letting $\alpha(t,\ \omega) = \{f(t,\ \omega,\ i_n)\}_{n=0}^{\infty}$. The fact that the stochastic processes f_t are almost surely pairwise independent implies that the random variables α_t are almost surely pairwise independent. By Theorem 5.2, we have for $L(P)$-almost all $\omega \in \Omega$, the distribution ν_ω on $X^{\mathbb{N}}$ induced by the random variable α_ω on T is equal to the distribution ν on $X^{\mathbb{N}}$ induced by the process α viewed as a random variable on $T \times \Omega$. Note that the Fubini theorem also implies that for any given $i \in [0,1]$, we have for $L(P)$-almost all $\omega \in \Omega$, $f_\omega(t,i) = \phi(\psi(i,\alpha_\omega(t)))$ for $L(\lambda)$-almost all $t \in T$. Since α and α_ω have the same distribution as random variables, we thus obtain that for $L(P)$-almost all $\omega \in \Omega$, the stochastic processes f_ω and f have the same finite dimensional distributions. □

As noted in Proposition 7.15, the almost sure pairwise independence condition is convenient for applications, since it is preserved under various transformations. For example, let Φ be a mapping from $T \times I \times X$ to another separable metric space Y such that the mapping g from $(T \times \Omega \times I,\ L(\mathcal{T} \otimes \mathcal{A}) \otimes \mathcal{I},\ L(\lambda \otimes P) \otimes \mu)$ to Y defined by $g(t,\omega,i) = \Phi(t,i,f(t,\ \omega,\ i))$ is $L(\mathcal{T} \otimes \mathcal{A}) \otimes \mathcal{I}$-measurable. If the stochastic processes f_t are almost surely pairwise independent, then so are the stochastic processes g_t. Same as before, the following notion is useful for the characterization of almost sure pairwise independence.

Definition 8.3. Let f be a hyperprocess from $(T \times \Omega \times I,\ L(\mathcal{T} \otimes \mathcal{A}) \otimes \mathcal{I},\ L(\lambda \otimes P) \otimes \mu)$ to a metric space X. We say that f satisfies the consistent law of large numbers in finite dimensional distribution (or simply the consistency law in finite dimensional distribution) if for any internal set $A \in \mathcal{T}$ with $L(\lambda)(A) > 0$, the hyperprocess f^A from $(A \times \Omega \times I,\ L(\mathcal{T}^A \otimes \mathcal{A}) \otimes \mathcal{I},\ L(\lambda^A \otimes P) \otimes \mu)$ to X satisfies the law in finite dimensional distribution in X. In other words, for $L(P)$-almost all $\omega \in \Omega$, and for all $i_1,\dots,i_n \in I$, the distribution $\tau_\omega^{A(i_1\cdots i_n)}$ on X^n induced by the random variable $(f^A(\cdot,\omega,\ i_1),\dots,f^A(\cdot,\omega,i_n))$ from $(A,\ L(\mathcal{T}^A),\ L(\lambda^A))$ to X^n is equal to the distribution $\tau^{A(i_1\cdots i_n)}$ on X^n induced by the random variable $(f^A(\cdot,\cdot,i_1),\dots,f^A(\cdot,\cdot,i_n))$ from $(A \times \Omega,\ L(\mathcal{T}^A \otimes \mathcal{A}),\ L(\lambda^A \otimes P))$ to X^n. Here, f^A is the restriction of f to $A \times \Omega \times I$, $\mathcal{T}^A$ 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.

Same as in Definitions 3.15 and 7.3, we can only consider those A satisfying $L(\lambda)(A) > 1 - \varepsilon$ for some fixed $\varepsilon \in (0,1)$ in the above definition. In Theorem 8.2, we only prove that f satisfies the law in finite dimensional distribution. It is clear that the same proof leads to the fact that f satisfies the consistency law in finite dimensional distribution. By Theorem 7.6, the almost sure pairwise independence assumption on the hyperprocess f is also necessary for the validity of the consistency law (see the following proposition whose proof is omitted). We will not attempt to write down as many characterizations of almost sure pairwise independence as in Section 7. From the proof of Theorem 8.2, it is straightforward to rewrite many of the results in Section 7 in the setting of hyperprocesses.

Theorem 8.4. *Let f be a hyperprocess from $(T \times \Omega \times I, L(\mathscr{T} \otimes \mathscr{A}) \otimes \mathscr{I}, L(\lambda \otimes P) \otimes \mu)$ to a separable metric space X, and let $\mathscr{E}$ be a separating class of Borel functions (real-valued or possibly complex valued) for all Borel probability measures on $X^{\mathbb{N}}$. Then the following are equivalent:*

(1) The stochastic processes f_t in the hyperprocess are almost surely pairwise independent;

(2) f satisfies the consistency law in finite dimensional distribution;

(3) For each $\gamma \in \mathscr{E}$, and for each sequence $\sigma = \{j_n\}_{n=0}^{\infty}$ in I, let α^{σ} be the process on $T \times \Omega$ defined by $\alpha^{\sigma}(t,\omega) = \gamma(f(t,\omega j_0), f(t,\omega,j_1), \ldots, f(t,\omega,j_n), \ldots)$; then the random variables α_t^{σ} are almost surely uncorrelated.

8.3. Universality for hyperprocesses

As for the study of processes, it is important to know whether there exist non-trivial hyperprocesses with almost surely pairwise independent stochastic processes. Otherwise, the results presented in Section 8.2 will be vacuous. To show the universality of atomless Loeb product spaces in the setting of hyperprocesses, we need the following lemma.

Lemma 8.5. *Let α be a process from a Loeb product space $(T_1 \times \Omega_1, L(\mathscr{T}_1 \otimes \mathscr{A}_1), L(\lambda_1 \otimes P_1))$ into a separable metric space X, and β a process from a probability measure space of the form $(T_2 \times \Omega_2, \mathscr{D}, \lambda_2 \otimes P_2)$ to X, where the probability spaces in the second product need not be Loeb spaces, and $\mathscr{D}$ is a σ-algebra which contains the product of σ-algebras on T_2 and Ω_2 such that the Fubini property is still satisfied. Let F be a Borel mapping from X into another separable metric space Y. Assume that α and β have the same hyper distribution. Then $F \circ \alpha$ and $F \circ \beta$ have the same hyper distribution.*

Proof: Note that the mapping $F_{\mathscr{M}}$ from $\mathscr{M}(X)$ to $\mathscr{M}(Y)$ definedby $F_{\mathscr{M}}(\nu) = \nu F^{-1}$ is Borel measurable. Let ϕ be any bounded Borel function on $\mathscr{M}(Y)$. Then $\phi \circ F_{\mathscr{M}}$ is a bounded Borel function on $\mathscr{M}(X)$. Since α and β have the same hyper distribution, we have

$$\int_{T_1} \phi \circ F_{\mathscr{M}}\left(L(P_1)\, \alpha_{t_1}^{-1} \right) \mathrm{d}L(\lambda_1) = \int_{T_2} \phi \circ F_{\mathscr{M}}\left(P_2\, \beta_{t_2}^{-1} \right) \mathrm{d}\lambda_2,$$

which implies that

$$\int_{T_1} \phi\left(\left(L(P_1)\, \alpha_{t_1}^{-1} \right) F^{-1} \right) \mathrm{d}L(\lambda_1) = \int_{T_2} \phi\left(\left(P_2\, \beta_{t_2}^{-1} \right) F^{-1} \right) \mathrm{d}\lambda_2,$$

and hence

$$\int_{T_1}\phi\Big(L(P_1)\big((F\circ\alpha)_{t_1}\big)^{-1}\Big)\mathrm{d}L(\lambda_1)=\int_{T_2}\phi\Big(P_2\big((F\circ\beta)_{t_2}\big)^{-1}\Big)\mathrm{d}\lambda_2,$$

Therefore $F\circ\alpha$ and $F\circ\beta$ have the same hyper distribution. □

By applying Theorem 6.2, we can obtain the following universality property of atomless Loeb product spaces in terms of hyperprocesses.

