数理解析研究所講究録
第1950巻 2015年 52-72

An operational characterization of the notion of probability by algorithmic randomness and its applications

Kohtaro Tadaki
Research and Development Initiative, Chuo University

Abstract. The notion of probability plays an important role in almost all areas of science. In modern mathematics, however, probability theory means nothing other than measure theory, and the operational characterization of the notion of probability is not established yet. In this paper, based on the toolkit of algorithmic randomness we present an operational characterization of the notion of probability. Algorithmic randomness, also known as algorithmic information theory, is a field of mathematics which enables us to consider the randomness of an individual infinite sequence. We use the notion of Martin-Löf randomness with respect to Bernoulli measure to present the operational characterization. As the first step of the research of this line, in this paper we only consider the case of finite probability space, i.e., the case where the sample space of the underlying probability space is finite, for simplicity. In the paper we make applications of our formalism to information theory, cryptography, and quantum mechanics.

1 Introduction

The notion of probability plays an important role in almost all areas of science. In modern mathematics, however, probability theory means nothing other than measure theory, and an operational characterization of the notion of probability is not established yet.

In the past century, however, there was a comprehensive attempt to provide such a characterization. Namely, von Mises developed a mathematical theory of repetitive events which is aimed at reformulating the theory of probability and statistics based on an operational characterization of the notion of probability [23, 24]. In a series of comprehensive works which began in 1919, von Mises developed this theory and, in particular, introduced the notion of *collective* as a mathematical idealization of a long

sequence of outcomes of experiments or observations repeated under a set of invariable conditions, such as the repeated tossing of a coin or of a pair of dice.

The collective plays a role as an operational characterization of the notion of probability, and is an infinite sequence of sample points in the sample space of a probability space. As the randomness property of the collective, von Mises assumes that all "reasonable" infinite subsequences of a collective satisfy the law of large numbers with the identical limit value, where the subsequences are selected using "acceptable selection rules." Wald [25, 26] later showed that for any countable collection of selection rules, there are sequences that are collectives in the sense of von Mises, but at the time it was unclear exactly what types of selection rules should be acceptable. There seemed to von Mises to be no canonical choice.

Later, with the development of computability theory and the introduction of generally accepted precise mathematical definitions of the notions of algorithm and computable function, Church [7] made the first explicit connection between computability theory and randomness by suggesting that a selection rule be considered acceptable if and only if it is computable. In 1939, however, Ville [22] revealed the defect of the notion of collective. Namely, he showed that for any countable collection of selection rules, there is a sequence that is random in the sense of von Mises but has properties that make it clearly nonrandom. (For the development of the theory of collectives from the point of view of the definition of randomness, see Downey and Hirschfeldt [8].)

In 1966, Martin-Löf [12] introduced the definition of random sequences, which is called *Martin-Löf randomness* nowadays, and plays a central role in the recent development of algorithmic randomness. At the same time, he introduced the notion of *Martin-Löf randomness with respect to Bernoulli measure* [12]. He then pointed out that this notion overcomes the defect of the collective in the sense of von Mises, and this can be regarded precisely as the collective which von Mises wanted to define. However, he did not develop probability theory based on Martin-Löf random sequence with respect to Bernoulli measure.

Algorithmic randomness is a field of mathematics which studies the definitions of random sequences and their property [14, 8]. However, the research on algorithmic randomness would seem only interested in the notions of randomness and their property, and not seem to have tried to develop probability theory based Martin-Löf randomness with respect to Bernoulli measure in an operational manner so far.

The subject of this paper is to make such an attempt. Namely, in this paper we present an operational characterization of the notion of probability based on Martin-Löf randomness with respect to Bernoulli measure. As the first step of the research of this line, we only consider the case of finite probability space, i.e., the case where the sample space of the underlying probability space is finite, for simplicity. The investigation of the case of general probability spaces is left to the future study. We emphasize that the Bernoulli measure which we consider in this paper is not required to be computable at all, while the measures considered in algorithmic randomness are usually computable. Thus, the results in this paper hold for any finite probability space.

In this paper, we make applications of our theory to information theory, cryptography, and quantum mechanics.

2 Preliminaries

2.1 Basic notation and definitions

We start with some notation about numbers and strings which will be used in this paper. $\#S$ is the cardinality of S for any set S. $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ is the set of natural numbers, and $\mathbb{N}^+$ is the set of positive integers. $\mathbb{Q}$ is the set of rationals, and $\mathbb{R}$ is the set of reals.

An *alphabet* is a nonempty finite set. We suppose that any alphabet which we consider in this paper has at least two elements. Let Ω be an alphabet. A *finite string over* Ω is a finite sequence of elements from the alphabet Ω. We denote by Ω^* the set of all finite strings over Ω, which contains the *empty string* denoted by λ. We denote by Ω^+ the set $\Omega - \{\lambda\}$. For any $\sigma \in \Omega^*$, $|\sigma|$ is the *length* of σ. Therefore $|\lambda| = 0$. A subset S of Ω^* is called *prefix-free* if no string in S is a prefix of another string in S. We write "r.e." instead of "recursively enumerable."

An *infinite sequence over* Ω is an infinite sequence of elements from the alphabet Ω, where the sequence is infinite to the right but finite to the left. We denote by Ω^∞ is the set of all infinite sequences over Ω.

Let $\alpha \in \Omega^\infty$. For any $n \in \mathbb{N}$, we denote by $\alpha\restriction_n \in \Omega^*$ the first n elements in the infinite sequence α and by $\alpha(n)$ the nth element in α. Thus, for example, $\alpha\restriction_4 = \alpha(1)\alpha(2)\alpha(3)\alpha(4)$, and $\alpha\restriction_0 = \lambda$. For any $S \subset \Omega^*$, the set

$$\{\alpha \in \Omega^\infty \mid \exists n \in \mathbb{N}\ \alpha\restriction_n \in S\}$$

is denoted by $[S]^\prec$. Note that (i) $[S]^\prec \subset [T]^\prec$ for every $S \subset T \subset \Omega^*$, and (ii) for every set $S \subset \Omega^*$ there exists a prefix-free set $P \subset \Omega^*$ such that $[S]^\prec = [P]^\prec$. For any $\sigma \in \Omega^*$, we denote by $[\sigma]^\prec$ the set $[\{\sigma\}]^\prec$, i.e., the set of all infinite sequences over Ω extending σ. Therefore $[\lambda]^\prec = \Omega^\infty$.

We briefly review measure theory. For the detail, see Billingsley [4]. A subset R of Ω^∞ is *open* if $R = [S]^\prec$ for some $S \subset \Omega^*$. In this paper we consider *the σ-field $\mathcal{F}$ generated by all open sets on Ω^∞*, which is defined as the intersection of all the σ-fields containing all open sets on Ω^∞. A *probability measure representation over* Ω is a function $r\colon \Omega^* \to [0, 1]$ such that (i) $r(\lambda) = 1$ and (ii) $r(\sigma) = \sum_{a \in \Omega} r(\sigma a)$ for every $\sigma \in \Omega^*$. A probability measure representation r induces the measure μ_r on the σ-field $\mathcal{F}$. In this paper, we use the following properties of the measure μ_r.

Proposition 1 (Properties of measure on Ω^∞).

(i) $\mu_r\left([P]^\prec\right) = \sum_{\sigma \in P} r(\sigma)$ *for every prefix-free set* $P \subset \Omega^*$. *Therefore* $\mu_r(\emptyset) = \mu_r\left([\emptyset]^\prec\right) = 0$ *and* $\mu_r(\Omega^\infty) = \mu_r\left([\{\lambda\}]^\prec\right) = 1$.

(ii) $\mu_r(\mathcal{C}) \leq \mu_r(\mathcal{D})$ *for every* $\mathcal{C}, \mathcal{D}$ *in the σ-field* $\mathcal{F}$ *with* $\mathcal{C} \subset \mathcal{D}$.

(iii) $\mu_r\left(\bigcup_i \mathcal{C}_i\right) = \sum_i \mu_r(\mathcal{C}_i)$ *for every sequence* $\{\mathcal{C}_i\}_{i \in \mathbb{N}}$ *in the σ-field* $\mathcal{F}$. □

A function $f\colon \mathbb{N} \to \Omega^*$ or $f\colon \mathbb{N} \to \mathbb{Q}$ is called *computable* if there exists a deterministic Turing machine which on every input $n \in \mathbb{N}$ halts and outputs $f(n)$. A computable function is also called a *total recursive function*. A real a is called *computable* if there exists a computable function $g\colon \mathbb{N} \to \mathbb{Q}$ such that $|a - g(k)| < 2^{-k}$ for all $k \in \mathbb{N}$. We say that $\alpha \in \Omega^\infty$ is *computable* if the mapping $\mathbb{N} \ni n \mapsto \alpha\restriction_n$ is a computable function, which is equivalent to that the real $0.\alpha$ in base-$\#\Omega$ notation is computable.

2.2 Algorithmic randomness

In the following we concisely review some definitions and results of algorithmic randomness [5, 6, 14, 8].

