Van Lambalgen's theorem fails for some computable measure

Bruno Bauwens[*][†]

December 22, 2015

Abstract

Van Lambalgen's theorem states that a pair (α, β) of bit sequences is Martin-Löf random if and only if α is Martin-Löf random and β is Martin-Löf random relative to α. In [Information and Computation 209.2 (2011): 183-197, Theorem 3.3], Hayato Takahashi generalized van Lambalgen's theorem for computable measures P on a product of two Cantor spaces; he showed that the equivalence holds for each β for which the conditional probability $P(\cdot|\beta)$ is computable. He asked whether this computability condition is necessary. We give a positive answer by providing a computable measure for which van Lambalgen's theorem fails. We also present a simple construction of a computable measure for which conditional measure is not computable. Such measures were first constructed by N. Ackerman, C. Freer and D. Roy in [Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science (LICS), pp. 107-116. IEEE (2011)].

Michiel van Lambalgen characterized Martin-Löf randomness of a pair of bit sequences:

Theorem 1 (van Lambalgen [6])**.** *The following are equivalent for a pair (α, β) of sequences:*

- *(α, β) is Martin-Löf random,*

- *α is Martin-Löf random and β is Martin-Löf random relative to α.*

One can replace uniform (Lebesgue) measure in the definition of Martin-Löf randomness by any other computable measure P. We call sequences that are random in this sense *P-random*. There exist two definitions of Martin-Löf randomness for a pair of sequences. The first states that (α, β) is random if the join $\alpha_1 \beta_1 \alpha_2 \beta_2 \ldots$ is random. The second definition uses the two dimensional variant of a Martin-Löf test, which is given by a family of uniformly effectively open sets $U_n \subseteq 2^{\mathbb{N}} \times 2^{\mathbb{N}}$ such that the uniform measure of U_n is at most 2^{-n}. Both approaches are equivalent.

To generalize van Lambalgen's theorem for computable measures P, the first approach seems not suitable. Why join two sequences in this specific way? What does it mean? Also, the most direct approach of replacing Martin-Löf randomness with P-randomness will make the theorem wrong for trivial reasons: There exist a computable P and a pair of sequences (α, β) such that $\alpha_1 \beta_1 \alpha_2 \beta_2 \ldots$ is P-random, while α is not P-random. Indeed, let P be the measure that concentrates all its mass on the single point $010101 \ldots$, i.e., $P(\{0101 \ldots\}) = 1$ and $P(S) = 0$ if $0101 \cdots \notin S$. The sequence $0101 \ldots$ is P-random, but $00 \ldots$ is not random.

To use the two-dimensional approach, we need to decompose the bivariate measure P into two univariate measures. It is natural to use the marginal and the conditional measure for P. In fact, such decompositions are omnipresent in probability theory, and it

nicely fits the statement of van Lambalgen's theorem, which uses in the second criterion a conditionally and an unconditionally random sequence.

We now define conditional measure. Let $2^{\mathbb{N}}$ denote Cantor space. For any string x, let $[x]$ be the (basic open) set containing all extensions of x. We say that a measure P on $2^{\mathbb{N}}$ is computable if the function that maps each string x to $P([x])$ is computable. Similar for measures P on $2^{\mathbb{N}} \times 2^{\mathbb{N}}$. Following Takahashi [10], we define for each measure P on $2^{\mathbb{N}} \times 2^{\mathbb{N}}$ and each measurable set $S \subseteq 2^{\mathbb{N}}$:

$$P_C(S|\beta) = \lim_{n \to \infty} \frac{P(S \times [\beta_1 \ldots \beta_n])}{P(2^{\mathbb{N}} \times [\beta_1 \ldots \beta_n])}.$$

Let the marginal distribution be $P_M(S) = P(2^{\mathbb{N}} \times S)$.

Remark: The definition of a conditional measure is usually given using the Radon-Nikodym theorem. In fact, this theorem defines a set of conditional measures, and each pair of such measures coincides on a set β of P_M-measure one. Using the Lebesgue differentiation theorem it can be shown that these conditional measures also coincide with $P_C(\cdot|\beta)$ for P_M-almost all β. We refer to the appendix for more details.