Proposition 8.6. *Let g be any hyperprocess from a triple product measure space $(\Sigma\times\Theta\times I, \mathscr{C}\otimes\mathscr{D}\otimes\mathscr{I}, \nu\otimes\tau\otimes\mu)$ to a Polish space X. If both $L(\lambda)$ and $L(P)$ are atomless, then there is a hyperprocess f from the mixed product space $(T\times\Omega\times I, L(\mathscr{T}\otimes\mathscr{A})\otimes\mathscr{I}, L(\lambda\otimes P)\otimes\mu)$ such that f and g have the same finite dimensional hyper distributions, and moreover for almost all $(t_1,t_2)\in T\times T$, the stochastic processes f_{t_1} and f_{t_2} are independent.*

Proof: By Lemma 8.1, there is a sequence $\{i_n\}_{n=0}^{\infty}$ of elements in [0,1] and two Borel functions ψ: $[0,1]\times X^{\mathbb{N}}\to\{0,1\}^{\mathbb{N}}$, ϕ:$\{0,1\}^{\mathbb{N}}\to X$ such that for all $i\in[0,1]$, $g(\sigma,\theta,i)=\phi(\psi(i,\{g(\sigma,\theta,i_n)\}_{n=0}^{\infty}))$ for $(\nu\otimes\tau)$-almost all $(\sigma,\theta)\in\Sigma\times\Theta$. Define a process α from $\Sigma\times\Theta$ into $X^{\mathbb{N}}$ by letting $\alpha(\sigma,\theta)=\{g(\sigma,\theta,i_n)\}_{n=0}^{\infty}$. Theorem 6.2 implies that there is a process β on $T\times\Omega$ such that the random variables β_t are almost surely pairwise independent, and α, β have the same hyper distribution. Define a hyperprocess f on $T\times\Omega\times I$ by letting $f(t,\omega,i)=\phi(\psi(i,\beta(t,\omega)))$. Then for any given $j_1,\dots,j_m\in I$, consider the hyper distributions of the processes $(g_{j_1},\dots,g_{j_m})$ and $(f_{j_1},\dots,f_{j_m})$. Let F be the mapping from $X^{\mathbb{N}}$ to X^m defined by

$$F\big(\{x_n\}_{n=0}^{\infty}\big)=\Big(\phi\Big(\psi\big(j_1,\{x_n\}_{n=0}^{\infty}\big)\Big),\cdots,\phi\Big(\psi\big(j_m,\{x_n\}_{n=0}^{\infty}\big)\Big)\Big).$$

Then F is a Borel mapping and $(g_{j_1},\dots,g_{j_m})=F\circ\alpha$ and $(f_{j_1},\dots,f_{j_m})=F\circ\beta$. Since α and β have the same hyper distribution, Lemma 8.5 shows that $F\circ\alpha$ and $F\circ\beta$ have the same hyper distribution. Therefore f and g have the same finite dimensional hyper distributions. The almost sure independence of the stochastic processes f_t are clear from that of the random variables β_t. □

8.4. Some structural results for stochastic processes

This section concerns with the representations of stochastic processes or hyperprocesses involving hyperfinite Loeb counting spaces. Some structural results on hyperprocesses with the same finite dimensional hyper distributions are also presented. We shall first consider the representation of one stochastic process in the following proposition. A proof can be given by using Lemma 8.1 and Corollary 5.4.

Proposition 8.7. *Let x be a stochastic process from* $(\Omega_1 \times I, L(\mathscr{A}_1) \otimes \mathscr{I}, L(P) \otimes \mu)$ *to a separable metric space X, where* $(\Omega_1, L(\mathscr{A}_1), L(P_1))$*, is a Loeb space with* $L(P_1)$ *being atomless. Then there is a hyperfinite Loeb counting space* $(\Omega_2, L(\mathscr{A}_2), L(P_2))$*, a measure preserving mapping* π*, and a process y on* $\Omega_2 \times I$ *such that for all* $t \in [0, 1]$*,* $x(\omega_1,t) = y(\pi(\omega_1),t)$ *for* $L(P_1)$*-almost all* $\omega_1 \in \Omega_1$*.*

Next, we present a result on the parametric representation of hyperprocesses. The result can be proven by Lemma 8.1 and Corollary 5.5. Note that the derived hyperprocess g below has a hyperfinite Loeb counting sample space.

Proposition 8.8. *Let f be a hyperprocess from a mixed product space* $(T \times \Omega_1 \times I, L(\mathscr{T} \otimes \mathscr{A}_1) \otimes \mathscr{I}, L(\lambda \otimes P_1) \otimes \mu)$ *to a separable metric space X. Assume that* $L(P_1)$ *is atomless. Then there is a hyperfinite Loeb counting space* $(\Omega, L(\mathscr{A}), L(P))$*, an internal function* π *from* $T \times \Omega_1$ *to* Ω *with* π_t *measure preserving from* $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ *to* $(\Omega, L(\mathscr{A}), L(P))$ *for each* $t \in T$*, and a hyperprocess g on* $T \times \Omega \times I$ *such that for all* $i \in [0,1]$*,* $f(t, \omega_1, i) = g(t,\pi(t,\omega_1),i)$ *for* $L(\lambda \otimes P_1)$*-almost all* $(t, \omega_1) \in T \times \Omega_1$*.*

To prove the parametric homogeneity for stochastic processes, we need a stronger version of Lemma 8.1, which allows us to take the same sequence of numbers to represent different stochastic processes.

Lemma 8.9. *Let* $x(\omega,i)$ *and* $y(\omega,i)$ *be stochastic processes from* $(\Lambda \times [0,1], \mathscr{B} \otimes \mathscr{I}, \nu \otimes \mu)$ *to a separable metric space X. Then there is a sequence* $\{i_n\}_{n=0}^{\infty}$ *of elements in [0, 1] and two Borel functions* $\psi: [0, 1] \times X^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}, \phi: \{0,1\}^{\mathbb{N}} \to X$ *such that for all* $i \in [0,1]$*,*

$$x(\omega,i) = \phi\Big(\psi\big(i,\{x(\omega,i_n)\}_{n=0}^{\infty}\big)\Big) \text{ and } y(\omega,t) = \phi\Big(\psi\big(i,\{y(\omega,i_n)\}_{n=0}^{\infty}\big)\Big)$$

for ν-almost all $\omega \in \Lambda$.

Proof: Let τ be the probability measure on {0, 1} such that $\tau(\{0\}) = 1/2$ and $\overline{\Lambda} = \{0,1\} \times \Lambda$. Define $\overline{\mathscr{B}} = \{\{j\} \times B: j = 0,1; B \in \mathscr{B}\}$ and $\overline{\nu} = \tau \times \nu$. Define a process z from $(\overline{\Lambda} \times [0,1], \overline{\mathscr{B}} \otimes \mathscr{I}, \overline{\nu} \otimes \mu)$ to X by letting $z((j, \omega), i) = x(\omega, i)$ for $j = 0$, and $z((j, \omega),i) = y(\omega, i)$ for $j = 1$. Apply Lemma 8.1 to the stochastic process z to obtain the desired result. □

Note that the above result is still valid for a given (finite or infinite) sequence of stochastic processes. The same proof works as well. We shall now present the parametric homogeneity property in the setting of hyperprocesses. For the case of two processes with the same finite dimensional distributions, see Keisler (1988).

Proposition 8.10. *Let f and g be hyperprocesses from a mixed product space $(T \times \Omega \times I,\ L(\mathscr{T} \otimes \mathscr{A}) \otimes \mathscr{I},\ L(\lambda \otimes P) \otimes \mu)$ to a separable metric space X such that for $L(\lambda)$-almost all $t \in T$, the stochastic processes f_t and g_t have the same finite dimensional distributions. Assume that P is the internal counting probability measure on Ω. Then there is an internal function π from $T \times \Omega$ to Ω such that for each $t \in T$, π_t is a permutation of Ω, and for all $i \in I$, $g(t,\omega,i) = f(t,\pi_t(\omega), i)$ for $L(\lambda \otimes P)$-almost all $(t,\omega) \in T \times \Omega$.*

Proof: Lemma 8.9 implies that there is a sequence $\{i_n\}_{n=0}^{\infty}$ of elements in [0,1] and two Borel functions $\psi:[0,1] \times X^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$, $\phi:\{0,1\}^{\mathbb{N}} \to X$ such that for all $i \in [0,1]$,

$$f(t,\omega,i) = \phi\Big(\psi\Big(i,\{f(t,\omega,i_n)\}_{n=0}^{\infty}\Big)\Big)$$

$$\text{and } g(t,\omega,i) = \phi\Big(\psi\Big(i,\{g(t,\omega,i_n)\}_{n=0}^{\infty}\Big)\Big)$$

for $L(\lambda \otimes P)$-almost all $(t,\omega) \in T \times \Omega$. Define two processes α and β from $T \times \Omega$ into $X^{\mathbb{N}}$ by letting $\alpha(t,\omega) = \{f(t,\omega,i_n)\}_{n=0}^{\infty}$ and $\beta(t,\omega) = \{g(t,\omega,i_n)\}_{n=0}^{\infty}$. Since for $L(\lambda)$-almost all $t \in T$, f_t and g_t have the same finite dimensional distributions, α_t and β_t have the same distribution for $L(\lambda)$-almost all $t \in T$. Thus Proposition 5.7 implies that there is an internal process π from $T \times \Omega$ to Ω such that for each $t \in T$, π_t is a permutation of Ω, and for $L(\lambda)$-almost all $t \in T$, $\beta_t = \alpha_t \circ \pi_t$. The rest is then clear. □

Finally, by using Lemma 8.9 and Proposition 6.3, we can characterize those hyperprocesses with a hyperfinite Loeb counting index space and having the same finite dimensional hyper distributions.