We use $\mathcal{L}$ to denote Lebesgue measure on $\{0,1\}^\infty$. Namely, $\mathcal{L} = \mu_r$ where the probability measure representation r is defined by the condition that $r(\sigma) = 2^{-|\sigma|}$ for every $\sigma \in \{0,1\}^*$. The idea in algorithmic randomness is to think of an infinite binary sequence as random if it is in no *effective null set.* An effective null set is a subset $\mathcal{S}$ of $\{0,1\}^\infty$ such that $\mathcal{L}(\mathcal{S}) = 0$ and $\mathcal{S}$ has some type of effective property. To specify an algorithmic randomness notion, one has to specify a type of effective null set, which is usually done by introducing a test concept. Failing the test is the same as being in the null set. In this manner, various randomness notions, such as 2-randomness, weak 2-randomness, Demuth randomness, Martin-Löf randomness, Schnorr randomness, Kurtz randomness, have been introduced so far, and a hierarchy of algorithmic randomness notions has been developed (see [14, 8] for the detail).

Among all randomness notions, *Martin-Löf randomness* is a central one. This is because in many respects, Martin-Löf randomness is well-behaved, in that the many properties of Martin-Löf random infinite sequences do match our intuition of what random infinite sequence should look like. Moreover, the concept of Martin-Löf randomness is robust in the sense that it admits various equivalent definitions that are all natural and intuitively meaningful, as we will see in Theorem 3. Martin-Löf randomness is defined as follows based on the notion of *Martin-Löf test.*

Definition 2 (Martin-Löf randomness, Martin-Löf [12]). *A subset $\mathcal{C}$ of $\mathbb{N}^+ \times \{0,1\}^*$ is called a* Martin-Löf test *if $\mathcal{C}$ is an r.e. set and for every $n \in \mathbb{N}^+$, $\mathcal{L}([\mathcal{C}_n]^\prec) \leq 2^{-n}$, where $\mathcal{C}_n = \{\, \sigma \mid (n,\sigma) \in \mathcal{C} \,\}$.*

For any $\alpha \in \{0,1\}^\infty$, we say that α is Martin-Löf random *if for every Martin-Löf test $\mathcal{C}$ there exists $n \in \mathbb{N}^+$ such that $\alpha \notin [\mathcal{C}_n]^\prec$.* □

Let $\mathcal{C}$ be a Martin-Löf test. Then, for each $k \in \mathbb{N}^+$, using (ii) of Proposition 1 we see that $\mathcal{L}\left(\bigcap_{n=1}^\infty [\mathcal{C}_n]^\prec\right) \leq \mathcal{L}\left([\mathcal{C}_k]^\prec\right) \leq 2^{-k}$. On letting $k \to \infty$, we have

$$\mathcal{L}\left(\bigcap_{n=1}^{\infty} [\mathcal{C}_n]^\prec\right) = 0.$$

Thus, the set $\bigcap_{n=1}^\infty [\mathcal{C}_n]^\prec$ forms an effective null set in the notion of Martin-Löf randomness. Definition 2 says that an infinite binary sequence α is Martin-Löf random if α is not in the effective null set $\bigcap_{n=1}^\infty [\mathcal{C}_n]^\prec$ for any Martin-Löf test $\mathcal{C}$.

The robustness of Martin-Löf randomness is mainly due to the fact that it admits characterizations based on the notion of program-size complexity, as shown in Theorem 3. The *program-size complexity* (or *Kolmogorov complexity*) $K(\sigma)$ of a finite binary string σ is defined as the length of the shortest binary input for a universal decoding algorithm U, called an *optimal prefix-free machine*, to output σ (see Chaitin [5] for the detail). By the definition, $K(\sigma)$ can be thought of as the randomness contained in the individual finite binary string σ.

Theorem 3 (Schnorr [16] and Chaitin [5]). *For every $\alpha \in \{0,1\}^\infty$, the following conditions are equivalent:*

(i) α *is Martin-Löf random.*

(ii) There exists $c \in \mathbb{N}$ *such that, for all* $n \in \mathbb{N}^+$, $n - c \leq K(\alpha\restriction_n)$. □

The condition (ii) means that the infinite binary sequence α is incompressible.

3 Martin-Löf randomness with respect to Bernoulli measure

In order to provide an operational characterization of the notion of probability we use a generalization of Martin-Löf randomness over Bernoulli measure.

Let Ω be an alphabet through out the rest of this paper. It plays a role of the set of all possible outcomes of experiments or observations. The *probability simplex on* Ω, denoted by $\mathbb{P}(\Omega)$, is the set of all functions $P\colon \Omega \to \mathbb{R}$ such that $P(a) \geq 0$ for every $a \in \Omega$ and $\sum_{a\in\Omega} P(a) = 1$. Bernoulli measure is given as follows.

Let $P \in \mathbb{P}(\Omega)$. Consider a function $r\colon \Omega^* \to [0,1]$ such that $r(a_1 \dots a_n) = \prod_{i=1}^n P(a_i)$ for every $n \in \mathbb{N}$ and $a_1, \dots, a_n \in \Omega$. The function r is a probability measure representation. The measure μ_r induced by r is *Bernoulli measure on* Ω^∞, denoted λ_P. Then Bernoulli measure λ_P on Ω^∞ has the following property: For every $\sigma \in \Omega^*$,

$$\lambda_P\left([\sigma]^{\prec}\right) = \prod_{a\in\Omega} P(a)^{N_a(\sigma)}, \tag{1}$$

where $N_a(\sigma)$ is the number of the occurrences of the element a in the finite string σ.[1]

Martin-Löf randomness with respect to Bernoulli measure is defined as follows. This notion was, in essence, introduced by Martin-Löf [12], as well as the notion of Martin-Löf randomness, which we describe in Definition 2.

Definition 4 (Martin-Löf randomness with respect to Bernoulli measure, Martin-Löf [12]). *Let* $P \in \mathbb{P}(\Omega)$. *A subset* $\mathcal{C}$ *of* $\mathbb{N}^+ \times \Omega^*$ *is called a* Martin-Löf P-test *if* $\mathcal{C}$ *is an r.e. set such that, for every* $n \in \mathbb{N}^+$, $\lambda_P\left([\mathcal{C}_n]^{\prec}\right) \leq 2^{-n}$, *where* $\mathcal{C}_n = \left\{\, \sigma \;\middle|\; (n,\sigma) \in \mathcal{C} \,\right\}$.

For any $\alpha \in \Omega^\infty$, *we say that* α *is* Martin-Löf P-random *if for every Martin-Löf* P*-test* $\mathcal{C}$ *there exists* $n \in \mathbb{N}^+$ *such that* $\alpha \notin [\mathcal{C}_n]^{\prec}$. □

Note that in Definition 4 we do not require that $P(a) > 0$ for all $a \in \Omega$. Therefore, $P(a_0)$ may be 0 for some $a_0 \in \Omega$. In the case where $\Omega = \{0,1\}$ and $P \in \mathbb{P}(\Omega)$ satisfies that $P(0) = P(1) = 1/2$, the Martin-Löf P-randomness results in the Martin-Löf randomness.

Since there are only countably infinitely many algorithms and every Martin-Löf P-test induces an effective null set, it is easy to show the following theorem.

Theorem 5. $\lambda_P(\mathrm{ML}_P) = 1$ *for every* $P \in \mathbb{P}(\Omega)$, *where* ML_P *is the set of all Martin-Löf* P*-random sequences.* □

[1] 0^0 is defined as 1 in the equation (1).

4 Ensemble

In this section, according to our former work [18] we give an operational characterization of the notion of probability for a finite probability space. We will identify the substance of the notion of probability for a finite probability space. For that purpose, we first review the notion of finite probability space, based on the notion of probability simplex. Let $P \in \mathbb{P}(\Omega)$. For each $A \subset \Omega$, we define $P(A)$ by

$$P(A) := \sum_{a \in A} P(a).$$

Then, P can be regarded as a *finite probability space* $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is the set of all subset of Ω. The set Ω is the *sample space*, and elements in Ω are called *sample points* or *elementary events*. A subset of Ω is called an *event*, and $P(A)$ is called the *probability* of A for every event A. In what follows, we regard each element in $\mathbb{P}(\Omega)$ as a finite probability space in this manner.

We propose to regard a Martin-Löf P-random sequence of sample points as an operational characterization of the notion of probability for a finite probability space. Thus, since the Martin-Löf P-randomness plays a central role in our formalism, in particular we call it *ensemble* for a finite probability space, as in Definition 6. The name "ensemble" comes from physics.

Definition 6 (Ensemble). *Let $P \in \mathbb{P}(\Omega)$. A Martin-Löf P-random sequence in Ω^∞ is called an* ensemble *for the finite probability space P.* □

Let $P \in \mathbb{P}(\Omega)$. Consider an infinite sequence $\alpha \in \Omega^\infty$ of outcomes which is obtained by an infinite reputation of trials described by the finite probability space P. The operational characterization of the notion of probability for the finite probability space P is thought to be completed if the property which the infinite sequence α has to satisfy is determined. We thus propose the following thesis.

