The 37th Symposium on Information Theory
and its Applications (SITA2014)
Unazuki, Toyama, Japan, Dec. 09–12, 2014

An Operational Characterization of the Notion of Probability by Algorithmic Randomness

Kohtaro Tadaki*

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 an 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 in the case where the sample space of the underlying probability space is finite.

Keywords— probability, algorithmic randomness, operational characterization, Martin-Löf randomness, Bernoulli measure

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 [14, 15]. 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 [16, 17] 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 [13] 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 [9] 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* [9]. He then pointed out that this notion overcomes the defect of collective, 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 [10, 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 algoritmic randomness are usually computable. Thus, the results in this paper hold for any finite probaility space.

Due to the 6-page limit, we omit some proofs. A full paper which describes all the proofs and other related results is in preparation.

* Research and Development Initiative, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan. E-mail: tadaki@kc.chuo-u.ac.jp WWW: http://www2.odn.ne.jp/tadaki/

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, $\alpha{\restriction}_4 = \alpha(1)\alpha(2)\alpha(3)\alpha(4)$. 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, 10, 8].

We denote by $\mathcal{L}$ 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 [10, 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 [9])**.** *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* $n \in \mathbb{N}^+$, $\mathcal{L}\left([\mathcal{C}_n]^\prec\right) \leq 2^{-n}$, *where* $\mathcal{C}_n = \left\{\, \sigma \;\middle|\; (n, \sigma) \in \mathcal{C} \,\right\}$.

For any $\alpha \in \{0,1\}^\infty$, *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 [12] 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 probaility 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 ocurrences 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, introduded by Martin-Löf [9], 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 [9])**.** *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 $\alpha \notin [\mathcal{C}_n]^\prec$.* □

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

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.* □

4 Substance of the Notion of Probabaility: Ensemble

In this section, 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 *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 the substance of the notion of probability as a Martin-Löf P-random sequence of elementary events. Thus, we introduce the notion of *ensemble* for a finite probability space, as in Definition 6, and regard it as the substance of the notion of probability.

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

Frist, we check that the law of large numbers holds for every ensemble for a finite probability space. Since P is not necessarily computable reals, we have to check whether the law of large numbers holds for any Martin-Löf P-random sequence. However, we can certainly prove it using the Chernoff bound.

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 $a \in \Omega$,*

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

□

The following is immediate from Theorem 7.

Corollary 8. *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 $P = Q$.* □

Note that the notion of probability is more than the law of large numbers. To see this, consider the case where $P(a) = 0$ for a particular $a \in \Omega$. Then we expect that the "elementary event" a never happens in experiments or obserbations. Thus, the following result 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 [9].

Theorem 9. *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.* □

5 Robustness of the Notion of Ensemble

In this section we show a certain robustness of the notion of ensemble. Consider an ensemble α for a finite probability space $P \in \mathbb{P}(\Omega)$:

$$\alpha = a_1,\ a_2,\ a_3,\ a_4,\ a_5,\ a_6,\ \ldots\ldots\ldots .$$

Suppose that this sequence is the outcomes of certain observations repeated infinitely. Consider another observer who wants to adopt the following sequence β as the outcomes of the observations:

$$\beta = a_2,\ a_3,\ a_5,\ a_7,\ a_{11},\ a_{13},\ \ldots\ldots\ldots,$$

where this observer only considers the nth elements in the original sequence α such that n is a prime number. If the notion of ensemble is the substance of the notion of probability, then β has to be an ensemble for P, as well. We can confirm this requirement in a more general setting by assuming that every observer can select elements from the original sequence α *only in an effective manner*.

Theorem 10 (Robustness of ensemble I)**.** *Let $P \in \mathbb{P}(\Omega)$, and let α be an ensemble for P. Then, for every total recursive function $f\colon \mathbb{N}^+ \to \mathbb{N}^+$, if f is an injection, then the infinite sequence*

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

is an ensemble for P. □

Thus, the notion of ensemble is closed under a computable shuffling. We can also show that the notion of ensemble is closed under the selection by computable selection rules as in the theory of collectives [14, 15, 16, 17, 7].

Theorem 11 (Robustness of ensemble II)**.** *Let $P \in \mathbb{P}(\Omega)$, and let α be an ensemble for P. Let $g\colon \Omega^* \to \{\mathrm{YES}, \mathrm{NO}\}$ be a total recursive 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*

$$\alpha(f(1))\alpha(f(2))\alpha(f(3))\alpha(f(4))\ldots\ldots\ldots$$

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

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

□

6 Conditional Probability and Independence

In this section, we operationally characterize the notions of conditional probability and independence 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 [15]. We can then show the following theorem.

Theorem 12. *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 12, it is convenient to prove the following theorem first, from which Theorem 12 follows.

Theorem 13. *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 occurences 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 ccurrences of a in σ, if exists, by b. Note that if σ has exactly n ccurences 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.

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 $\Omega - B$ ocurring 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 [15].

Theorem 14. *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 occurences of elements of $\Omega - B$ in α by a. It follows from Theorem 13 that β is Martin-Löf Q-random. In addition, note that

$$1 - Q(a) = P(B), \tag{2}$$

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 (2), 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. □

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 $P(A|B) = P(A)$.

We give the characterization of the notion of the independency between two events by the notion of ensemble. 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 15. *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 probaability 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 12 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 14 that $\mathrm{Filtered}_B(\alpha)$ is an ensemble for the finite probability space P_B. Therefore, by Theorem 12, 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 probaability 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 probaability 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 8 we see that $\mathcal{C}(P, A) = Q = \mathcal{C}(P_B, A)$, and therefore $P(A) = P_B(A)$. This completes the proof. □

7 Application to Information Theory

In this section, we consider 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 $a_1, a_2, \dots, a_N \in \Omega$ is called a *message*. On the other hand, the finite binary string

$C(a_1)C(a_2)\dots C(a_N)$ is called the *coded message* for a message $a_1, a_2, \dots, a_N$.

Let $P \in \mathbb{P}(\Omega)$ be a finite probability space, and let $X_1, X_2, \dots, X_N$ 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 $a_1, a_2, \dots, a_N$ generated by the random variables $X_1, X_2, \dots, X_N$ 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 probaility 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 $X_1, X_2, X_3, \dots$ 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

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

We can then show the following theorem.

Theorem 16. *Let $P \in \mathbb{P}(\Omega)$, and let C be an instantaneous code for Ω. Suppose that α is an emsemble 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 16 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.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number 23340020 and by "Research and Development of the Public Key Systems for Secure Communication between Organizations", the Commissioned Research of National Institute of Information and Communications Technology (NICT). This work was completed 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] P. Martin-Löf, "The definition of random sequences," *Information and Control*, vol. 9, pp. 602–619, 1966.

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

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

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

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

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

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

[16] 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.

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