Proposition 8.11. *Let f and g be hyperprocesses from $(T \times \Omega_1 \times I,\ L(\mathscr{T} \otimes \mathscr{A}_1) \otimes \mathscr{I},\ L(\lambda \otimes P_1) \otimes \mu)$ to a separable metric space X. Assume that λ is the internal counting probability measure on the hyperfinite set T and $L(P_1)$ is atomless. Then f and g have the same finite dimensional hyper distributions if and only if there exists an internal permutation π of T, a hyperfinite Loeb counting space $(\Omega, L(\mathscr{A}), L(P))$, and a hyperprocess h from $(T \times \Omega \times I,\ L(\mathscr{T} \otimes \mathscr{A}) \otimes \mathscr{I},\ L(\lambda \otimes P) \otimes \mu)$ to X with the following property: for any $i \in I$,*

$$f(t,\omega_1,i) = h\big(t,p_t(\omega_1),i\big) \text{ and } g(t,\omega_1,i) = h\big(\pi(t),q_t(\omega_1),i\big),$$

for $L(\lambda \otimes P_1)$-almost all $(t,\omega_1) \in T \times \Omega$, where p and q are internal mappings from $T \times \Omega_1$ to Ω, and for each fixed $t \in T$, p_t and q_t are Loeb measure preserving from $(\Omega_1, L(\mathscr{A}_1), L(P_1))$ to $(\Omega, L(\mathscr{A}), L(P))$.

8.5. The folk law and some related remarks

If we have a continuum of objects which act almost independently and the movement of each of them along time is governed by a stochastic process with the same finite dimensional distributions as a fixed stochastic process Y, then the positions of all the objects in almost all states of nature, are still governed by processes with the same finite dimensional distributions as Y. Thus, what we can observe at time t about the distribution of the positions of all the objects is precisely the prior distribution of the random variable Y_t. This observation has been regarded as a folk result, and are already widely used in the economic literature. As noted earlier in the introduction of the paper, the result does not make much sense if the usual continuum is used to model the large number of objects. The following theorem provides a firm mathematical foundation for such a folk law in the hyperfinite setting. The result is also an analog of Propositions 4.7 and 7.16 in the setting of hyperprocesses.

Theorem 8.12. *Let f be a hyperprocess from $(T \times \Omega \times I, L(\mathcal{T} \otimes \mathcal{A}) \otimes \mathcal{I}, L(\lambda \otimes P) \otimes \mu)$ to a separable metric space X. Then the stochastic processes f_t are almost surely pairwise independent and have the same finite dimensional distributions with a common stochastic process h almost surely on a product measure space $(\Lambda \times I, \mathcal{B} \otimes \mathcal{I}, \nu \otimes \lambda)$ if and only if so are the empirical stochastic processes f_ω.*

Proof: We only consider the 'only if' part. The other part follows from symmetry. We first show that the hyperprocess f viewed as a stochastic process on $T \times \Omega$ has the same finite dimensional distributions as h. Note that almost sure pairwise independence is not needed for this point. Choose $i_1, \dots, i_n$ from I. For any bounded continuous functions ϕ on the n-fold Cartesian product X^n of X, we have

$$\begin{aligned}\int\int_{T\times\Omega} \phi(f_{i_1}, \cdots, f_{i_n}) \mathrm{d}L(\lambda \times P) &= \int_T \int_\Omega \phi((f_{i_1})_t, \cdots, (f_{i_n})_t) \mathrm{d}L(P) \mathrm{d}L(\lambda) \\ &= \int_T \int_\Lambda \phi(h_{i_1}, \cdots, h_{i_n}) \mathrm{d}\nu \mathrm{d}L(\lambda) \\ &= \int_\Lambda \phi(h_{i_1}, \cdots, h_{i_n}) \mathrm{d}\nu.\end{aligned}$$

Hence, f and h have the same finite dimensional distributions.

Next, by the almost sure pairwise independence assumption on the stochastic processes f_t, Theorem 8.2 implies that for $L(P)$-almost all $\omega \in \Omega$, the empirical process f_ω has the same finite dimensional distributions as f, and hence h.

It remains to show that the empirical stochastic processes f_ω are almost surely pairwise independent. As in the proof of Theorem 8.2, f can be expressed as $f(t,\omega,i) = \phi(\psi(i, \alpha(t,\omega)))$, where $\alpha(t,\omega) = \{f(t,\omega,i_n)\}_{n=0}^{\infty}$. It is easy to see that the random variables α_t are pairwise independent and identically distributed almost surely. By Proposition 7.16, the sample functions α_ω are also pairwise independent and identically distributed almost surely, and so are the empirical processes f_ω. □

As an example, we consider the case that there is a relatively large number of small particles suspended in some fluid. Assume that the number of particles is small compared with the number of molecules in the fluid. For a given particle, the number of particles interacting with this particle is considered to be small. Thus, we can assume that the particles act almost independently. Since the total number of particles is assumed to be comparatively small, the environment for the movement of each particle is similar to that of one particle case. Hence, the erratic movement of each of such particles can still be modeled by the same Brownian motion B as the one particle case by modifying its initial positions. Then the distribution of all the particles is still governed by B with some modification at time 0. Such an observation is hardly surprising from the classical law of large numbers. However, it is rather curious that the usual mathematical framework of the theory of stochastic processes cannot provide a rigorous and meaningful formulation of the problem as indicated in the beginning of the paper. Of course, one could say that it is already well known that the empirical process of a large finite number of particles is close to the Brownian motion. But in a continuum model (here, we use our special continuum model, the hyperfinite model), the fact that the empirical process is *exactly* the Brownian motion tells us much more in the sense that tremendously many other results on Brownian motion are now directly applicable to the empirical process. The importance and the convenience of the idealization of a large finite number of particles to a continuum of particles in this setting cannot be underestimated. From another perspective, if there is any objection to the use of continuum as the number of particles here, then one could also question the use of continuum models in other aspects. For example, the total number of fluid molecules in the tank is finite, thus the problem about the motion of a particle is in fact a many-body problem, which seems hopeless to solve. So we have certainly already used continuous mathematics to model a large finite phenomenon for the same problem.

9. Some asymptotic interpretations

9.1. Introduction

In Sections 9.2 and 9.3, we shall translate several results presented in the previous sections to obtain information about triangular arrays or sequences of random variables. Only real-valued random variables will be considered. In Section 9.2, we consider asymptotic orthogonality as well as asymptotic uncorrelatedness. It was noted by Doob (1953) (p. 148) that the expectations of a sequence of orthogonal random variables may not vanish. Proposition 9.1, however, shows that the asymptotic orthogonality of a sequence (or more generally, a triangular array) of random variables is sufficient to imply that the expectations tend to zero asymptotically. Proposition 9.2 shows that asymptotic uncorrelatedness is necessary and sufficient for a triangular array of random variables to satisfy the consistency law. In Section 9.3, it is shown that asymptotic pairwise independence is necessary and sufficient for a triangular array of random variables to satisfy the consistency law in distribution.