Thesis 1. *Let $P \in \mathbb{P}(\Omega)$. An ensemble for P is an operational characterization of the finite probability space P.* □

Let us consider the validity of Thesis 1. In what follows we check that the notion of ensemble satisfies the necessary conditions which the notion of probability is considered to have to satisfy from our intuitive understanding of the notion of probability. Let $P_0 \in \mathbb{P}(\Omega)$, and consider an infinite sequence $\alpha_0 \in \Omega^\infty$ of outcomes which is obtained by an infinite reputation of trials described by the finite probability space P_0.

The first necessary condition which the notion of probability is considered to have to satisfy is that *the law of large numbers holds for α_0*. Theorem 7 below confirms that this certainly holds. Note that we have to check whether the law of large numbers holds for *any* Martin-Löf P-random sequence since P is not computable reals, in general. However, we can certainly prove it using the Chernoff bound as follows.

Theorem 7 (The law of large numbers). *Let $P \in \mathbb{P}(\Omega)$. For every $\alpha \in \Omega^\infty$, if α is an ensemble for P then, for every $a \in \Omega$,*

$$\lim_{n \to \infty} \frac{N_a(\alpha\restriction_n)}{n} = P(a).$$

□

In order to prove Theorem 7, we need the following theorem, Chernoff bound, which is a modification of the form given in Section 1.2.2 of Goldreich [10].

Theorem 8 (Chernoff bound). *Let P in $\mathbb{P}(\{0,1\})$. Then for each ε with $0 < \varepsilon \leq P(0)P(1)$ and each $n \in \mathbb{N}^+$, we have*

$$\lambda_P\left([S_n]^{\prec}\right) < 2e^{-\frac{\varepsilon^2}{2P(0)P(1)}n},$$

where S_n is the set of all $\sigma \in \{0,1\}^n$ such that $|N_1(\sigma)/n - P(1)| > \varepsilon$. □

of Theorem 7. Let $a \in \Omega$. We define $Q \in \mathbb{P}(\{0,1\})$ such that $Q(1) = P(a)$ and $Q(0) = 1 - P(a)$. Let β be the infinite binary sequence obtained from α by replacing all a by 1 and all symbols other than a by 0 in α. It follows from Theorem 15 below that β is Martin-Löf Q-random and $N_1(\beta\restriction_n) = N_a(\alpha\restriction_n)$ for every n.

Assume contrarily that $\lim_{n\to\infty} N_a(\alpha\restriction_n)/n \neq P(a)$. Then $\lim_{n\to\infty} N_1(\beta\restriction_n)/n \neq P(a)$ and therefore there exists $\varepsilon > 0$ such that $|N_1(\beta\restriction_n)/n - P(a)| > 2\varepsilon$ for infinitely many n. It follows from Theorem 8 that

$$\Pr\left[\left|\frac{N_1(\beta\restriction_n)}{n} - P(a)\right| > \varepsilon\right] < 2e^{-\frac{\varepsilon^2}{2P(a)(1-P(a))}n}.$$

Since p_i is not necessarily computable, we choose $r_L, r_R \in \mathbb{Q}$ with $p_i - 2\varepsilon < r_L < p_i - \varepsilon$ and $p_i + \varepsilon < r_R < p_i + 2\varepsilon$. For each $n \in \mathbb{N}^+$, let S_n be the set $\{x \in \{0,1\}^n \mid r_L < N_1(x)/n < r_R\}$ and let $T_n = \bigcup_{m=n}^{\infty} S_m$. Then $\beta \in [T_n]^{\prec}$ and

$$\lambda_Q\left([T_n]^{\prec}\right) \leq \sum_{m=n}^{\infty} 2e^{-cm} = \frac{2e^{-cn}}{1-e^{-c}},$$

where $c \in \mathbb{Q}$ with $0 < c < \varepsilon^2/(2P(a)(1-P(a)))$. Then it is easy to show that there exists a total recursive function $f\colon \mathbb{N}^+ \to \mathbb{N}^+$ such that $2e^{-cf(n)}/(1-e^{-c}) \leq 2^{-n}$. Thus, β is Martin-Löf Q-random since the set $\{(n,\sigma) \mid n \in \mathbb{N}^+ \ \&\ \sigma \in T_{f(n)}\}$ is Martin-Löf Q-test and $\beta \in [T_n]^{\prec}$ for every n. Hence we have a contradiction, and the result follows. □

The following is immediate from Theorem 7.

Corollary 9. *Let $P, Q \in \mathbb{P}(\Omega)$. If there exists $\alpha \in \Omega^\infty$ which is both an ensemble for P and an ensemble for Q, then $P = Q$.* □

The second necessary condition which the notion of probability is considered to have to satisfy is that *an elementary event with probability zero never occurs in the infinite sequence* α_0. Note that the notion of probability is more than the law of large numbers. To see this, consider the finite probability space $P \in \mathbb{P}(\{a,b\})$ such that $P(a) = 0$ and $P(b) = 1$, and consider the infinite sequence $\alpha = b,a,b,b,b,b,b,b,b,b,b,b,\ldots\ldots$. Since $\lim_{n\to\infty} N_a(\alpha\restriction_n)/n = 0 = P(a)$, the law of large numbers certainly holds for α. However, the elementary event a with probability zero has occurred in α once. This contradicts our intuition that an elementary event with probability zero never occurs. The example shows that the law of large numbers is insufficient to characterizes the notion of probability. Thus, *the notion of probability is more than the law of large numbers.*

Theorem 10 below states that an elementary event with probability zero never occurs in an ensemble, and thus shows that the notion of ensemble coincides with our intuition about the notion of probability in this respect. The result was, in essence, pointed out by Martin-Löf [12].

Theorem 10. *Let $P \in \mathbb{P}(\Omega)$, and let $a \in \Omega$. Suppose that α is an ensemble for the finite probability space P and $P(a) = 0$. Then α does not contain a.*

Proof. Assume contrarily that α contains a. Then there exists a prefix $\sigma \in \Omega^+$ of α which contains a. For each n, we define T_n as $\{\sigma\}$. Then, since $P(a) = 0$, we have $\lambda_P\left([T_n]^{\prec}\right) = 0$ for all $n \in \mathbb{N}^+$, and T is r.e., obviously. Thus, $\{T_n\}$ is Martin-Löf P-test. On the other hand, $\alpha \in [T_n]^{\prec}$ for all n, and therefore α is not Martin-Löf P-random. Hence, we have a contradiction, and the proof is completed. □

The following corollary is immediate from Theorem 10. It states that an elementary event with probability one always happens in an ensemble, and thus the notion of ensemble coincides with our intuition about the notion of probability in this respect.

Corollary 11. *Let $P \in \mathbb{P}(\Omega)$, and let $a \in \Omega$. Suppose that α is an ensemble for the finite probability space P and $P(a) = 1$. Then α consists only of a.* □

In what follows we consider the third necessary condition which the notion of probability is considered to have to satisfy. Assume that an observer A performs an infinite reputation of trials described by a finite probability space $P \in \mathbb{P}(\Omega)$, and thus is generating an infinite sequence $\alpha \in \Omega^\infty$ of outcomes of observations: $\alpha = a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, \ldots\ldots$. According to our thesis, Thesis 1, α is an ensemble for P. Consider another observer B who wants to adopt the following subsequence β of α as the outcomes of the observations: $\beta = a_2, a_3, a_5, a_7, a_{11}, a_{13}, a_{17}, \ldots\ldots$, where the observer B only takes into account the nth elements in the original sequence α such that n is a prime number. According to Thesis 1, β has to be an ensemble for P, as well. However, is this true?

Consider this problem in a general setting. Assume as before that an observer A performs an infinite reputation of trials described by a finite probability space $P \in \mathbb{P}(\Omega)$, and thus is generating an infinite sequence $\alpha \in \Omega^\infty$ of outcomes of observations: $\alpha = a_1, a_2, a_3, a_4, a_5, a_6, a_7, \ldots\ldots$. According to Thesis 1, α is an ensemble for P. Now, let $f\colon \mathbb{N}^+ \to \mathbb{N}^+$ be an injection. Consider another observer B who wants to adopt the following sequence β of α as the outcomes of the observations: $\beta = a_{f(1)}, a_{f(2)}, a_{f(3)}, a_{f(4)}, a_{f(5)}, \ldots\ldots$. According to our thesis, β has to be an ensemble for P, as well. However, is this true?

We can confirm this by restricting the ability of B, that is, by assuming that every observer can select elements from the original sequence α *only in an effective manner.* This means that the function $f\colon \mathbb{N}^+ \to \mathbb{N}^+$ has to be a computable function. Theorem 12 below shows this result.

Theorem 12 (Closure property under a computable shuffling, Tadaki [18])**.** *Let $P \in \mathbb{P}(\Omega)$, and let α be an ensemble for P. Then, for every injective function $f\colon \mathbb{N}^+ \to \mathbb{N}^+$, if f is computable, then the infinite sequence*

$$\alpha_f := \alpha(f(1))\alpha(f(2))\alpha(f(3))\alpha(f(4))\ldots\ldots$$

is an ensemble for P.

