Recent progress on conditional randomness

Hayato Takahashi*
Random Data Lab.

Abstract

We review the recent progress on the definition of randomness with respect to conditional probabilities and a generalization of van Lambalgen theorem (Takahashi 2006, 2008, 2009, 2011). In addition we show a new result on the random sequences when the conditional probabilities are mutually singular, which is a generalization of Kjos Hanssen's theorem (2010). Finally we propose a definition of random sequences with respect to conditional probability and argue the validity of the definition from the Bayesian statistical point of view.

Keywords: Martin-Löf random sequences, Lambalgen theorem, conditional probability, Bayesian statistics

1 Introduction

The notion of conditional probability is one of the main idea in probability theory. In order to define conditional probability rigorously, Kolmogorov introduced measure theory into probability theory [4]. The notion of randomness is another important subject in probability and statistics.

In statistics, the set of random points is defined as the compliment of give statistical tests. In practice, data is finite and statistical test is a set of small probability, say 3%, with respect to null hypothesis. In order to discuss whether a point is random or not rigorously, we study the randomness of sequences (infinite data) and null sets as statistical tests. The random set depends on the class of statistical tests. Kolmogorov brought an idea of recursion theory into statistics and proposed the random set as the compliment of the effective null sets [5, 6, 8]. The effective null set is defined as the

*Tokyo 121-0062 Japan. hayato.takahashi@ieee.org
This work was done when the author was with Gifu University and supported by KAKENHI (24540153).

limit of recursively enumerable sets that goes to zero effectively. Many standard statistical tests are shown to be effective null set. One of the advantage of this definition is that the universal (nonparametric) character of the class of tests, which leads to universal (nonparametric) theory of statistics, for example see [7]. Another advantage of the definition is that the random set is characterized with entropy, i.e., Kolmogorov complexity. It is impossible to define randomness of finite strings rigorously, however it is possible to argue asymptotic theory of random sequences with the help of complexity theory.

Kolmogorov showed that conditional probabilities exist with probability one, however it was not known whether conditional probability exists for each given parameter. Therefore only a few research have been made about conditional randomness with very restricted conditions [16]. For survey on conditional randomness, see [1].

In this paper we review the recent progress on the definition of randomness with respect to conditional probabilities and a generalization of van Lambalgen theorem [10, 12, 11, 13]. In addition we show a new result on the random sequences when the conditional probabilities are mutually singular, which is a generalization of Kjos Hanssen's theorem [3]. There are three variations of Martin-Löf (ML) randomness for conditional probabilities, i.e., section of global ML-random set at the parameter, ML-random set defined form conditional probabilities, and ML-random set defined form conditional probabilities with oracle of the parameter. Generalized form of van Lambalgen theorem and generalized form of Kjos Hanssen's theorem show the relations of above three variants of randomness. In particular if the conditional probability is computable with oracle of the parameter then section of global ML-random set at the parameter and ML-random set defined form conditional probabilities with oracle of the parameter are equal, and in addition if conditional probabilities are mutually singular then the above three variations are equal.

Finally we propose a definition of random sequences with respect to conditional probabiities as the section of ML-random set at the parameter and argue the validity of the definition from the Bayesian statistical point of view.

2 Conditional probabilities and generalized Lambalgen theorem

Since conditional probabilities are defined through a joint probability of product space, we study the randomness of a joint probability of product space. We consider the space $\Omega := \{0,1\}^\infty$ and computable probability on $(\Omega \times \Omega, \mathcal{B}_2)$. $\mathcal{B}_2$ is the σ-algebra generated from cylinder sets $\{\Delta(x) \times \Delta(y) | x, y \in S\}$. Note that the completion of $\mathcal{B}_2$ does not change the random set and it does not matter the following argument. S is the set of the finite binary strings and $\Delta(x) := \{x\omega | \omega \in \Omega\}$, where $x\omega$ is the concatenation of x and ω. Let λ be the empty word. We write $x \sqsubseteq y$ if x is a prefix of y. $\mathbb{N}$ is the set of natural numbers and $\mathbb{Q}$ is the set of rational numbers. In order to clear the difference between strings and sequences, we use symbols such as x, y for strings S and x^∞, y^∞ for sequences Ω.