Plenty of work has been devoted to the study of random variables with mixing or weak dependence conditions (see, for example, Billingsley, 1968; Lai, 1977; Philipp and Stout, 1975; Rosenblatt, 1956 and Shao, 1993). Since the notion of weak dependence is often used to mean that any random variable in a given large collection of random variables is approximately independent in some sense to most other random variables in the collection, it is likely that our asymptotic pairwise independence is a most general version of weak dependence which implies the validity of the law of large numbers.

We shall now fix some notation for Sections 9.2 and 9.3. Let $(\Omega, \mathscr{A}, P)$ be a fixed probability space which will be used as the common sample space of the triangular array of random variables to be considered. For each $n \geq 1$, let $(T_n, \mathscr{T}_n, \lambda_n)$ be finite probability spaces which will be the n-th index set, and we shall assume that the number of points in T_n goes to infinity as $n \to \infty$; let g_n be a real-valued process from $T_n \times \Omega$ to $\mathbb{R}$ such that $g_n(t, \cdot)$ is a random variable on Ω for each $t \in T_n$. The measure λ_n provides the weights for each point in the finite index set T_n and $\mathscr{T}_n$ is the power set of T_n. Such a sequence of processes $g = \{g_n\}_{n \geq 1}$ will be called a triangular array of random variables.

Note that the usual definition of a triangular array of random variables is set to be a sequence of random variables in the form, $x_1^n, x_2^n, \dots, x_n^n$, $n = 1,2,\dots$, which corresponds to our case when $T_n = \{1,2,\dots, n\}$ and λ_n is the uniform probability measure on T_n. So integrals on T_n are just the arithmetic averages. Here, we consider general weighted averages rather than the special arithmetic averages, since the proofs are the same.

9.2. Asymptotic orthogonality and asymptotic uncorrelatedness

A triangular array of random variables g, i.e., a sequence of processes g_n on $T_n \times \Omega$, $n \geq 1$, is said to be uniformly integrable if

$$\lim_{m \to \infty} \sup_{1 \leq n < \infty} \iint_{|g_n| > m} |g_n| \mathrm{d}\lambda_n \otimes P = 0.$$

We shall work with a fixed triangular array of random variables f, i.e., for each $n \geq 1$ and $t \in T_n$, $f_n(t, \cdot)$ is a real-valued random variable on Ω. The first proposition simply says that the asymptotic orthogonality of random variables in f implies that the means of the random variables tend to zero asymptoticly as well as the fact that the sample functions are asymptoticly orthogonal.

Proposition 9.1. *Assume that the triangular array of random variables f is uniformly square integrable, i.e., the sequence of processes f_n^2, $n \geq 1$, are uniformly integrable. For any $\varepsilon > 0$ and $n \geq 1$, define*

$$T_\varepsilon^n = \left\{ (t_1, t_2) \in T_n \times T_n : \left| \int_\Omega f_n(t_1, \omega) f_n(t_2, \omega) \mathrm{d}P \right| > \varepsilon \right\}$$

$$M_\varepsilon^n = \left\{ t \in T_n : \left| \int_\Omega f_n(t, \omega) \mathrm{d}P \right| > \varepsilon \right\}$$

$$\Omega_\varepsilon^n = \left\{ (\omega_1, \omega_2) \in \Omega \times \Omega : \left| \int_{T_n} f_n(t, \omega_1) f_n(t, \omega_2) \mathrm{d}\lambda_n(t) \right| > \varepsilon \right\}$$

.

If for any $\varepsilon > 0$, $\lim_{n\to\infty} (\lambda_n \otimes \lambda_n)(T_\varepsilon^n) = 0$, then for any $\varepsilon > 0$,

(1) $\lim_{n\to\infty} \lambda_n (M_\varepsilon^n) = 0$;

(2) $\lim_{n\to\infty} (P \otimes P)(\Omega_\varepsilon^n) = 0$.

Proof: We transfer the sequence to the non-standard universe to obtain a sequence $\{f_n\}_{n \in {}^*\mathbb{N}}$ of internal processes on the associated internal sequence $\{(T_n \times {}^*\Omega, \mathscr{T}_n \otimes {}^*\mathscr{A}, \lambda_n \otimes {}^*P) : n \in {}^*\mathbb{N}\}$ of probability spaces. The uniform integrability of the processes f_n^2 implies that for any $m \in {}^*\mathbf{N}_\infty$, $\sup_{n \in {}^*\mathbb{N}} \int_{f_n^2 > m} |f_n^2| \mathrm{d}(\lambda_n \otimes {}^*P) \simeq 0$. By Proposition 2.35 of Hurd and Loeb (1985) (p. 187), we know that each f_n^2 is S-integrable. That is,

$$\int\int_{T_n \times {}^*\Omega} {}^\circ f_n^2 \mathrm{d}L(\lambda_n \otimes {}^*P) = {}^\circ\int\int_{T_n \times {}^*\Omega} f_n^2 \mathrm{d}(\lambda_n \otimes {}^*P) < \infty.$$

By the assumption, we know that for any $\varepsilon \in \mathbb{R}^+$, $(\lambda_n \otimes \lambda_n)(T_\varepsilon^n) \simeq 0$ for $n \in {}^*\mathbb{N}_\infty$.

Next, we fix an $n \in {}^*\mathbb{N}_\infty$. By spillover, $L(\lambda_n \otimes \lambda_n)(T_\varepsilon^n) = 0$ for some positive infinitesimal ε. Thus, for $L(\lambda_n \otimes \lambda_n)$-almost all (t_1, t_2), $\int_{^*\Omega} {}^\circ f_n(t_1, \omega) {}^\circ f_n(t_2, \omega) dL({}^*P) = 0$. By Theorem 4.5, we know that the means of the random variables ${}^\circ f_n(t, \cdot)$ are almost surely 0. This implies that $\int_{^*\Omega} f_n(t, \omega) \mathrm{d}^*P \simeq 0$ for $L(\lambda_n)$-almost all $t \in T_n$. Thus for any $\varepsilon \in \mathbb{R}^+$, $\lambda_n(\{t \in T_n : |\int_{^*\Omega} f_n(t, \omega) \mathrm{d}^*P| > \varepsilon\})$ is an infinitesimal. Theorem 4.5 also shows that the sample functions ${}^\circ f_n(\cdot, \omega)$ are almost surely orthogonal. Thus, for any $\varepsilon \in \mathbb{R}^+$,

$$(P \otimes P)\left(\left\{(\omega_1, \omega_2) : \left|\int_{T_n} f_n(t, \omega_1) f_n(t, \omega_2) \mathrm{d}\lambda_n\right| > \varepsilon\right\}\right)$$

is also an infinitesimal.

Now, fix an $\varepsilon \in \mathbb{R}^+$. The above paragraph shows that $\lambda_n(M_\varepsilon^n) \simeq 0$ and also $(P \otimes P)(\Omega_\varepsilon^n) \simeq 0$ for any $n \in {}^*\mathbb{N}_\infty$. Hence, $\lim_{n\to\infty} \lambda_n(M_\varepsilon^n) = 0$ and $\lim_{n\to\infty}(P \otimes P)(\Omega_\varepsilon^n) = 0$. □

In the next proposition, we translate the equivalence of Theorem 4.6 (2) and (8) to the discrete case. In particular, asymptotic uncorrelatedness is necessary and sufficient to ensure a triangular array of random variables to satisfy the consistent law of large numbers. Note that item (2) given below is the asymptotic version of the phenomenon ‘‘No betting system can beat the house’’.

Proposition 9.2. *Assume that the triangular array of random variables f is uniformly square integrable. For any $\varepsilon > 0$ and $n \geq 1$, define*

$$T_\varepsilon^n = \Big\{ (t_1, t_2) \in T_n \times T_n : \Big| \int_\Omega f_n(t_1, \omega) f_n(t_2, \omega) \mathrm{d}P(\omega) - \int_\Omega f_n(t_1, \omega) \mathrm{d}P \int_\Omega f_n(t_2, \omega) \mathrm{d}P(\omega) \Big| > \varepsilon \Big\}$$

$$L_\varepsilon^{A_n} = \Big\{ \omega \in \Omega : \Big| \int_{A_n} f_n(t, \omega) \mathrm{d}\lambda_n(t) - \int\int_{A_n \times \Omega} f_n \mathrm{d}(\lambda_n \otimes P) \Big| > \varepsilon \cdot \lambda_n(A_n) \Big\}.$$

Then the following are equivalent:

(1) For any $\varepsilon > 0$, $\lim_{n \to \infty} (\lambda_n \otimes \lambda_n)(T_\varepsilon^n) = 0$;

(2) There is a number $\delta \in (0,1)$ such that for any subset $A_n \subseteq T_n$ with $\lambda_n(A_n) > 1 - \delta$, $\lim_{n \to \infty} P(L_\varepsilon^{A_n}) = 0$ for any $\varepsilon > 0$.