Proof. We show the contraposition. Suppose that α_f is not Martin-Löf P-random. Then there exists a Martin-Löf P-test $S \subset \mathbb{N}^+ \times \Omega^*$ such that $\alpha_f \in [S_n]^{\prec}$ for every n. For each $\sigma \in \Omega^+$, let $F(\sigma)$ be the set of all $\tau \in \Omega^+$ such that $|\tau| = \max f(\{1, 2, \ldots, |\sigma|\})$ and

$\sigma = \tau(f(1))\tau(f(2))\dots\tau(f(|\sigma|))$. We then define T to be $\{(n, F(\sigma)) \mid n \in \mathbb{N}^+ \ \&\ \sigma \in S_n\}$. Since f is an injection and $\sum_{a\in\Omega} P(a) = 1$, it is easy to see that $\lambda_P\left([F(\sigma)]^{\prec}\right) = \lambda_P\left([\sigma]^{\prec}\right)$. Therefore $\lambda_P\left([T_n]^{\prec}\right) = \lambda_P\left([S_n]^{\prec}\right) \leq 2^{-n}$. Thus, since T is r.e., we see that T is Martin-Löf P-test. On the other hand, $\alpha \in [T_n]^{\prec}$ for every n, and therefore α is not Martin-Löf P-random. This completes the proof. □

In other words, Theorem 12 states that ensembles for P are closed under a *computable shuffling.*

As the forth necessary condition which the notion of probability is considered to have to satisfy, we can consider the condition that the infinite sequence $\alpha_0 \in \Omega^\infty$ of outcomes which is obtained by an infinite reputation of trials described by the finite probability space $P_0 \in \mathbb{P}(\Omega)$ is closed under *the selection by a computable selection function,* as considered in the theory of collectives [23, 24, 25, 26, 7]. Theorem 13 below confirms that this condition certainly holds for every ensemble. Thus, ensembles for P are closed under the selection by a computable selection function. For the notion of the selection by a computable selection function and its meaning, see e.g., Downey and Hirschfeldt [8].

Theorem 13 (Closure property under the selection by a computable selection function, Tadaki [18]). *Let $P \in \mathbb{P}(\Omega)$, and let α be an ensemble for P. Let g be a computable selection function, that is, let $g\colon \Omega^* \to \{\mathrm{YES}, \mathrm{NO}\}$ be a computable function. Suppose that $g(\alpha\restriction_k)$ is defined for every $k \in \mathbb{N}$ and $\{k \in \mathbb{N} \mid g(\alpha\restriction_k) = \mathrm{YES}\}$ is an infinite set. Then the infinite sequence*

$$\alpha_f := \alpha(f(1))\alpha(f(2))\alpha(f(3))\alpha(f(4))\dots\dots\dots$$

is an ensemble for P, where the computable function $f\colon \mathbb{N}^+ \to \mathbb{N}^+$ is defined by

$$f(n) = \min\{m \in \mathbb{N}^+ \mid \#\{k \leq m \mid g(\alpha\restriction_k) = \mathrm{YES}\} = n\} + 1.$$

Proof. We show the contraposition. Suppose that α_f is not Martin-Löf P-random. Then there exists a Martin-Löf P-test $S \subset \mathbb{N}^+ \times \Omega^*$ such that $\alpha_f \in [S_n]^{\prec}$ for every n. For each $\sigma, \tau \in \Omega^+$, we say σ is selected by g from τ if $\tau\restriction_k \in \operatorname{dom} g$ for all $k = 0, 1, \dots, |\tau| - 1$ and there exists a strictly increasing function $h\colon \{0, 1, \dots, |\sigma| - 1\} \to \mathbb{N}$ such that (i) $\{k \in \{0, 1, \dots, |\tau| - 1\} \mid g(\tau\restriction_k) = \mathrm{YES}\} = h(\{0, 1, \dots, |\sigma| - 1\})$, (ii) $h(|\sigma| - 1) + 1 = |\tau|$, and (iii) $\tau(h(k) + 1) = \sigma(k + 1)$ for all $k = 0, 1, \dots, |\sigma| - 1$. For each $\sigma \in \Omega^+$, let $F(\sigma)$ be the set of all $\tau \in \Omega^*$ such that σ is selected by g from τ. We also set $F(\lambda) = \{\lambda\}$.

We show that

$$\lambda_P\left([F(\sigma)]^{\prec}\right) \leq \lambda_P\left([\sigma]^{\prec}\right) \tag{2}$$

for all $\sigma \in \Omega^*$ by the induction on $|\sigma|$. First, the inequality (2) holds for the case of $|\sigma| = 0$, obviously. For an arbitrary $n \in \mathbb{N}$, assume that (2) holds for all $\sigma \in \Omega^n$. Let $\sigma \in \Omega^{n+1}$ and let ρ be the prefix of σ of length n. Let $a = \sigma(|\sigma|)$. Note that $G(\tau) = \{\upsilon \mid \tau\upsilon a \in F(\sigma)\}$ is a prefix-free set for each $\tau \in F(\rho)$. Therefore, we have

$$\begin{aligned}
\lambda_P\left([F(\sigma)]^{\prec}\right) &= \sum_{\tau\in F(\sigma)} \lambda_P\left([\tau]^{\prec}\right) = \sum_{\tau\in F(\rho)} \sum_{\upsilon\in G(\tau)} \lambda_P\left([\tau\upsilon a]^{\prec}\right) \\
&= \sum_{\tau\in F(\rho)} \sum_{\upsilon\in G(\tau)} \lambda_P\left([\tau\upsilon]^{\prec}\right) P(a) \leq \sum_{\tau\in F(\rho)} \lambda_P\left([\tau]^{\prec}\right) P(a) \\
&\leq \lambda_P\left([\rho]^{\prec}\right) P(a) = \lambda_P\left([\sigma]^{\prec}\right),
\end{aligned}$$

where the second inequality follows from the assumption.

We then define $T \subset \mathbb{N}^+ \times \Omega^*$ by the condition that $T_n = \bigcup_{\sigma \in S_n} F(\sigma)$ for each $n \in \mathbb{N}^+$. Then, by (2), we have

$$\lambda_P\left([T_n]^{\prec}\right) \leq \sum_{\sigma \in S_n} \lambda_P\left([F(\sigma)]^{\prec}\right) \leq \sum_{\sigma \in S_n} \lambda_P\left([\sigma]^{\prec}\right) = \lambda_P\left([S_n]^{\prec}\right) \leq 2^{-n}.$$

Thus, since T is r.e., we see that T is Martin-Löf P-test. On the other hand, $\alpha \in [T_n]^{\prec}$ for every n, and therefore α is not Martin-Löf P-random. This completes the proof. □

5 Conditional probability and the independence between events

In this section, we operationally characterize the notions of conditional probability and the independence between events in a finite probability space in terms of ensembles.

Let $P \in \mathbb{P}(\Omega)$, and let $A \subset \Omega$ be an event in the finite probability space P. For each ensemble α for P, $\mathrm{C}_A(\alpha)$ is defined as the infinite binary sequence such that, for every i, its ith element $\mathrm{C}_A(\alpha)(i)$ is 1 if $\alpha(i) \in A$ and 0 otherwise. The pair (P, A) induces a finite probability space $\mathcal{C}(P, A) \in \mathbb{P}(\{0,1\})$ such that $\mathcal{C}(P, A)(1) = P(A)$ and $\mathcal{C}(P, A)(0) = 1 - P(A)$. Note that the notions of $\mathrm{C}_A(\alpha)$ and $\mathcal{C}(P, A)$ in our theory together correspond to the notion of *mixing* in the theory of collectives by von Mises [24]. We can then show the following theorem.

Theorem 14. *Let $P \in \mathbb{P}(\Omega)$, and let $A \subset \Omega$. Suppose that α is an ensemble for the finite probability space P. Then $\mathrm{C}_A(\alpha)$ is an ensemble for the finite probability space $\mathcal{C}(P, A)$.* □

In order to prove Theorem 14, it is convenient to prove the following theorem first, from which Theorem 14 follows.

Theorem 15. *Let $P \in \mathbb{P}(\Omega)$. Let α be an ensemble for P, and let a and b be distinct elements in Ω. Suppose that β is the infinite sequence in $(\Omega - \{b\})^\infty$ obtained by replacing all occurrences of b by a in α. Then β is an ensemble for Q, where $Q \in \mathbb{P}(\Omega - \{b\})$ such that $Q(d) = P(a) + P(b)$ if $d = a$ and $Q(d) = P(d)$ otherwise.*

Proof. We show the contraposition. Suppose that β is not a Martin-Löf Q-random sequence. Then there exists a Martin-Löf Q-test S such that $\beta \in [S_n]^{\prec}$ for every n. For each $\sigma \in (\Omega - \{b\})^*$, let $f(\sigma)$ be the set of all $\tau \in \Omega^*$ such that τ is obtained by replacing some occurrences of a in σ, if exists, by b. Note that if σ has exactly n occurrences of a then $\#f(\sigma) = 2^n$. We then define T to be $\{(n, f(\sigma)) \mid \sigma \in S_n\}$. Since $Q(a) = P(a) + Q(b)$, it is easy to see that $\lambda_Q\left([\sigma]^{\prec}\right) = \lambda_P\left([f(\sigma)]^{\prec}\right)$. Therefore $\lambda_P\left([T_n]^{\prec}\right) = \lambda_Q\left([S_n]^{\prec}\right) \leq 2^{-n}$. Since T is r.e., we see that T is Martin-Löf P-test. On the other hand, $\alpha \in [T_n]^{\prec}$ for every n, and therefore α is Martin-Löf P-random. This completes the proof. □

We show that the notion of conditional probability in a finite probability space can be represented by an ensemble in a natural manner. For that purpose we recall the notion of conditional probability in a finite probability space.