This specific conditional measure is especially suitable to generalize van Lambalgen's theorem: if β is P_M-random, then $P_C(\cdot|\beta)$ is defined and is a measure [10, Theorem 4.1] (see also [8, Lemma 10]). In [1, Theorem 29, p14] it is shown that for computable P, the measure P_C might not be computable. The measure that satisfies the conditions of our main result satisfies a similar property:

Corollary 2 (of the proof of Theorem 4)**.** *There exists a computable measure P on $2^{\mathbb{N}} \times 2^{\mathbb{N}}$ such that the set of β for which $P_C(\cdot|\beta)$ is not computable relative to β, has nonzero P_M-measure.*

The corollary is proven after Theorem 4. Similar examples of such measures were invented by Jason Rute [7]. In the example from [1], definitions of computability of functions and measures from computable analysis are used. They can be used on general spaces but are rather difficult to formulate. Functions that are not computable in this sense include all functions with a discontinuity. Therefore, the example in [1] is made in such a way that $P_C(S|\beta)$ is continuous in β for all measurable sets S. We present a simple variant of the construction of such a measure in Theorem 6 below. The proof of this theorem does not rely on other parts of this note.

Hayato Takahashi generalized van Lambalgen's theorem as follows:

Theorem 3 (Takahashi [11, 12])**.** *For any computable bivariate measure P and any β such that $P_C(\cdot|\beta)$ is computable relatively to β, the following are equivalent:*

- *(α, β) is P-Martin-Löf random,*

- *β is P_M-random and α is $P_C(\cdot|\beta)$-random relative to β.*

For an alternative exposition of the proof and for related results, I refer to the upcoming article [9]. One might ask whether the theorem only holds for β for which $P_C(\cdot|\beta)$ is computable relative to β? In this note we show that we can not drop this assumption, hence, van Lambalgen's theorem fails for some computable measure. To formulate the result, we need a definition of randomness relative to a non-computable measure. There exist two types of Martin-Löf tests [5]:

- A *uniform* P-Martin-Löf test is a P-Martin-Löf test that is effectively open relative to each oracle that computes P.

- A *Hippocratic* or *blind* P-Martin-Löf test is a Martin-Löf test that is effectively open without any oracle.

If P is computable, then both types of tests define the same set of random sequences. Otherwise, the second type of tests defines a weaker notion of randomness, which we will use below. We call a sequence *blind P-random* if no blind Martin-Löf test succeeds on it. Our main result is:
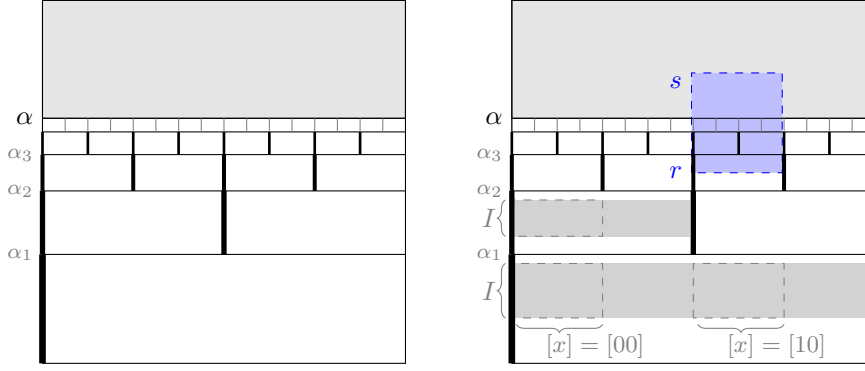
Figure 1: Left: The measure P. Right: Some values for $P(I \times [x])$. From bottom to top: For $I \subseteq [0, \alpha_1]$, $P(I \times [00]) = \mu(I)$ and $P(I \times [10]) = 0$. For $I \subseteq [\alpha_1, \alpha_2]$, we have $P(I \times [00]) = \mu(I)/2$. For $\alpha_{|x|} \leq r < s \leq 1$ we have $P([r, s] \times [x]) = \mu([r, s] \times [x])$.