Proof: The proof of (1) ⇒ (2) is similar to that of Proposition 9.1. So we only consider (2) ⇒ (1). Assume (2). We transfer the relevant sequences to internal sequences as in the proof of Proposition 9.1. Fix any $n \in {}^*\mathbb{N}_\infty$. Denote $g = {}^\circ f_n$. We shall omit the sub-index n in rest of this paragraph for simplicity. (2) implies that for any fixed $A \in \mathscr{T}$ with $L(\lambda)(A) > 1 - \delta/2$, $P(L_\varepsilon^A) \simeq 0$ for some positive infinitesimal ε. Hence, for almost all $\omega \in {}^*\Omega$,

$$\int_A f(t, \omega) \mathrm{d}\lambda \simeq \int\int_{A \times {}^*\Omega} f \mathrm{d}(\lambda \otimes {}^*P),$$

which implies that

$$\int_A g^A(t, \omega) \mathrm{d}L(\lambda^A)(t) = \int\int_{A \times {}^*\Omega} g \mathrm{d}L(\lambda^A \otimes {}^*P).$$

Thus, the process g satisfies the consistency law by Lemma 3.16. By Theorem 4.6, the random variables g_t are almost surely uncorrelated. Hence, for any $\varepsilon \in \mathbb{R}^+$,

$$\Big| \int_{{}^*\Omega} f(t_1, \omega) f(t_2, \omega) \mathrm{d}^*P(\omega) - \int_{{}^*\Omega} f(t_1, \omega) \mathrm{d}^*P(\omega) \int_{{}^*\Omega} f(t_2, \omega) \mathrm{d}^*P(\omega) \Big| > \varepsilon$$

for $L(\lambda \otimes \lambda)$-almost all $(t_1, t_2) \notin T \times T$.

Now we resume the index n and also fix an $\varepsilon \in \mathbb{N}^+$ as before. The previous paragraph shows that $(\lambda_n \otimes \lambda_n)(T_\varepsilon^n) \simeq 0$ for any $n \in {}^*\mathbb{N}_\infty$. Hence, $\lim_{n \to \infty}(\lambda_n \otimes \lambda_n)(T_\varepsilon^n) = 0$, and (1) follows. □

9.3. Intrinsic characterizations of asymptotic pairwise independence

This section presents intrinsic characterizations of asymptotic pairwise independence. We start with some technical definitions. Let $\mathscr{F}$ be the space of distribution functions of real-valued random variables. As usual, we assume distribution functions are right continuous. For $F, F' \in \mathscr{F}$, let $d(F, F')$ to be the infimum of all those h for which $F(x-h)-h \le F'(x) \le F(x+h)+h$ whatever be $x \in \mathbb{R}$. Then d define a metric and the space $(\mathscr{F}, d)$ is called Lévy's space (see Loéve, 1977a, p. 228). We shall refer to d as the Lévy metric. Note that convergence in $(\mathscr{F}, d)$ is equivalent to convergence in distribution. For distribution functions on $\mathbb{R}^2$, we can define a similar distance d_2 such that $d_2(F, F')$ is the infimum of all those positive number h for which $F(x-h, y-h)-h \le F'(x,y) \le F(x+h,y+h)+h$ whatever be $(x,y) \in \mathbb{R}^2$, where F and F' are distribution functions on $\mathbb{R}^2$. For a given distribution function $F(x,y)$ on $\mathbb{R}^2$ with marginal distributions $F_1(x)$ and $F_2(y)$, we can use d_2 to define a number $\rho_2(F)$ to measure the degree of independence of random variables with joint distribution F by letting $\rho_2(F) = d(F(x,y), F_1(x) \cdot F_2(y))$. Similarly, we can define $\rho_1(F)$ to measure the degree of diagonal independence of random variables with joint distribution F by letting $\rho_1(F)$ be the infimum of all those $h>0$ for which $F_1(x-h) \cdot F_2(x-h) - h \le F(x,x) \le F_1(x+h) \cdot F_2(x+h) + h$ whatever be $x \in \mathbb{R}$. Note that the above metrics d and d_2 are also defined for internal distribution functions induced by internally measurable functions, and so are ρ_1 and ρ_2. The following technical lemma is needed in the proof of Proposition 9.4.

Lemma 9.3. *Let f and g be internally measurable functions from some internal probability spaces (possibly different) to* $^*\mathbb{R}$ *with internal distribution functions F and G respectively. Assume that their standard parts* $^\circ f$ *and* $^\circ g$ *exist. Let* F', G' *be the distribution functions of* $^\circ f$ *and* $^\circ g$ *respectively. Then* $d(F,G) \simeq d(F',G')$.

Proof: For any real number $h > d(F',G')$, choose $\delta \in \mathbb{R}$ such that $h > \delta > d(F',G')$. Then for any finite $x \in {}^*\mathbb{R}$,

$$F(x) \le F'(^\circ x) \le G'(^\circ x + \delta) + \delta \le G(x+h) + \delta < G(x+h) + h,$$

and similarly

$$F(x) \ge F'(^\circ x - h + \delta) \ge G'(^\circ x - h) - \delta$$
$$\ge G(x-h) - \delta > G(x-h) - h.$$

Since $^\circ f(\cdot)$ exists almost everywhere, for $x \in {}^*\mathbb{R}_\infty$, we have $F(x) \simeq 1$ if $x>0$, and $F(x) \simeq 0$ if $x<0$; then, it is certainly true that $G(x-h)-h \le F(x) \le G(x+h)+h$. Hence, $d(F,G) \le h$. Since h is an arbitrary number satisfying $h >$

$d(F',G')$, we must have $^\circ d(F,G) \leq d(F',G')$. The same proof leads to $^\circ d(F,G) \geq d(F',G')$. $\square$

In the above lemma, if the internal random variables are taking values in $^*\mathbb{R}^2$, we have a similar result. The following proposition shows that for a triangular array of random variables, asymptotic pairwise independence is usually necessary if one needs an asymptotic version of the law of large numbers in distribution. It also reveals the equivalence of asymptotic pairwise diagonal independence and asymptotic pairwise independence. We still work with a fixed triangular array of random variables f, which are assumed to be uniformly tight, that is,

$$\lim_{m \to \infty} \sup_{1 \leq n \leq \infty} (\lambda_n \otimes P)(|f_n| > m) = 0.$$

Proposition 9.4. *Let A_n be a subset of T_n and F^{A_n} be the distribution function of $f_n^{A_n}$ on $(A_n \times \Omega, \mathscr{T}_n^{A_n} \otimes \mathscr{A}, \lambda_n^{A_n} \otimes P)$. Here, $f_n^{A_n}$ is the restriction of f_n to $A_n \times \Omega, T_n^{A_n}$ is the power set of A_n, and $\lambda_n^{A_n}$ is the probability measure on A_n rescaled from λ_n. For any t_n^1, $t_n^2 \in T_n$ and $\omega_n \in \Omega$, let $F_{t_n^1}, F_{t_n^2}, F_{t_n^1 t_n^2}$ and $F_\omega^{A_n}$ be the distribution functions of the random variables $f_n(t_n^1, \cdot)$, $f_n(t_n^2, \cdot)$, $(f_n(t_n^1, \cdot), f_n(t_n^2, \cdot))$ on (Ω, P), and $f_n^{A_n}(\cdot, \omega)$ on $(A_n, \mathscr{T}_n^{A_n}, \lambda_n^{A_n})$ respectively. For any $\varepsilon > 0$ and $n \geq 1$, define*

$$T_n^1(\varepsilon) = \left\{ (t_n^1, t_n^2) \in T_n \times T_n : \rho_1(F_{t_n^1 t_n^2}) > \varepsilon \right\}$$
$$T_n^2(\varepsilon) = \left\{ (t_n^1, t_n^2) \in T_n \times T_n : \rho_2(F_{t_n^1 t_n^2}) > \varepsilon \right\}.$$
$$\mathscr{D}^{A_n}(\varepsilon) = \left\{ \omega \in \Omega : \mathrm{d}(F_\omega^{A_n}, F^{A_n}) > \varepsilon \right\}$$

Then the following are equivalent:

(1) $\lim_{n \to \infty} (\lambda_n \otimes \lambda_n)(T_n^1(\varepsilon)) = 0$ for any $\varepsilon > 0$;

(2) There is $\delta \in (0, 1)$ such that for any $A_n \subseteq T_n$, if $\lambda_n(A_n) > 1 - \delta$, then for any $\varepsilon > 0$, $\lim_{n \to \infty} P(\mathscr{D}^{A_n}(\varepsilon)) = 0$;

(3) $\lim_{n \to \infty} (\lambda_n \otimes \lambda_n)(T_n^2(\varepsilon)) = 0$ for any $\varepsilon > 0$.