Let $P \in \mathbb{P}(\Omega)$, and let $B \subset \Omega$ be an event in the finite probability space P. Suppose that $P(B) > 0$. Then, for each event $A \subset \Omega$, the *conditional probability of A given B*, denoted by $P(A|B)$, is defined as $P(A \cap B)/P(B)$. This notion defines a finite probability space $P_B \in \mathbb{P}(B)$ such that $P_B(a) = P(\{a\}|B)$ for every $a \in B$.

When an infinite sequence $\alpha \in \Omega^\infty$ contains infinitely many elements from B, $\mathrm{Filtered}_B(\alpha)$ is defined as the infinite sequence in B^∞ obtained from α by eliminating all elements in $\Omega - B$ occurring in α. If α is an ensemble for the finite probability space P and $P(B) > 0$, then α contains infinitely many elements from B due to Theorem 7. Therefore, $\mathrm{Filtered}_B(\alpha)$ is defined in this case. Note that the notion of $\mathrm{Filtered}_B(\alpha)$ in our theory corresponds to the notion of *partition* in the theory of collectives by von Mises [24].

We can then show Theorem 16 below, which states that ensembles are closed under conditioning. For the proof of the theorem, see Tadaki [18].

Theorem 16 (Closure property under conditioning). *Let $P \in \mathbb{P}(\Omega)$, and let $B \subset \Omega$ be an event in the finite probability space P with $P(B) > 0$. For every ensemble α for P, $\mathrm{Filtered}_B(\alpha)$ is an ensemble for the finite probability space P_B.*

Proof. In the case of $B = \Omega$, $P_B = P$ and $\mathrm{Filtered}_B(\alpha) = \alpha$. Therefore the result is obvious. Thus, in what follows, we assume B is a proper subset of Ω.

First, we choose any one $a \in \Omega - B$ and define $Q \in \mathbb{P}(B \cup \{a\})$ by the condition that $Q(d) = \sum_{b \in \Omega - B} P(b)$ if $d = a$ and $Q(d) = P(d)$ otherwise. Let β be the infinite sequence in $(B \cup \{a\})^\infty$ obtained by replacing all occurrences of elements of $\Omega - B$ in α by a. It follows from Theorem 15 that β is Martin-Löf Q-random. In addition, note that

$$1 - Q(a) = P(B), \tag{3}$$

and therefore $Q(a) < 1$. Thus, it is sufficient to show that if $\mathrm{Filtered}_B(\alpha)$ is not Martin-Löf P_B-random then β is not Martin-Löf Q-random.

Assume that $\mathrm{Filtered}_B(\alpha)$ is not Martin-Löf P_B-random. Then there exists a Martin-Löf P_B-test S such that $\mathrm{Filtered}_B(\alpha) \in [S_n]^\prec$ for every n. For each $\sigma \in \Omega^+$, let $F(\sigma)$ be the set of all finite strings in $B \cup \{a\}$ of the form $a^{k_1}\sigma_1 a^{k_2}\sigma_2 \dots a^{k_L}\sigma_{k_L}$ for some $k_1, k_2, \dots, k_L \in \mathbb{N}$, where $\sigma = \sigma_1\sigma_2\dots\sigma_L$ with $\sigma_i \in B$. Then, by (3), we see that

$$\begin{aligned}
\lambda_Q\left([F(\sigma)]^\prec\right) &= \sum_{k_1,k_2,\dots,k_L=0}^{\infty} \lambda_Q\left(\left[a^{k_1}\sigma_1 a^{k_2}\sigma_2 \dots a^{k_L}\sigma_{k_L}\right]^\prec\right) \\
&= \sum_{k_1,k_2,\dots,k_L=0}^{\infty} \lambda_Q\left([\sigma]^\prec\right) Q(a)^{k_1} Q(a)^{k_2} \dots Q(a)^{k_L} \\
&= \lambda_Q\left([\sigma]^\prec\right) \left(\sum_{k=0}^{\infty} Q(a)^k\right)^L = \lambda_Q\left([\sigma]^\prec\right) \frac{1}{(1-Q(a))^L} \\
&= \lambda_Q\left([\sigma]^\prec\right) \frac{1}{P(B)^L} = \lambda_{P_B}\left([\sigma]^\prec\right).
\end{aligned}$$

We then define T to be $\{(n, F(\sigma)) \mid \sigma \in S_n\}$. It follows that $\lambda_Q\left([T_n]^\prec\right) = \lambda_{P_B}\left([S_n]^\prec\right) \leq 2^{-n}$. Thus, since T is r.e., we see that T is Martin-Löf P_B-test. On the other hand, $\beta \in [T_n]^\prec$ for every n, and therefore β is not Martin-Löf P-random. This completes the proof. □

As an application of Theorem 16, we can consider the Von Neumann extractor as follows.

Example 17 (Von Neumann extractor). *Consider a Bernoulli sequence in the sense of normal probability theory. Recall that the Von Neumann extractor takes successive pairs of consecutive bits from the Bernoulli sequence. If the two bits matches, no output is generated. If the bits differs, the value of the first bit is output. The Von Neumann extractor can be shown to produce a uniform binary output. For the detail, see [27].*

In our framework, the Von Neumann extractor operates as follows: Let $P \in \mathbb{P}(\{0,1\})$ *and let* α *be an ensemble for* P. *Then* α *can be regarded as an ensemble for* $Q \in \mathbb{P}(\{00,01,10,11\})$ *where* $Q(ab) = P(a)P(b)$ *for every* $a, b \in \{0,1\}$. *Consider the event* $B = \{01,10\}$. *It follows from Theorem 16 that* $\mathrm{Filtered}_B(\alpha)$ *is an ensemble for* $P_B \in \mathbb{P}(\{01,10\})$ *with* $P_B(01) = P_B(10) = 1/2$. *Namely,* α *is, in essence, Martin-Löf random. Hence, a random individual infinite sequence is certainly extracted by the Von Neumann extractor in our framework.* □

Let $P \in \mathbb{P}(\Omega)$. For any events $A, B \subset \Omega$ in the finite probability space P, we say that A and B are *independent* if $P(A \cap B) = P(A)P(B)$. In the case of $P(B) > 0$, A and B are independent if and only if $P(A|B) = P(A)$.

Theorem 18 below gives operational characterizations of the notion of the independence between two events in terms of ensembles. Let $\alpha, \beta \in \Omega^\infty$. We say that α and β are *equivalent* if there exists $P \in \mathbb{P}(\Omega)$ such that α and β are both an ensemble for P. The following theorem gives an operational characterization of the notion of the independency between two events by the notion of ensemble.

Theorem 18. *Let* $P \in \mathbb{P}(\Omega)$, *and let* $A, B \subset \Omega$ *be events in the finite probability space* P. *Suppose that* $P(B) > 0$. *Then the following conditions are equivalent to one another.*

(i) The events A *and* B *are independent.*

(ii) For every ensemble α *for the finite probability space* P, $\mathrm{C}_A(\alpha)$ *is equivalent to* $\mathrm{C}_{A\cap B}(\mathrm{Filtered}_B(\alpha))$.

(iii) There exists an ensemble α *for the finite probability space* P *such that* $\mathrm{C}_A(\alpha)$ *is equivalent to* $\mathrm{C}_{A\cap B}(\mathrm{Filtered}_B(\alpha))$.

Proof. Suppose that α is an arbitrary ensemble for the finite probability space P. Then, on the one hand, it follows from Theorem 14 that $\mathrm{C}_A(\alpha)$ is Martin-Löf $\mathcal{C}(P, A)$-random. On the other hand, it follow from $P(B) > 0$ and Theorem 16 that $\mathrm{Filtered}_B(\alpha)$ is an ensemble for the finite probability space P_B. Therefore, by Theorem 14, we see that $\mathrm{C}_{A\cap B}(\mathrm{Filtered}_B(\alpha))$ is Martin-Löf $\mathcal{C}(P_B, A)$-random.

Assume that the condition (i) holds. Then $P_B(A) = P(A)$. Therefore, for an arbitrary ensemble α for the finite probability space P, $\mathrm{C}_A(\alpha)$ and $\mathrm{C}_{A\cap B}(\mathrm{Filtered}_B(\alpha))$ are equivalent. Thus, we have the implication (i) $\Rightarrow$ (ii).