P on $(\Omega \times \Omega, \mathcal{B}_2)$ is called computable if there is a computable function $A : S \times S \times \mathbb{N} \to \mathbb{Q}$ such that $\forall x, y \in S, k \in \mathbb{N}\ |P(x,y) - A(x,y,k)| < \frac{1}{k}$. Intuitively speaking, P is computable if it is approximated with arbitrary precision with Turing machine. Computable probabilities on $(\Omega, \mathcal{B})$ are defined similar manner, where $\mathcal{B}$ is the σ-algebra generated from $\{\Delta(x) | x \in S\}$.

Let $U \subseteq S \times \mathbb{N}$. U is called test (effective null set) with respect to $(\Omega, \mathcal{B}, P)$ if

$$U \text{ is a recursively enumerable set (r.e. set)}, \qquad (1)$$
$$\forall n\ \tilde{U}_n \supseteq \tilde{U}_{n+1}, \text{ and } P(\tilde{U}_n) < 2^{-n}$$

where $U_n = \{x | (x, n) \in U\}$ and $\tilde{U}_n = \cup_{x \in U_n} \Delta(x)$. In the following we write $\tilde{A} := \cup_{x \in A} \Delta(x)$ for $A \subseteq S$. A set A is called recursively enumerable if there is a computable $f : \mathbb{N} \to A$ such that $f(\mathbb{N}) = A$.

The set of Martin-Löf random (ML-random) sequences w.r.t. P is defined as the compliment of the effective null sets w.r.t. P. We denote it by $\mathcal{R}^P$, i.e., $\mathcal{R}^P := (\cup_{U:test} \cap_n \tilde{U}_n)^c$. Note that if P is not computable, it is also called Hippocratic randomness [3]. For simplicity, throughout the paper, we call $\mathcal{R}^P$ as ML-random set and if computability assumption of P is necessary, we always state it. Tests and ML-random set $\mathcal{R}^P$ with respect to $(\Omega \times \Omega, \mathcal{B}_2, P)$ are defined in similar manner. If we consider the class of tests that is r.e. with oracle y^∞ in (1), the random set defined with the extended class of tests is called ML-random with oracle y^∞ and denote it with $\mathcal{R}^{P, y^\infty}$.

Lambalgen's theorem [15] says that a pair of sequences $(x^\infty, y^\infty) \in \Omega^2$ is ML-random w.r.t. the product of uniform measures iff x^∞ is ML-random and y^∞ is ML-random with oracle x^∞. In [16, 10, 12, 11, 13] Lambalgen's theorem is generalized for computable correlated probabilities.

Let $X = Y = \Omega$ and P be a computable probability on $X \times Y$. P_X and P_Y are marginal distributions on X and Y, respectively. In the following we write $P(x, y) := P(\Delta(x) \times \Delta(y))$ and $P(x|y) := P(\Delta(x)|\Delta(y))$ for $x, y \in S$.

Let $\mathcal{R}^P$ be the set of ML-random points and $\mathcal{R}^P_{y^\infty} := \{x^\infty \mid (x^\infty, y^\infty) \in \mathcal{R}^P\}$. In [10, 11] ML-random sequences satisfies martingale convergence theorem and from this fact we can show the existence of conditional probabilities as follows.

Theorem 1 ([10, 12, 11]) *Let P be a computable probability on $X \times Y$ and*

$$\forall x \in S,\ y^\infty \in \mathcal{R}^{P_Y}\ \ P(x|y^\infty) := \lim_{y \to y^\infty} P(x|y) \text{ if the right-hand-side exist.}$$

Then $P(\cdot|y^\infty)$ is a probability on $(\Omega, \mathcal{B})$ for each $y^\infty \in \mathcal{R}^{P_Y}$.

Let $\mathcal{R}^{P(\cdot|y^\infty), y^\infty}$ be the set of ML-random set w.r.t. $P(\cdot|y^\infty)$ with oracle y^∞.