Proof: We extend the triangular array of random variables and the probability spaces as in the proof of Proposition 9.1. The uniform tightness of the processes f_n implies that for each $n \in \mathbb{N}_\infty$, the standard part of the $f_n(t_n, \omega)$ exists for almost all $(t_n, \omega) \in T_n \times \Omega$.

We first consider (1) $\Rightarrow$ (2). (1) implies that for any $\varepsilon \in \mathbb{R}^+$, $(\lambda_n \otimes \lambda_n)(T_n^1(\varepsilon)) \simeq 0$ for all $n \in {}^*\mathbb{N}_\infty$. As in the proof of Proposition 9.2, fix $n \in {}^*\mathbb{N}_\infty$ and omit the sub-index n in rest of this paragraph. By spillover, we can obtain that $L(\lambda \otimes$

$\lambda)(T^1(h/2)) = 0$ for some positive infinitesimal h. Thus for $(t^1, t^2) \notin T^1(h/2), \rho_1(F_{t^1 t^2}) \leq h/2$, and hence for any $x \in {}^*\mathbb{R}$,

$$\begin{aligned} &{}^*P\left((f_{t^1})^{-1}\right)\left({}^*(-\infty, x-h]\right){}^*P\left((f_{t^2})^{-1}\left({}^*(-\infty, x-h]\right)\right) - h \\ &\leq {}^*P\left((f_{t^1})^{-1}\left({}^*(-\infty, x]\right) \cap (f_{t^2})^{-1}\left({}^*(-\infty, x]\right)\right) \\ &\leq {}^*P\left((f_{t^1})^{-1}\left({}^*(-\infty, x+h]\right)\right){}^*P\left((f_{t^2})^{-1}\left({}^*(-\infty, x+h]\right)\right) + h. \end{aligned}$$

Therefore, for any $x \in \mathbb{R}$,

$$\begin{aligned} &L({}^*P)\left(({}^\circ f_{t^1})^{-1}((-\infty, x]) \cap ({}^\circ f_{t^2})^{-1}((-\infty, x])\right) \\ &= \lim_{m\to\infty} L({}^*P)\left((f_{t^1})^{-1}\left({}^*(-\infty, x+\frac{1}{m}]\right) \cap (f_{t^2})^{-1}\left({}^*(-\infty, x+\frac{1}{m}]\right)\right) \\ &\leq \lim_{m\to\infty} L({}^*P)\left((f_{t^1})^{-1}\left({}^*(-\infty, x+\frac{1}{m}+h]\right)\right) \\ &\quad \times L({}^*P)\left((f_{t^2})^{-1}\left({}^*(-\infty, x+\frac{1}{m}+h]\right)\right) \\ &= L({}^*P)\left(({}^\circ f_{t^1})^{-1}((-\infty, x])\right) L({}^*P)\left(({}^\circ f_{t^2})^{-1}((-\infty, x])\right). \end{aligned}$$

Similarly, we can also obtain

$$\begin{aligned} &L({}^*P)\left(({}^\circ f_{t^1})^{-1}((-\infty, x]) \cap ({}^\circ f_{t^2})^{-1}((-\infty, x])\right) \\ &= \lim_{m\to\infty} L({}^*P)\left((f_{t^1})^{-1}\left({}^*(-\infty, x+\frac{1}{m}]\right) \cap (f_{t^2})^{-1}\left({}^*(-\infty, x+\frac{1}{m}]\right)\right) \\ &\geq \lim_{m\to\infty} L({}^*P)\left((f_{t^1})^{-1}\left({}^*(-\infty, x+\frac{1}{m}-h]\right)\right) \\ &\quad \times L({}^*P)\left((f_{t^2})^{-1}\left({}^*(-\infty, x+\frac{1}{m}-h]\right)\right) \\ &= L({}^*P)\left(({}^\circ f_{t^1})^{-1}((-\infty, x])\right) L({}^*P)\left(({}^\circ f_{t^2})^{-1}((-\infty, x])\right). \end{aligned}$$

Hence, for any $x \in \mathbb{R}$,

$$\begin{aligned} &L({}^*P)\left(({}^\circ f_{t^1})^{-1}((-\infty, x]) \cap ({}^\circ f_{t^2})^{-1}((-\infty, x])\right) \\ &= L({}^*P)\left(({}^\circ f_{t^1})^{-1}((-\infty, x])\right) L({}^*P)\left(({}^\circ f_{t^2})^{-1}((-\infty, x])\right) \end{aligned}$$

Whence, ${}^\circ f_{t^1}$ and ${}^\circ f_{t^2}$ are diagonally independent.

For any fixed $A \in \mathscr{T}$ with $\lambda(A) > 1/2$, we take $\delta = 1/2$, the same proof shows that ${}^\circ f^A_{t^1}$ and ${}^\circ f^A_{t^2}$ are diagonally independent. By Theorem 3.10, the distribution function of ${}^\circ f^A_\omega$ is equal to the distribution function of ${}^\circ f^A$ for

$L({}^*P)$-almost all $\omega \in {}^*\Omega$. Lemma 9.3 implies that $d(F_\omega^A, F^A) \simeq 0$ for almost all $\omega \in {}^*\Omega$.

Now, we resume the index n and also fix an $\varepsilon \in \mathbb{R}^+$. The previous paragraph shows that ${}^*P(\mathscr{D}^{A_n}(\varepsilon)) \simeq 0$ for any $n \in {}^*\mathbb{N}_\infty$. Hence, $\lim_{n \to \infty} P(\mathscr{D}^{A_n}(\varepsilon)) = 0$, for all $\lambda_n(A_n) > 1/2$, and (2) follows.

Next, we prove that (2) ⇒ (3). Fix any $n \in {}^*\mathbb{N}_\infty$ and omit the sub-index n in this paragraph. Then for any fixed $A \in \mathscr{T}$ with $L(\lambda)(A) > 1 - \delta/2$, by spillover, (2) implies that ${}^*P(\mathscr{D}^A(\varepsilon)) \simeq 0$ for some positive infinitesimal ε. Hence, for almost all $\omega \in \Omega$, $d(F_\omega^A, F^A) \simeq 0$. As remarked in the paragraph following Lemma 9.3, we can obtain that for almost all $\omega \in {}^*\Omega$, the distribution function of the random variable $({}^\circ f)_\omega^A$ on $(A, L(\lambda^A))$ is equal to the distribution function of the random variable $({}^\circ f)^A$ on $(A \times {}^*\Omega, L(\lambda^A \otimes {}^*P))$. Therefore the process ${}^\circ f$ satisfies the consistency law in distribution. By Theorem 7.6, the random variables $({}^\circ f)_t$ are almost surely pairwise independent. Essentially the same computation as in the proof of (1) ⇒ (2) shows that $\rho_2(F_{t^1 t^2}) \simeq 0$ for $L(\lambda \otimes \lambda)$-almost all $(t^1, t^2) \in T \times T$.

Now, we resume the index n and also fix an $\varepsilon \in \mathbb{R}^+$ as before. The above paragraph shows that $(\lambda_n \otimes \lambda_n)(T_n^2(\varepsilon)) \simeq 0$ for any $n \in {}^*\mathbb{N}_\infty$, and therefore $\lim_{n \to \infty} (\lambda_n \otimes \lambda_n)(T_n^2(\varepsilon)) = 0$. Hence, (3) follows.