Since there exists an ensemble α for the finite probability space P by Theorem 5, the implication (ii) $\Rightarrow$ (iii) is obvious.

Finally, the implication (iii) $\Rightarrow$ (i) is shown as follows. Assume that the condition (iii) holds. Then $\mathrm{C}_A(\alpha)$ and $\mathrm{C}_{A\cap B}(\mathrm{Filtered}_B(\alpha))$ are Martin-Löf Q-random for some ensemble α for the finite probability space P and some $Q \in \mathbb{P}(\{0,1\})$. Thus, $\mathrm{C}_A(\alpha)$ is

Martin-Löf $\mathcal{C}(P, A)$-random, and $\mathrm{C}_{A\cap B}(\mathrm{Filtered}_B(\alpha))$ is Martin-Löf $\mathcal{C}(P_B, A)$-random. Using Corollary 9 we see that $\mathcal{C}(P, A) = Q = \mathcal{C}(P_B, A)$, and therefore $P(A) = P_B(A)$. This completes the proof. □

6 Independence of random variables

In this section, we operationally characterize the notion of the independence of random variables in a finite probability space in terms of ensembles.

A *random variable* on Ω is a function $X\colon \Omega \to \Omega'$ where Ω' is an alphabet. Let $X_1\colon \Omega \to \Omega_1, \ldots, X_n\colon \Omega \to \Omega_n$ be random variables on Ω. For any predicate $F(v_1, \ldots, v_n)$ with variables $v_1, \ldots, v_n$, we use $F(X_1, \ldots, X_n)$ to denote the event $\{a \in \Omega \mid F(X_1(a), \ldots, X_n(a))\}$. We say that the random variables $X_1, \ldots, X_n$ are *independent* if for every $x_1 \in \Omega_1, \ldots, x_n \in \Omega_n$ it holds that $P(X_1 = x_1 \;\&\; \ldots \;\&\; X_n = x_n) = P(X_1 = x_n) \cdots P(X_n = x_n)$.

Let $\alpha \in \Omega^\infty$, and $X\colon \Omega \to \Omega'$ be a random variable on Ω. We define $X(\alpha)$ as an infinite sequence β over Ω' such that $\beta(i) = X(\alpha(i))$ for every $i \in \mathbb{N}^+$. Using Theorem 15 we can show the following theorem.

Theorem 19 (Closure property under the mapping by a random variable). *Let $X\colon \Omega \to \Omega'$ be a random variable on Ω, and let $P \in \mathbb{P}(\Omega)$. If α is an ensemble for P then $X(\alpha)$ is an ensemble for $P' \in \mathbb{P}(\Omega')$ where $P'(x) = P(X = x)$ for every $x \in \Omega'$.* □

We introduce the notion of the *independence* of ensembles as follows. Let $\Omega_1, \ldots, \Omega_n$ be alphabets. For any $\alpha_1 \in \Omega_1^\infty, \ldots, \alpha_n \in \Omega_n^\infty$, we use $\alpha_1 \times \cdots \times \alpha_n$ to denote an infinite sequence over $\Omega_1 \times \cdots \times \Omega_n$ such that $\alpha(i) = (\alpha_1(i), \ldots, \alpha_n(i))$ for every $i \in \mathbb{N}^+$.

Definition 20 (Independence of ensembles). *Let $\Omega_1, \ldots, \Omega_n$ be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1), \ldots, P_n \in \mathbb{P}(\Omega_n)$. Let $\alpha_1, \ldots, \alpha_n$ be ensembles for $P_1, \ldots, P_n$, respectively. We say that $\alpha_1, \ldots, \alpha_n$ are* independent *if $\alpha_1 \times \cdots \times \alpha_n$ is an ensemble for $P \in \mathbb{P}(\Omega_1 \times \cdots \times \Omega_n)$ where $P(a_1, \ldots, a_n) = P_1(a_1) \cdots P_n(a_n)$ for every $a_1 \in \Omega_1, \ldots, a_n \in \Omega_n$.* □

Note that the notion of the independence of ensembles in our theory corresponds to the notion of *independence* of collectives in the theory of collectives by von Mises [24]. The following theorem gives equivalent characterizations of the notion of the independence of random variables in terms of that of ensembles.

Theorem 21. *Let $X_1\colon \Omega \to \Omega_1, \ldots, X_n\colon \Omega \to \Omega_n$ be random variables on Ω, and let $P_1 \in \mathbb{P}(\Omega_1), \ldots, P_n \in \mathbb{P}(\Omega_n)$. Then the following conditions are equivalent to one another.*

(i) The random variables $X_1, \ldots, X_n$ are independent.

(ii) For every ensemble α for the finite probability space P, the ensembles $X_1(\alpha), \ldots, X_n(\alpha)$ are independent.

(iii) There exists an ensemble α for P such that the ensembles $X_1(\alpha), \ldots, X_n(\alpha)$ are independent. □

In the rest of this section we consider the notion of *Martin-Löf P-randomness relative to an oracle.* The *relativized computation* is a generalization of normal computation. Let $\beta_1, \ldots, \beta_m$ be arbitrary infinite sequences over an alphabet. In the relativized computation, a (deterministic) Turing machine is allowed to refer to $(\beta_1, \ldots, \beta_m)$ as an oracle during the computation. Namely, in the relativized computation, a Turing machine can query (k, i) at any time and then obtains the response $\beta_k(i)$ during the computation. Such a Turing machine is called an *oracle Turing machine.* The relativized computation is more powerful than normal computation, in general.

We can define the notion of *Martin-Löf P-test relative to* $\beta_1, \ldots, \beta_m$ where the Turing machine which computes the Martin-Löf P-test is an oracle Turing machine which can refer to the sequence $\beta_1, \ldots, \beta_m$ during the computation. Using the notion of Martin-Löf P-tests relative to $\beta_1, \ldots, \beta_m$, we can define the notion of *Martin-Löf P-randomness relative to* $\beta_1, \ldots, \beta_m$ in the same manner as the second part of Definition 4. Obviously, the following holds.

Proposition 22. *Let $\beta_1, \ldots, \beta_m$ be infinite sequences over an alphabet, and let $P \in \mathbb{P}(\Omega)$. For every $\alpha \in \Omega^\infty$, if α is Martin-Löf P-random relative to $\beta_1, \ldots, \beta_m$, then α is Martin-Löf P-random.* □

The converse does not necessarily hold. In the case where α is Martin-Löf P-random, the converse means that the Martin-Löf P-randomness of α is independent of $\beta_1, \ldots, \beta_m$ in a certain sense.

For any $P \in \mathbb{P}(\Omega)$, we say that P is *computable* if $P(a)$ is a computable real for every $a \in \Omega$. The following theorem gives an equivalent characterization of the notion of the Independence of ensembles in terms of Martin-Löf P-randomness relative to an oracle. Its proof is obtained by modifying the proof of van Lambalgen's Theorem [21].

Theorem 23. *Let $P_1 \in \mathbb{P}(\Omega_1), \ldots, P_n \in \mathbb{P}(\Omega_n)$. Let $\alpha_1, \ldots, \alpha_n$ be ensembles for $P_1, \ldots, P_n$, respectively. Suppose that $P_1, \ldots, P_n$ are computable. Then the ensembles $\alpha_1, \ldots, \alpha_n$ are independent if and only if for every $k = 1, \ldots, n$ it holds that α_k is Martin-Löf P_k-random relative to $\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_n$.* □

Combining Theorem 21 with Theorem 23 we obtain the following theorem.

Theorem 24. *Let $X_1 \colon \Omega \to \Omega_1, \ldots, X_n \colon \Omega \to \Omega_n$ be random variables on Ω, and let $P \in \mathbb{P}(\Omega)$. For each $k = 1, \ldots, n$, let $P_k \in \mathbb{P}(\Omega_k)$ be a finite probability space such that $P_k(x) = P(X_k = x)$ for every $x \in \Omega_k$. Suppose that P is computable. Then the following conditions are equivalent to one another.*

(i) The random variables $X_1, \ldots, X_n$ are independent.

(ii) For every ensemble α for P and every $k = 1, \ldots, n$ it holds that $X_k(\alpha)$ is Martin-Löf P_k-random relative to $X_1(\alpha), \ldots, X_{k-1}(\alpha), X_{k+1}(\alpha), \ldots, X_n(\alpha)$.

(iii) There exists an ensemble α for P such that for every $k = 1, \ldots, n$ it holds that $X_k(\alpha)$ is Martin-Löf P_k-random relative to $X_1(\alpha), \ldots, X_{k-1}(\alpha), X_{k+1}(\alpha), \ldots, X_n(\alpha)$. □

7 Application to information theory

In this section, we make some application of our formalism to information theory. Instantaneous codes play a basic role in the noiseless source coding problem in information theory, as described in what follows.