Theorem 4. *There exists a bivariate computable measure P on $2^{\mathbb{N}} \times 2^{\mathbb{N}}$ for which $P_C(\cdot|\beta)$ exists for all β; moreover, there exists a pair of sequences (α, β) such that the pair is P-Martin-Löf random and α is not blind $P_C(\cdot|\beta)$-Martin-Löf random (even without oracle β).*

Definitions Let μ be the uniform measure, thus, $\mu([x]) = 2^{-|x|}$ for any string x. We also use μ for the product of two uniform measures over $2^{\mathbb{N}} \times 2^{\mathbb{N}}$. Real numbers in $[0, 1]$ that are not binary rational, are interpreted as elements of $2^{\mathbb{N}}$. For binary rational numbers α and β, we associate $[\alpha, \beta]$ with the corresponding basic open set in Cantor space (thus only containing the binary representation of α with a tail of zeros, and a tail of ones for β).

Proof. Let $\alpha_1, \alpha_2, \ldots,$ be an increasing computable sequence of binary rational numbers that converges to a Martin-Löf random real α. Such a real exists (and can be Turing complete, see e.g. [3, Theorem 4.3]). To construct the bivariate measure P, modify the uniform measure on $2^{\mathbb{N}} \times 2^{\mathbb{N}}$ as illustrated in figure 1 left: concentrate all measure in the horizontal strip $[0, \alpha_1] \times 2^{\mathbb{N}}$ uniformly in its left-most horizontal position, i.e., in $[0, \alpha_1] \times 000\ldots$; concentrate the measure in the intervals $[\alpha_1, \alpha_2] \times [0]$ and $[\alpha_1, \alpha_2] \times [1]$ uniformly in their leftmost positions, i.e., in $[\alpha_1, \alpha_2] \times 000\ldots$ and $[\alpha_1, \alpha_2] \times 1000\ldots$; and so on.[1]

Before presenting the formal definition, let us illustrate the construction of P. Consider an interval $I \subseteq [0, \alpha_1]$. We have $P(I \times [x]) = 0$ if x contains at least one 1, and $P(I \times [x]) = \mu(I)$ otherwise, see figure 1 right. For $I \subseteq [\alpha_1, \alpha_2]$ we have $P(I \times [0x]) = P(I \times [1x]) = 0$ if x contains at least one 1 and $\mu(I)/2$ otherwise.

We define the measure more formally for every basic open set $I \times [y] \subseteq 2^{\omega} \times 2^{\omega}$. We consider several cases:

- If $I \subseteq [\alpha, 1]$, then $P(I \times [y]) = \mu(I \times [y])$.

- If $I \subseteq [\alpha_n, \alpha_{n+1}]$ and $|y| \geq n$, then let $y = wx$ where w represents the first n bits of y. If:

    - x contains at least one 1, then $P(I \times [wx]) = 0$,
    - otherwise, i.e. if x is empty or contains only zeros, $P(I \times [wx]) = \mu(I \times [w])$.

- Otherwise, we partition the basic open set in (countably many) other basic open sets that satisfy one of the conditions above. The measure is the sum of the measures of all sets in the partition.

---

[1] The construction has some similarities with the measure constructed in the proof of Proposition 6.3 in [2]: the measure has also singularities that approach a left computable real. However, I believe there is no deeper correspondence between this measure and the measure constructed here.
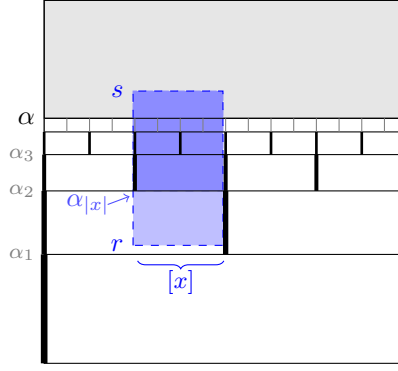
Figure 2: The Martin-Löf test V_n for P is obtained by trimming each basic set enumerated in a set U_n.