Theorem 2 ([10, 12, 11, 13]) *Let P be a computable probability on $X \times Y$. Then*

$$\mathcal{R}^P_{y^\infty} \supseteq \mathcal{R}^{P(\cdot|y^\infty), y^\infty} \text{ for all } y^\infty \in \mathcal{R}^{P_Y}. \tag{2}$$

Fix $y^\infty \in \mathcal{R}^{P_Y}$ and suppose that $P(\cdot|y^\infty)$ is computable with oracle y^∞. Then

$$\mathcal{R}^P_{y^\infty} = \mathcal{R}^{P(\cdot|y^\infty), y^\infty}. \tag{3}$$

It is known that there are non-computable conditional probabilities [9] and in [2] Bauwens showed an example that violates the equality in (3) when the conditional probability is not computable with oracle y^∞. In [14] an example that for all y^∞, the conditional probabilities are not computable with oracle y^∞ and (3) holds.

In Vovk and V'yugin [16] (3) is shown under the condition that conditional probabilities exist for all parameters and each conditional probability is computable with oracle of parameter. Bayesian theory consists of prior and parametric model and many models satisfy the assumption in [16], e.g., Bernoulli and finite order Markov processes satisfy the model. However in our model we start with computable global P, and prior and parametric model are derived from P as marginal distribution and conditional probabilities, respectively. In particular we can argue the case that conditional probability is not computable with oracle of given parameter. It is possible that conditional probabilities and parametric model may differ at null set of parameters [10].

3 Bayesian statistics

We apply randomness theory to Bayesian statistics and show a point wise theory for Bayesian statistics. In Bayesian statistics, we suppose a probability on parameter space and it is called prior. Let X and Y be sample and parameter spaces, respectively. Let P be a probability on $X \times Y$ then marginal distributions

$$P_X = \int_Y P(\cdot|y^\infty)dP_Y$$

and P_Y are called mixture and prior distributions, respectively.

Theorem 3 ([10, 12, 11]) *Let P be a computable probability on $X \times Y$.*

$$P(\mathcal{R}^P_{y^\infty}) = 1 \textit{ if } y^\infty \in \mathcal{R}^{P_Y} \textit{ and } \mathcal{R}^P_{y^\infty} = \emptyset \textit{ else.}$$

$$\mathcal{R}^{P_X} = \cup_{y^\infty \in \mathcal{R}^{P_Y}} \mathcal{R}^P_{y^\infty}. \tag{4}$$

(4) shows that the natural properties of Bayesian mixture.

Next we study mutually singular conditional probabilities. Kjos Hanssen [3] showed that for Bernoulli model $P(\cdot|\theta)$,

$$\mathcal{R}^{P(\cdot|\theta)} = \mathcal{R}^{P(\cdot|\theta),\theta} \text{ for all } \theta. \tag{5}$$

We generalize Kjos Hanssen's theorem (5) for mutually singular conditional probabilities. Strictly speaking, prior space of Bernoulli model is $[0,1]$, however we consider Ω as prior space.

In [10, 13] equivalent conditions for mutually singular conditional probabilities are shown.

Theorem 4 ([10, 12, 13]) *Let P be a computable probability on $X \times Y$, where $X = Y = \Omega$. The following six statements are equivalent:*
(1) $P(\cdot\,|\,y) \perp P(\cdot\,|\,z)$ if $\Delta(y) \cap \Delta(z) = \emptyset$ for $y, z \in S$.
(2) $\mathcal{R}^{P(\cdot\,|\,y)} \cap \mathcal{R}^{P(\cdot\,|\,z)} = \emptyset$ if $\Delta(y) \cap \Delta(z) = \emptyset$ for $y, z \in S$.
(3) $P_{Y|X}(\cdot\,|\,x)$ converges weakly to I_{y^∞} as $x \to x^\infty$ for $(x^\infty, y^\infty) \in \mathcal{R}^P$, where I_{y^∞} is the probability that has probability of 1 at y^∞.
(4) $\mathcal{R}^P_{y^\infty} \cap \mathcal{R}^P_{z^\infty} = \emptyset$ if $y^\infty \neq z^\infty$.
(5) There exists $f : X \to Y$ such that $f(x^\infty) = y^\infty$ for $(x^\infty, y^\infty) \in \mathcal{R}^P$.
(6) There exists $f : X \to Y$ and $Y' \subseteq Y$ such that $P_Y(Y') = 1$ and $f = y^\infty$, $P(\cdot\,; y^\infty) - a.s.$ for $y^\infty \in Y'$.