For any $\varepsilon \in \mathbb{R}^+$, it is clear that $T_n^1(\varepsilon) \subseteq T_n^2(\varepsilon)$, and hence (3) ⇒ (1). □

In Section 7.4, many other equivalent conditions for almost sure pairwise independence are presented. Asymptotic interpretations for them can be given by the standard procedure shown above. We simply note that in general, exact results for processes on Loeb product spaces correspond to asymptotic results for triangular arrays of random variables. For example, although Γ-uncorrelatedness and pairwise semi-independence have no relevance with each other (see Example 7.2), and both are strictly weaker than pairwise independence, the asymptotic versions of all three notions are, however, still equivalent. We shall leave such translations to those interested readers.

10. Concluding remarks

Though the transfer of a standard result to a non-standard model cannot be regarded as a new result, the interplay of internal and external entities is, however, still very useful in some situations for the discovery of completely standard results. In fact, it was pointed out by Henson and Keisler (1986) that the power of non-standard analysis comes from the saturation principle and the use of external sets. Non-standard 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 non-standard results in the standard models. Section 9 already illustrated how the measure–theoretic framework developed in earlier sections can be used to obtain large finite results for triangular arrays of random variables.

From another perspective, it might be possible to obtain the type of the results in Section 9 by working directly with the triangular arrays of random variables. However, it is the interplay between the external constructions which provide a natural framework for the discovery of the results in the limiting case and the internal structures which allow us to translate the derived internal results into the asymptotic case that gives the strength for the discovery of such asymptotic results systematically. From a purely mathematical point of view, the results presented for the limiting case, i.e., on Loeb product spaces, are more valuable, since they are genuine properties about the limiting case itself. Translation to the asymptotic case is often more cumbersome and less natural, just like one translates the equality $\frac{de^x}{dx} = e^x$ into some asymptotic property involving the sequence $\{(1+\frac{1}{n})^n\}_{n=1}^{\infty}$.

Note that 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. As noted by Anderson (1976), Keisler (1984) and Hoover and Perkins (1983), such a process can be lifted to a process which is measurable with respect to the relevant product space of Loeb spaces (for important applications of such a procedure, see Albeverio et al., 1986; Keisler, 1984 and Perkins, 1981). This means that the classical theory is included in the theory of processes which are measurable with respect to the product Loeb spaces, i.e., with respect to the $\mathscr{U}$. 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 non-trivial singular parts with respect to the $\mathscr{U}$. It will be seen in Appendix A that there is indeed no hope of developing a similar theory for processes with a Lebesgue interval parameter set.

Finally, note that 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.

Acknowledgements

The author is especially grateful to Robert Anderson, Truman Bewley, Donald Brown, John Geanakoplos, Ward Henson, Ali Khan, Peter Loeb, Heracles Polemarchakis, Qiman Shao, and Martin Shubik for useful conversations and help. The author would like to respectfully dedicate this paper to his teacher and friend Peter Loeb on his sixtieth birthday.

Appendix A

Note that many of the results in Sections 3–6 are still valid (except Theorem 6.2), if we work with arbitrary probability spaces with suitable properties. However, we may usually obtain vacuous results. To illustrate, we consider the unit Lebesgue interval $(T,\mathcal{T},\lambda)$. Let $(\Omega,\mathcal{A},P)$ be a standard probability space and $(T\times\Omega,\ \mathcal{T}\otimes\mathcal{A},\ \lambda\otimes P)$ the associated product probability space. Note that the same notations were used to denote internal probability spaces in the main body of the paper. Let $(T\times\Omega,\ \mathcal{B}_2,\ \mu_2)$ be an extension of the product space $(T\times\Omega,\ \mathcal{T}\otimes\mathcal{A},\ \lambda\otimes P)$ such that the Fubini type result still holds for $\mathcal{B}_2$-measurable functions. Then, by using the proof in Remark 3.3, one can prove an analog of Theorem 3.1 as well as Proposition 3.5 in this setting by the Fubini property. We formulate the following lemma without proof.

Lemma A.1. *For a given real-valued integrable process f on $(T\times\Omega,\ \mathcal{B}_2,\ \mu_2)$, if the conditional expectation $E(f|_{\mathcal{T}\otimes\mathcal{A}})$ of f with respect to $\mathcal{T}\otimes\mathcal{A}$ is essentially a function h on T, i.e., a function independent of particular samples, then for P-almost all $\omega\in\Omega$, the following identity*

$$\int_T \phi(t) f_\omega(t)\mathrm{d}\lambda(t) = \int_T \phi(t) h(t)\mathrm{d}\lambda(t)$$

holds for all bounded $\mathcal{T}$-measurable functions ϕ. Moreover, $h(t)=\int_\Omega f_t(\omega)\mathrm{d}P$ for λ-almost all $t\in T$.

Since the Fubini property implies that f_ω is measurable for P-almost all $\omega\in\Omega$, one can obtain the following corollary by using the uniqueness of Radon–Nikodym derivatives.

Corollary A.2. *Let f be a real-valued integrable process f on $(T\times\Omega,\ \mathcal{B}_2,\ \mu_2)$. If $E(f|\mathcal{T}\otimes\mathcal{A})$ is a function h on T, then for P-almost all $\omega\in\Omega$, $f_\omega=h$, and it follows from the Fubini property that for λ-almost all $t\in T$, f_t is essentially the constant function with value h(t).*

Thus $\mathcal{B}_2$-measurable functions are still essentially $\mathcal{T}\otimes\mathcal{A}$-measurable, i.e., $(T\times\Omega,\ \mathcal{T}\otimes\mathcal{A},\ \lambda\otimes P)$ does not admit non-trivial extension such that the Fubini property is preserved.

To consider an analog of Theorems 3.7 and 3.8, let $(T_1\times T_2\times\Omega,\ \mathcal{B}_3,\ \mu_3)$ be an extension of the product space of $(T_1,\ \mathcal{T}_1,\ \lambda_1)$ with $(T_2\times\Omega,\ \mathcal{T}_2\otimes\mathcal{A},\ \lambda_2\otimes P)$

such that the Fubini property holds for the different combinations of marginal measures, where $(T_1, \mathscr{T}_1, \lambda_1)$ and $(T_2, \mathscr{T}_2, \lambda_2)$ are copies of $(T, \mathscr{T}, \lambda)$. We also assume that the marginal measure spaces on $T_1 \times \Omega$ and $T_2 \times \Omega$ are the same measure space $(T \times \Omega, \mathscr{B}_2, \mu_2)$. Let $(T \times T, \mathscr{T}_2, \lambda_2)$ be the marginal on $T \times T$. The following lemma is an analog of Theorem 3.8. For simplicity, we only consider bounded processes.

Lemma A.3. *Let f be a bounded real-valued process on $(T \times \Omega, \mathscr{B}_2, \mu_2)$. Assume that the random variables f_t are almost surely uncorrelated, i.e., for λ_2-almost all $(t_1, t_2) \in T \times T$, f_{t_1} and f_{t_2} are uncorrelated. Then for P-almost all $\omega \in \Omega$, $\int_T f_\omega(t) d\lambda(t) = \int\int_{T \times \Omega} f d\mu_2$.*

By an analog of Theorems 3.17 and 3.21 together with Corollary A.2, we can obtain the following corollary.

Corollary A.4. *Let f be a bounded real-valued process on $(T \times \Omega, \mathscr{B}_2, \mu_2)$. If the random variables f_t are almost surely uncorrelated, then for λ-almost all $t \in T$, f_t is essentially a constant function.*

Thus the space $(T \times \Omega, \mathscr{B}_2, \mu_2)$ only supports trivial processes with uncorrelated random variables. This means that no matter how far we extend the measure on the sample space in the example considered by Judd (1985), we still cannot obtain non-trivial processes with the desired properties. So there is no way to develop a similar theory, as presented in the main body of this paper, for processes with a Lebesgue interval parameter set. In terms of the these facts, Theorem 6.2 is very important in the sense that we are not working on an empty theory.

Let $\mathscr{F}(\Omega, X)$ be the space of all the equivalence classes of $\mathscr{A}$-measurable functions from Ω to a compact metric space X with metric d. For any ϕ and ψ in $\mathscr{F}(\Omega, X)$, let

$$\rho(\phi,\psi) = \inf\{\delta \in \mathbb{R}^+ : P(\{\omega : d(\phi(\omega),\psi(\omega)) \le \delta\}) \ge 1 - \delta\}.$$

Then ρ defines a metric on $\mathscr{F}(\Omega, X)$ which induces the topology of convergence in measure. One can also view the non-measurability result for the iid case via the type of Lusin's theorem given below in Lemma A.5. To be precise, if we are given a non-trivial iid process f from $T \times \Omega$ to X, then Lemma A.5 implies that f cannot be measurable with respect to $\mathscr{T} \otimes \mathscr{A}$, since it can be checked that $\rho(f_{t_1}, f_{t_2})$ is equal to a positive constant for all $t_1 \neq t_2$. The result can be proven by modifying the proof of its classical counterpart.