Note that for any string x, the P-measure of $[\alpha_{|x|}, \alpha] \times [x]$ equals the uniform measure, i.e.,

$$P([\alpha_{|x|}, \alpha] \times [x]) = \mu([\alpha_{|x|}, \alpha] \times [x]).$$

The same holds for any set $[r, s] \times [x]$ with $\alpha_{|x|} \leq r < s \leq 1$: $P([r, s] \times [x]) = \mu([r, s] \times [x])$, see figure 1 right. P is computable, because $\alpha_{|x|}$ is computable from x, and the horizontal line at $\alpha_{|x|}$ splits any interval $I \times [x]$ in at most two parts, see figure 2; for each part the P-measure is easily calculated: the measure of the part above $\alpha_{|x|}$ equals its uniform measure and the measure in the part below can be partitioned into finitely many pieces which each satisfies one of the cases in the definition of P.

We now determine the conditional measure at a horizontal coordinate β. Unless β contains finitely many ones, $P_C(\cdot|\beta)$ is the uniform measure with support $[\alpha, 1]$. For such β, the point α is not blind $P_C(\cdot|\beta)$-random: the open sets $U_n =]0, \alpha + 2^{-n}[$ contain α for all n, are uniformly effectively open and have $P(\cdot|\beta)$-measure $O(2^{-n})$. If β contains finitely many 1's, the conditional measure is piecewise constant and nonzero on infinitely many intervals below α. Hence, $P_C(\cdot|\beta)$ is defined for all β.

We choose β, such that (α, β) is Martin-Löf random relative to the uniform measure. By the original version of van Lambalgen's theorem, it suffices to choose β to be random relative to α. Clearly, β contains infinitely many ones and as argued in the previous paragraph, α is not $P_C(\cdot|\beta)$-random.

It remains to show that the pair (α, β) is also P-random. Let $(U_n)_{n \in \mathbb{N}}$ be a Martin-Löf test relative to P. It suffices to convert this test to a Martin-Löf test $(V_n)_{n \in \mathbb{N}}$ relative to the uniform measure such that U_n and V_n have the same intersection with the horizontal line at height α. More precisely, it suffices for each V_n to be uniformly effectively open such that:

- $U_n \cap (\{\alpha\} \times 2^{\mathbb{N}}) = V_n \cap (\{\alpha\} \times 2^{\mathbb{N}})$,

- $\mu(V_n) \leq P(U_n)$.

(Indeed, this implies that if (α, β) was not P-random, then it is also not random relative to the uniform measure and this would contradict the construction.) Construction of V_n: Each time an interval $[r, s] \times [x]$ is enumerated in a set U_n, enumerate its upper part starting from $\alpha_{|x|}$ in V_n, i.e., $[\max\{r, \alpha_{|x|}\}, s] \times [x]$ if $s > \alpha_{|x|}$ and nothing otherwise, see figure 2. Note that U_n and V_n have the same intersection with the line at height α because enumerated intervals are only modified below $\alpha_{|x|} < \alpha$. The sets V_n are uniformly effectively open. Hence, the first condition is satisfied. Finally, observe that $\mu(V_n) \leq P(U_n)$: for each enumerated interval $[r, s] \times [x]$, nothing is changed unless $r < \alpha_{|x|}$, and in this case we have $\mu([\alpha_{|x|}, s] \times [x]) = P([\alpha_{|x|}, s] \times [x]) \leq P([r, s] \times [x])$. Because V_n and U_n are the union of corresponding rectangles, the second condition is also satisfied.