The above theorem shows that consistency of posterior distribution for each pair of random sequences and their equivalent conditions. It is interesting to see that the condition (1) and (6) do not have algorithmic notion.

Generalized form of Kjos Hanssen's theorem (5) is as follows.

Theorem 5 *Let P be a computable probability on $X \times Y$, where $X = Y = \Omega$. Under one of the condition of Theorem 4, we have*

$$\mathcal{R}^P_{y^\infty} \supseteq \mathcal{R}^{P(\cdot|y^\infty)} \supseteq \mathcal{R}^{P(\cdot|y),y^\infty} \text{ for all } y^\infty \in \mathcal{R}^{P_Y}.$$

Fix $y^\infty \in \mathcal{R}^{P_Y}$ and suppose that $P(\cdot|y^\infty)$ is computable with oracle y^∞. Then

$$\mathcal{R}^P_{y^\infty} = \mathcal{R}^{P(\cdot|y^\infty)} = \mathcal{R}^{P(\cdot|y^\infty),y^\infty}.$$

Before we prove Theorem 5, we need a lemma.

Lemma 1 *Let P be a computable probability on $X \times Y$, where $X = Y = \Omega$. Fix $y^\infty \in \mathcal{R}^{P_Y}$ then*

$$\forall y \sqsubset y^\infty \ \mathcal{R}^{P(\cdot|y^\infty)} \subseteq \mathcal{R}^{P(\cdot|y)}.$$

Proof) In [11] Corollary 4.1, it is shown that there is an integer M such that

$$y^\infty \in \mathcal{R}^{P_Y} \Rightarrow \sum_n P((\tilde{U}_n)_{y^\infty}|y^\infty) < M < \infty, \tag{6}$$

where U is a test with respect to P and $(\tilde{U}_n)_{y^\infty} = \{x^\infty|(x^\infty, y^\infty) \in \tilde{U}_n\}$.

From (6), we have for all k

$$P(\sum_n I_{(\tilde{U}_n)_{y^\infty}} > k|y^\infty) < \frac{M}{k}, \tag{7}$$

where I is the characteristic function, i.e., $I_{(\tilde{U}_n)_{y^\infty}}(x^\infty) = 1$ if $x^\infty \in (\tilde{U}_n)_{y^\infty}$ else 0.

Let $y = \lambda$ then we have $P(\cdot|\lambda) = P_X$. Let U^X be a test with respect to P_X and $U^{X\times\lambda} := \{(x, \lambda, n)|(x, n) \in U^X\}$. We see that $U^{X\times\lambda}$ is a test with respect to P. Since $(\tilde{U}^{X\times\lambda}_n)_{y^\infty} = \tilde{U}^X_n$, from (7) we have

$$\forall k \ P(\sum_n I_{\tilde{U}^X_n} > k|y^\infty) < \frac{M}{k}.$$

Thus we have a test V^X with respect to $\mathcal{R}^{P(\cdot|y^\infty)}$ such that $\forall k \ \tilde{V}^X_k = \{\sum_n I_{\tilde{U}^X_n} > M2^k\}$ and the lemma proved for $y = \lambda$.

We can show the lemma for any finite prefix $y \sqsubset y^\infty$ in the similar way.

■

Proof of Theorem 5) As with the same way of Theorem 3, we have

$$\forall y\ \mathcal{R}^{P(\cdot|y)} = \cup_{y^\infty \in \mathcal{R}^{P_Y} \cap \Delta(y)} \mathcal{R}^P_{y^\infty}. \tag{8}$$

For example consider $P(\cdot|X \times \Delta(y))$ and its ML-random set and observe that $\mathcal{R}^{P(\cdot|X\times\Delta(y))} = \mathcal{R}^P \cap (X \times \Delta(y))$.