Let Ω be an alphabet, as in the preceding sections. An *instantaneous code* C for Ω is an injective mapping from Ω to $\{0,1\}^*$ such that $C(\Omega) := \{C(a) \mid a \in \Omega\}$ is a prefix-free set. A sequence $a_1, a_2, \ldots, a_N \in \Omega$ is called a *message*. On the other hand, the finite binary string $C(a_1)C(a_2)\ldots C(a_N)$ is called the *coded message* for a message $a_1, a_2, \ldots, a_N$.

Let $P \in \mathbb{P}(\Omega)$ be a finite probability space, and let $X_1, X_2, \ldots, X_N$ be independent identically distributed random variables drawn from the probability mass function $P(a)$ with $a \in \Omega$. In the source coding problem, the probability space P is called an *information source* which emits a symbol in Ω. The objective of the noiseless source coding problem is to minimize the length of the coded message for a message $a_1, a_2, \ldots, a_N$ generated by the random variables $X_1, X_2, \ldots, X_N$ as $N \to \infty$. For that purpose, it is sufficient to consider the *average codeword length* $L_P(C)$ of an instantaneous code C for a finite probability space P defined by

$$L_P(C) := \sum_{a \in \Omega} P(a)\,|C(a)|$$

independently on the value of N. We can then show that $L_P(C) \geq H(P)$ for every instantaneous code C for Ω and every finite probability space $P \in \mathbb{P}(\Omega)$, where $H(P)$ is the *Shannon entropy* of P defined by

$$H(P) := -\sum_{a \in \Omega} P(a) \log_2 P(a).$$

Hence, *the Shannon entropy gives the data compression limit for the noiseless source coding problem based on instantaneous codes.* For this reason, it is important to consider the notion of absolutely optimality of an instantaneous code, where we say that an instantaneous code C for Ω is *absolutely optimal* for a finite probability space $P \in \mathbb{P}(\Omega)$ if $L_P(C) = H(P)$.

As an application of our formalism, we regard a "typical" infinite sequence in Ω^∞ which is a realization of the infinite sequence of the random variables $X_1, X_2, X_3, \ldots$ as an ensemble for the finite probability space P. For any $\alpha \in \Omega^\infty$ we denote by $\mathrm{Coded}_C(\alpha)$ the infinite binary sequence

$$C(\alpha(1))C(\alpha(2))C(\alpha(3))\ldots\ldots.$$

We can then show the following theorem.

Theorem 25. *Let $P \in \mathbb{P}(\Omega)$, and let C be an instantaneous code for Ω. Suppose that α is an ensemble for P. Then the following conditions are equivalent:*

(i) The instantaneous code C is absolutely optimal for the finite probability space P.

(ii) $\mathrm{Coded}_C(\alpha)$ is Martin-Löf random. □

Recall from Theorem 3 that Martin-Löf random sequences are precisely the infinite binary sequences which cannot be compressible any more. Thus, Theorem 25 rephrases in a sharp manner the basic result of the noiseless source coding problem that the Shannon entropy gives the data compression limit, in the form of our formalism.

8 Application to cryptography

In this section, we make some application of our formalism to cryptography by presenting new equivalent characterizations of the notion of perfect secrecy in terms of our formalism.

The notion of *perfect secrecy* plays a basic role in cryptography. First, we review the definition of encryption schemes to which the notion of perfect secrecy is applied.

Definition 26 (Encryption scheme). *Let $\mathcal{M}$, $\mathcal{K}$, and $\mathcal{C}$ be alphabets. An* encryption scheme *over a message space $\mathcal{M}$, a key space $\mathcal{K}$, and a ciphertext space $\mathcal{C}$ is a tuple $\Pi = (P_{\mathrm{K}}, \mathsf{Enc}, \mathsf{Dec})$ such that*

(i) $P_{\mathrm{K}} \in \mathbb{P}(\mathcal{K})$,

(ii) $\mathsf{Enc}\colon \mathcal{M} \times \mathcal{K} \to \mathcal{C}$,

(iii) $\mathsf{Dec}\colon \mathcal{C} \times \mathcal{K} \to \mathcal{M}$, *and*

(iv) $\mathsf{Dec}(\mathsf{Enc}(m,k),k) = m$ *for every* $m \in \mathcal{M}$ *and* $k \in \mathcal{K}$. □

Let $\Pi = (P_{\mathrm{K}}, \mathsf{Enc}, \mathsf{Dec})$ be as in Definition 26, and let $Q \in \mathbb{P}(\mathcal{M})$, which serves as a probability distribution over message space $\mathcal{M}$ for the encryption scheme Π. We consider a finite probability space $P_{\Pi,Q} \in \mathbb{P}(\mathcal{M} \times \mathcal{K})$ defined by the condition that $P_{\Pi,Q}(m,k) = Q(m)P_{\mathrm{K}}(k)$ for every $m \in \mathcal{M}$ and $k \in \mathcal{K}$. We then define random variables $M_{\Pi,Q}$ and $C_{\Pi,Q}$ on $\mathcal{M} \times \mathcal{K}$ by $M_{\Pi,Q}(m,k) = m$ and $C_{\Pi,Q}(m,k) = \mathsf{Enc}(m,k)$, respectively. The notion of perfect secrecy is then defined as follows.

Definition 27 (Perfect secrecy, Shannon [17]). *Let $\mathcal{M}$, $\mathcal{K}$, and $\mathcal{C}$ be alphabets. Let $\Pi = (P_{\mathrm{K}}, \mathsf{Enc}, \mathsf{Dec})$ be an encryption scheme over a message space $\mathcal{M}$, a key space $\mathcal{K}$, and a ciphertext space $\mathcal{C}$. The encryption scheme Π is* perfectly secret *if for every $Q \in \mathbb{P}(\mathcal{M})$ it holds that the random variables $M_{\Pi,Q}$ and $C_{\Pi,Q}$ are independent.* □

Using Theorems 21 and 24 we can show the following theorem, which characterizes the notion of perfect secrecy in terms of the notions of the independence of ensembles and Martin-Löf P-randomness relative to an oracle.

Theorem 28 (New equivalent characterizations of perfect secrecy). *Let $\mathcal{M}$, $\mathcal{K}$, and $\mathcal{C}$ be alphabets. Let $\Pi = (P_{\mathrm{K}}, \mathsf{Enc}, \mathsf{Dec})$ be an encryption scheme over a message space $\mathcal{M}$, a key space $\mathcal{K}$, and a ciphertext space $\mathcal{C}$. For each $Q \in \mathbb{P}(\mathcal{M})$, let $R_Q \in \mathbb{P}(\mathcal{C})$ be a finite probability space such that $R_Q(c) = P_{\Pi,Q}(C_{\Pi,Q} = c)$ for every $c \in \mathcal{C}$. Suppose that P_{K} is computable. Then the following conditions are equivalent to one another.*

(i) The encryption scheme Π is perfectly secret.

(ii) For every $Q \in \mathbb{P}(\mathcal{M})$ and every ensemble α for $P_{\Pi,Q}$, the ensembles $M_{\Pi,Q}(\alpha)$ and $C_{\Pi,Q}(\alpha)$ are independent.

(iii) *For every $Q \in \mathbb{P}(\mathcal{M})$ there exists an ensemble α for $P_{\Pi,Q}$ such that the ensembles $M_{\Pi,Q}(\alpha)$ and $C_{\Pi,Q}(\alpha)$ are independent.*

(iv) *For every computable $Q \in \mathbb{P}(\mathcal{M})$ and every ensemble α for $P_{\Pi,Q}$ it holds that $M_{\Pi,Q}(\alpha)$ is Martin-Löf Q-random relative to $C_{\Pi,Q}(\alpha)$, and $C_{\Pi,Q}(\alpha)$ is Martin-Löf R_Q-random relative to $M_{\Pi,Q}(\alpha)$.*

(v) *For every computable $Q \in \mathbb{P}(\mathcal{M})$ there exists an ensemble α for $P_{\Pi,Q}$ such that $M_{\Pi,Q}(\alpha)$ is Martin-Löf Q-random relative to $C_{\Pi,Q}(\alpha)$, and $C_{\Pi,Q}(\alpha)$ is Martin-Löf R_Q-random relative to $M_{\Pi,Q}(\alpha)$.* □

Note that the finite probability space P_{K}, which serves as a probability distribution over key space $\mathcal{K}$, is normally computable in modern cryptography.

9 Application to quantum mechanics

The notion of probability plays a crucial role in quantum mechanics. It appears in quantum mechanics as the so-called *Born rule*, i.e., *the probability interpretation of the wave function.* In modern mathematics which describes quantum mechanics, however, probability theory means nothing other than measure theory, and therefore any operational characterization of the notion of probability is still missing in quantum mechanics. In this sense, the current form of quantum mechanics is considered to be imperfect as a physical theory which must stand on operational means.

As a major application of our theory, in this section we present an alternative rule to the Born rule based on the notion of ensemble for the purpose of making quantum mechanics perfect. Namely, we use the notion to state the alternative rule to the Born rule for specifying the property of the results of quantum measurements in an operational way. As the first step of the research of this line, we only consider, for simplicity, the case where the set of all possible outcomes of a quantum measurement is finite, as in the preceding sections. We remark, however, that in our treatment the state space itself of a quantum system is allowed to have an infinite dimension, as long as the set of all possible outcomes of a quantum measurement performed over the quantum system is finite.