Lemma A.5. *Let f be a process from $(T \times \Omega, \mathscr{T} \otimes \mathscr{A}, \lambda \otimes P)$ to a compact metric space X with a metric d. Define a mapping F from T to $\mathscr{F}(\Omega, X)$ by letting $F(t) = f_t$. Then for any $\varepsilon > 0$, there is a compact subset K_ε of the unit interval T such that $\lambda(K_\varepsilon) > 1 - \varepsilon$ and the restriction $F|_{K_\varepsilon}$ is continuous from K_ε to the metric space $(\mathscr{F}(\Omega, X), \rho)$.*

References

Albeverio, S., Fenstad, J.E., Hoegh-Krohn, R., Lindstrom, T.L., 1986. Nonstandard Methods in Stochastic Analysis and Mathematical Physics. Academic Press, Orlando, FL.

Anderson, R.M., 1976. A nonstandard representation for Brownian motion and Ito integration. Isr. J. Math. 25, 15–46.

Anderson, R.M., 1991. Non-standard analysis with applications to economics. In: Hildenbrand, W., Sonnenschein, H. (Eds.), Handbook of Mathematical Economics IV. North-Holland, New York.

Basilevsky, A., 1994. Statistical Factor Analysis and Related Methods. Wiley, New York.

Bertola, G., 1994. Flexibility, investment, and growth. J. Monetary Econ. 34, 215–238.

Bewley, T.F., 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.

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

Breiman, L., 1968. Probability. Addison-Wesley, Reading, MA.

Brown, D.J., Robinson, A., 1972. A limit theorem on the cores of large standard exchange economies. Proc. Natl. Acad. Sci. USA 69, 1258–1260.

Chung, K.L., 1974, Elementary Probability Theory with Stochastic Processes. Springer-Verlag, New York.

Cohn, D.J., 1980. Measure Theory. Birkhäuser, Boston.

Conway, J.B., 1985. A Course in Functional Analysis. Springer-Verlag, New York.

Cutland, N.J., 1983. Nonstandard measure theory and its applications. Bull. London Math. Soc. 15, 529–589.

Diamond, D.W., Dybvig, P.H., 1983. Bank runs, deposit insurance and liquidity. J. Political Econ. 91, 401–419.

Doob, J.L., 1937. Stochastic processes depending on a continuous parameter. Trans. Am. Math. Soc. 42, 107–140.

Doob, J.L., 1953, Stochastic Processes. Wiley, New York.

Durrett, R., 1991. Probability: theory and examples. Wadsworth, Belmont, CA.

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

Green, E.J., 1994. Individual-level randomness in a nonatomic population, mimeo. Univ. of Minnesota.

Henson, C.W., Keisler, H.J., 1986. On the strength of nonstandard analysis. J. Symbolic Logic 51, 377–386.

Heuser, H.G., 1982. Functional Analysis. Wiley, New York.

Hildenbrand, W., 1971. Random preferences and equilibrium analysis. J. Econ. Theory 3, 414–429.

Hoover, D.N., Keisler, H.J., 1984. Adapted probability distributions. Trans. Am. Math. Soc. 286, 159–201.

Hoover, D.N., Perkins, E., 1983. Nonstandard constructions of the stochastic integral and applications to stochastic differential equations, I, II. Trans. Am. Math. Soc. 275, 1–58.

Hurd, A.E., Loeb, P.A., 1985. An Introduction to Nonstandard Real Analysis. Academic Press, Orlando, FL.

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

Kadison, R.V., Ringrose, J.R., 1983. Fundamentals of the Theory of Operator Algebras I. Academic Press, New York.

Keisler, H.J., 1977. Hyperfinite model theory. In: Gandy, R.O., Hyland, J.M.E. (Eds.), Logic Colloquium 76. North-Holland, Amsterdam.

Keisler, H.J., 1984. An infinitesimal approach to stochastic analysis. Memoirs Am. Math. Soc. 48 (297).

Keisler, H.J., 1988. Infinitesimals in probability theory. In: Cutland, N.J. (Ed.), Nonstandard Analysis and its Applications. Cambridge Univ. Press, Cambridge.

Khan, M.A., Sun, Y.N., 1996a. Non-atomic games on Loeb spaces. Proc. Natl. Acad. Sci. USA 93, 15518–15521.

Khan, M.A., Sun, Y.N., 1996b. Non-cooperative games with many players, Mimeo, Johns Hopkins University.

Khan, M.A., Sun, Y.N., 1997a. On Loeb measure spaces and their significance for non-cooperative game theory. In: Alber, M., Hu, B., Rosenthal, J. (Eds.), Current and Future Directions in Applied Mathematics. Birkhäuser, pp. 183–218.

Khan, M.A., Sun, Y.N., 1997b. The capital asset pricing model and arbitrage pricing theory: a unification. Proc. Natl. Acad. Sci. USA 94, 4229–4232.

Khan, M.A., Yamazaki, A., 1981. On the cores of economies with indivisible commodities and a continuum of traders. J. Econ. Theory 24, 218–225.

Lai, T.L., 1977. Convergence rates and r-quick versions of strong law for stationary mixing sequences. Ann. Probability 5, 693–706.

Lintner, J., 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. 47, 13–37.

Loeb, P.A., 1975. Conversion from nonstandard to standard measure spaces and applications in probability theory. Trans. Am. Math. Soc. 211, 113–122.

Loeb, P.A., 1985. A nonstandard functional approach to Fubini's theorem. Proc. Am. Math. Soc. 93, 343–346.

Loéve, M., 1977a. Probability Theory I, 4th edn. Springer-Verlag, New York.

Loéve, M., 1977b. Probability Theory II, 4th edn. Springer-Verlag, New York.

Lucas, R.E., 1980. Equilibrium in a pure currency economy. Econ. Inquiry 18, 203–220.

Lucas, R.E., Prescott, E.C., 1974. Equilibrium search and unemployment. J. Econ. Theory 7, 188–209.

Mas-Collel, A., 1977. Indivisible commodities and general equilibrium theory. J. Econ. Theory 16, 443–456.

Nelson, E., 1987. Radically Elementary Probability Theory. Princeton Univ. Press, Princeton, NJ.

Noe, T.H., Rebello, M.J., 1994. The dynamics of business ethics and economic activity. Am. Econ. Rev. 84, 531–547.

Papoulis, A., 1965. Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York.

Parthasarathy, K.R., 1967. Probability Measures on Metric Spaces. Academic Press, New York.

Perkins, E., 1981. A global intrinsic characterization of local time. Ann. Probability 9, 800–817.

Philipp, W., Stout, W.F., 1975. Almost sure invariance principles for partial sums of weakly dependent random variables, Memoirs Am. Math. Soc., No. 161.

Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. Proc. Natl. Acad. Sci. USA 42, 43–47.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing. J. Econ. Theory 13, 341–360.

Royden, H.L., 1968. Real Analysis. Macmillan, New York.

Shao, Q.M., 1993. Complete convergence for α-mixing sequences. Stat. Probability Lett. 16, 279–287.

Sharpe, W., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. J. Finance 33, 885–901.

Stroyan, K.D., Bayod, J.M., 1986, Foundations of Infinitesimal Stochastic Analysis. North-Holland, New York.

Sun, Y.N., 1995. Semi-independence, multiplicative properties and the central limit theorem, National Univ. of Singapore, preliminary draft.

Sun, Y.N., 1996a. Hyperfinite law of large numbers. Bull. Symbolic Logic 2, 189–198.

Sun, Y.N., 1996b. Distributional properties of correspondences on Loeb spaces. J. Functional Analysis 139, 68–93.

Sun, Y.N., 1997a. Integration of correspondences on Loeb spaces. Trans. Am. Math. Soc. 349, 129–153.

Sun, Y.N., 1997b. The complete removal of individual uncertainty: multiple optimal choices and random exchange economies, mimeo.