Suppose that $\mathcal{R}^P_{y^\infty} \cap \mathcal{R}^P_{z^\infty} = \emptyset$ if $y^\infty \neq z^\infty$. Then from (8), we have

$$\cap_{y \to y^\infty} \mathcal{R}^{P(\cdot|y)} = \mathcal{R}^P_{y^\infty}.$$

From Lemma 1, we have the first statement of the theorem. The latter part follows from Theorem 2 and the first statement. ■

4 Bayesian definition of random sequences

In this paper we discussed three variations of ML-randomness with respect to conditional probabilities, i.e., $\mathcal{R}^P_{y^\infty}$, $\mathcal{R}^{P(\cdot|y^\infty)}$, and $\mathcal{R}^{P(\cdot|y^\infty),y^\infty}$. These relation are shown in Theorem 2 and 5. However the only proven variants that satisfy the properties of Bayesian mixture (4) is $\mathcal{R}^P_{y^\infty}$. Therefore as in [10, 11] I propose that $\mathcal{R}^P_{y^\infty}$ as *the Bayesian definition of random sequences with respect to conditional probabilities*.

Note that $\mathcal{R}^{P(\cdot|y^\infty)}$ and $\mathcal{R}^{P(\cdot|y^\infty),y^\infty}$ are defined from local conditional probability $P(\cdot|y^\infty)$ and y^∞, and $\mathcal{R}^P_{y^\infty}$ is defined from global P and y^∞. It is interesting to see that these three sets are equal if the conditional probabilities are mutually singular and computable with oracle y^∞.

One might think that a sequence is a *good* random one if it is in the compliment of the large class of statistical tests. The class of tests of local conditional probabilities are larger than the class of tests of the section of global probability at the parameter under the conditions of Theorem 2 and 5. However from Theorem 2, 3, and 5, the properties of Bayesian mixture (4) might be violated for local conditional randomness, which may lead to a completely new idea in randomness theory.

$\mathcal{R}^P_{y^\infty}$ satisfies many probability law related to Bayesian statistics and information theory, for more details, see [11, 13].

References

[1] B. Bauwens, A. Shen, and H. Takahashi. Conditional probabilities and van Lambalgen theorem revisited. arxiv:1607.04240, 2016.

[2] Bruno Bauwens. Conditional measure and the violation of van Lambalgen's theorem for Martin-Löf randomness. http://arxiv.org/abs/1103.1529, 2015.

[3] Bjørn Kjos Hanssen. The probability distribution as a computational resource for randomness testing. *Journal of Logic and Analysis*, 2(10):1–13, 2010.

[4] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, volume 2 of *Eng. Math.* Springer Verlag, Berlin, 1933.

[5] A. N. Kolmogorov. On tables of random numbers. *Sankhyā*, 25:369—376, 1963.

[6] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Probl. Inf. Transm.*, 1(1):1–7, 1965.

[7] M. Li and P. Vitányi. *An introduction to Kolmogorov complexity and Its applications.* Springer, New York, third edition, 2008.

[8] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–609, 1966.

[9] D. M. Roy. *Computability, inference and modeling in probabilistic programming.* PhD thesis, MIT, 2011.

[10] H. Takahashi. Bayesian approach to a definition of random sequences and its applications to statistical inference. In *2006 IEEE International Symposium on Information Theory*, pages 2180–2184, July 2006.

[11] H. Takahashi. On a definition of random sequences with respect to conditional probability. *Inform. and Compt.*, 206:1375–1382, 2008.

[12] H. Takahashi. Some problems of algorithmic randomness on product space. In *The 8th Workshop on Stochastic Numerics*, volume 1620, pages 175–196. RIMS Kôkyûroku, Kyoto University, 2009.

[13] H. Takahashi. Algorithmic randomness and monotone complexity on product space. *Inform. and Compt.*, 209:183–197, 2011.

[14] H. Takahashi. Generalization of van lambalgen's theorem and blind randomness for conditional probabilities, 2014. arxiv:1310.0709v3.

[15] M. van Lambalgen. *Random sequences.* PhD thesis, Universiteit van Amsterdam, 1987.

[16] V. G. Vovk and V. V. V'yugin. On the empirical validity of the Bayesian method. *J. R. Stat. Soc. B*, 55(1):253–266, 1993.