Recall that the Born rule of quantum mechanics is given as the following postulate:

Postulate 1 (The Born rule). *Quantum measurement is described by an* observable, *M, a Hermitian operator on the state space of the system being measured. The observable has a spectral decomposition*

$$M = \sum_{k=1}^{N} \lambda_k E_k,$$

where E_k is the projector onto the eigenspace of M with eigenvalue λ_k. The possible outcomes of the measurement correspond to the eigenvalues, λ_k, of the observable. If the state of the quantum system is $|\Psi\rangle$ immediately before the measurement then the probability *that result λ_k occurs is given by $\langle\Psi|E_k|\Psi\rangle$, and the state of the system after the measurement is*

$$\frac{E_k|\Psi\rangle}{\sqrt{\langle\Psi|E_k|\Psi\rangle}}.$$

□

Thus, the Born rule uses the notion of probability. However, the operational characterization of the notion of probability is not given in the Born rule, and therefore its relation to an specific infinite sequence of outcomes of quantum measurements which are being generated by an infinitely repeated measurements is unclear. We fix this point based on the notion of ensemble.

Let $\Omega = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ be an alphabet consisting of reals, which serves as *the set of all possible measurement outcomes.* Let us identify the form of the postulate of quantum measurements as it ought to be, from a general point of view. Consider the sequence of the outcomes of quantum measurements, which is an element of Ω^∞. All that the experimenter of quantum measurements can obtain through the measurements about quantum system is such a specific infinite sequence of outcomes of the measurements which are being generated by infinitely repeated measurements. Thus, the object about which the postulate of quantum measurements makes a statement should be the properties of an specific infinite sequence of outcomes of the measurements. Suggested by this consideration, we propose to replace the Born rule, Postulate 1, by the following postulate:

Postulate 2. *Quantum measurement is described by an* observable, *M, a Hermitian operator on the state space of the system being measured. The observable has a spectral decomposition*

$$M = \sum_{k=1}^{N} \lambda_k E_k,$$

where E_k is the projector onto the eigenspace of M with eigenvalue λ_k. The possible outcomes of the measurement is in the spectrum $\Omega = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ of M. Suppose that the measurements are repeatedly performed over identical quantum systems whose states are all $|\Psi\rangle$, and the infinite sequence $\alpha \in \Omega^\infty$ of measurement outcomes is being generated. Then α is an ensemble for a finite probability space P such that $P \in \mathbb{P}(\Omega)$ and $P(\lambda_k) = \langle\Psi|E_k|\Psi\rangle$ for every $k = 1, \ldots, N$. For each of the measurements, the state of the system immediately after the measurement is

$$\frac{E_k|\Psi\rangle}{\sqrt{\langle\Psi|E_k|\Psi\rangle}}, \tag{4}$$

where λ_k is the corresponding measurement outcome. □

As seen in Section 4, we can see that Postulate 2 is certainly a refinement of the Born rule from the point of view of our intuitive understanding of the notion of probability. For example, according to Postulate 2 we can show that the law of large numbers, i.e., the frequency interpretation, holds for the infinite sequence $\alpha \in \Omega^\infty$ in Postulate 2. On the other hand, we verify the *self-consistency* of Postulate 2 on some level, which suggests that Postulate 2 is not too strong.

So far we have only considered the case of pure states. According to Postulate 2, the result of the quantum measurements forms a Martin-Löf P-random sequence of states each of which is of the form (4). In the conventional quantum mechanics, this result is described as a mixed state. Suggested by this, we can give *a mathematical definition of the notion of a mixed state* in terms of the notion of ensemble. We can then replace the Born rule about mixed states by a rule based on algorithmic randomness.

Finally, we consider the validity of our new rules, in particular, based on *the many-worlds interpretation of quantum mechanics* (*MWI*, for short) [9]. MWI is more than just an interpretation of quantum mechanics. It aims to recover the predictions of quantum mechanics without assuming the Born rule, Postulate 1. For that purpose, MWI usually assumes that our world is 'typical' or 'random' among many coexisting worlds. However, the proposal of MWI by Everett was nearly a decade earlier than the advent of algorithmic randomness, and this assumption of 'typicality' was not rigorous. The notion of 'typicality' or 'randomness' is just the research object of algorithmic randomness. Based on a generalization of the notion of Martin-Löf P-randomness, we can introduce a postulate, called *the principle of typicality*, which is a refinement of the assumption of 'typicality' by Everett. We can then show that all of our new rules can be derived from the principle of typicality in a unified way. In particular, the principle of typicality is equivalent to Postulate 2 in the case of pure states.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Numbers 23340020 and 24540142. This work was partially done while the author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2014.

References

[1] R. B. Ash, *Information Theory*. Dover Publications, Inc., New York, 1990.

[2] L. Bienvenu, W. Merkle, and A. Nies, Solovay functions and K-triviality, Proceedings of the 28th Symposium on Theoretical Aspects of Computer Science (STACS 2011), pp.452–463, 2011.

[3] V. Brattka, J. Miller, and A. Nies, "Randomness and differentiability," preprint, 2012.

[4] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons, Inc., New York, 1995.

[5] G. J. Chaitin, "A theory of program size formally identical to information theory," *J. Assoc. Comput. Mach.*, vol. 22, pp. 329–340, 1975.

[6] G. J. Chaitin, *Algorithmic Information Theory*. Cambridge University Press, Cambridge, 1987.

[7] A. Church, "On the concept of a random sequence," *Bulletin of the American Mathematical Society*, vol. 46, pp. 130–135, 1940.

[8] R. G. Downey and D. R. Hirschfeldt, *Algorithmic Randomness and Complexity*. Springer-Verlag, New York, 2010.

[9] H. Everett, III, ""Relative State" formulation of quantum mechanics," *Rev. Mod. Phys.*, vol. 29, no. 3, pp. 454–462, 1957.

[10] O. Goldreich, *Foundations of Cryptography: Volume 1 - Basic Tools.* Cambridge University Press, New York, 2001.

[11] J. Katz and Y. Lindell, *Introduction to Modern Cryptography.* Chapman & Hall/CRC Press, 2007.

[12] P. Martin-Löf, "The definition of random sequences," *Information and Control*, vol. 9, pp. 602–619, 1966.

[13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Edition. Cambridge University Press, Cambridge, 2010.

[14] A. Nies, *Computability and Randomness.* Oxford University Press, Inc., New York, 2009.

[15] M. B. Pour-El and J. I. Richards, *Computability in Analysis and Physics.* Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1989.

[16] C.-P. Schnorr, "Process complexity and effective random tests," *J. Comput. System Sci.*, vol. 7, pp. 376–388, 1973.

[17] C. E. Shannon, "Communication theory of secrecy systems," *Bell Systems Technical Journal*, vol. 28, pp. 656–715, 1949.

[18] K. Tadaki, An operational characterization of the notion of probability by algorithmic randomness. Proceedings of the 37th Symposium on Information Theory and its Applications (SITA2014), 5.4.1, pp. 389–394, December 9-12, 2014, Unazuki, Toyama, Japan.

[19] K. Tadaki, An operational characterization of the notion of probability by algorithmic randomness and its application to cryptography. Proceedings of the 32nd Symposium on Cryptography and Information Security (SCIS2015), 2D4-3, January 20-23, 2015, Kokura, Japan.

[20] K. Tadaki, A refinement of quantum mechanics by algorithmic randomness. Presentation at Quantum Computation, Quantum Information, and the Exact Sciences (QCOMPINFO2015), January 30-31, 2015, Ludwig-Maximilians-Universität München, Munich, Germany.

[21] M. van Lambalgen, *Random Sequences.* Ph.D. dissertation, University of Amsterdam, 1987.

[22] J. Ville, "Étude Critique de la Notion de Collectif," Monographies des Probabilités. Calcul des Probabilités et ses Applications. Gauthier-Villars, Paris, 1939.

[23] R. von Mises, *Probability, Statistics and Truth.* Dover Publications, Inc., New York, 1957.

[24] R. von Mises, *Mathematical Theory of Probability and Statistics.* Academic Press Inc., New York, 1964.

[25] A. Wald, "Sur la notion de collectif dans la calcul des probabilités," *Comptes Rendus des Seances de l'Académie des Sciences*, vol. 202, pp. 180–183, 1936.

[26] A. Wald, "Die Wiederspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung," *Ergebnisse eines Mathematischen Kolloquiums*, vol. 8, pp. 38–72, 1937.

[27] Wikipedia contributors, "Randomness extractor," Wikipedia, The Free Encyclopedia, `http://en.wikipedia.org/wiki/Randomness_extractor` (accessed December 17, 2014).

Research and Development Initiative
Chuo University
1–13–27 Kasuga, Bunkyo-ku
Tokyo 112-8551
JAPAN
E-mail address: tadaki@kc.chuo-u.ac.jp

中央大学・研究開発機構　只木 孝太郎