$\square$

4

In the proof of Corollary 2, we use the following observation:

Lemma 5 (De Leeuw, Moore, Shannon and Shapiro [4]). *Let Q be a computable measure on $2^{\mathbb{N}}$ and let $\alpha \in 2^{\mathbb{N}}$. If there exists a set of positive Q-measure of sequences that compute α, then α is computable.*

Proof. Because there exists countably many machines, there exists a unique machine that computes α from a set of sequences with positive Q-measure. Let $c > 0$ be a lower bound for this Q-measure. We can enumerate a binary tree containing all strings x that can be computed on this machine from a set of oracles that has Q-measure at least c. This tree contains at most $1/c$ infinite branches and each such branch is computable. □

of Corollary 2. Let α and P be as constructed above. P_M is computable. The binary rational sequences have P_M-measure $\alpha < 1$, because P concentrates all measure below α on the binary rational sequences and above α, these sequences have measure zero. For each β that is not binary rational, the measure $P_C(\cdot|\beta)$ equals the uniform measure with support $[\alpha, 1]$. Let R be this measure. The function $x \mapsto R([x])$ computes α, hence R is not computable. Lemma 5 implies that the set of β that computes R has P_M-measure zero. Hence, at most a P_M-measure zero of sequences β that are not binary rational, compute $P_C(\cdot|\beta)$. The other non-rational sequences have measure $1 - \alpha > 0$ and satisfy the conditions of the corollary. □

Unfortunately, for any $[x]$ below α, the function $P_C([x]|\cdot)$ is nowhere continuous. It is only continuous in the set of points that are not binary rational, and the set of binary rational points is not negligable (it has P_M-measure α). Therefore, we present another example of such a measure for which the conditional measure is continuous, even for all β.

Theorem 6. *There exists a computable measure P on $\mathbb{N} \times 2^{\mathbb{N}}$ such that:*

- *for each $S \subseteq \mathbb{N}$, the function $P_C(S|\cdot)$ is defined and continuous on $2^{\mathbb{N}}$,*

- *the set of β for which $P_C(\cdot|\beta)$ is not computable relative to β, has P_M-measure one.*

Proof. Let A be a computably enumerable set that is not computable (for example the Halting problem). Fix an algorithm that enumerates the elements of A, and for each $n \in A$ let t_n be the time at which this algorithm enumerates n. The idea of the construction of P is the same as in [1]: if $n \notin A$, then the measure $P(\{n\} \times \cdot)$ is uniformly distributed over $2^{\mathbb{N}}$. Otherwise, the measure is non-uniform, but only at a very small scale, i.e., for $|x| \leq t_n$, the values of $P(\{n\} \times \cdot)$ do not depend on whether $x \in A$ or not, and only for $|x| > t_n$ the values are different. In this way, we guarantee that P is computable: if $|x| > t_n$, a program that computes $P(\{n\} \times [x])$ on input (n, x) can discover whether $n \in A$ and compute the different value. Because the conditional measure is defined in the limit, $P_C(\cdot|\beta)$ depends on this small scale structure, and therefore, the conditional measure can encode non-computable information.

To define P, we use the functions f_0 and f_1 which are defined graphically in the figure below. Note that the average of f_i over $2^{\mathbb{N}}$ is 1 for $i = 0, 1$. For $\beta \in 2^{\mathbb{N}}$, let β_t be the tth bit of β. Note that

$$\beta \mapsto f_i(\beta_{t+1}\beta_{t+2}\dots)$$

is the function obtained by repeating f_i with period 2^{-t}. These functions are all continuous and have average 1.
Let us first define P using the following density, see figure 3:

$$f(n, \beta) = \begin{cases} 2^{-n} f_0(\beta_{t_n+1}\beta_{t_n+2}\dots) & \text{if } n \in A \\ 2^{-n} & \text{otherwise.} \end{cases}$$

Thus, $P(\{n\} \times [x]) = \int_{[x]} f(n, \beta)\mathrm{d}\beta$.

Let $P(n|\beta)$ be short for $P_C(\{n\}|\beta)$. Is this function continuous in β? The marginal density $f_M = \sum_{i \in \mathbb{N}} f(i, \cdot)$ is continuous, because it is a uniformly convergent sum of
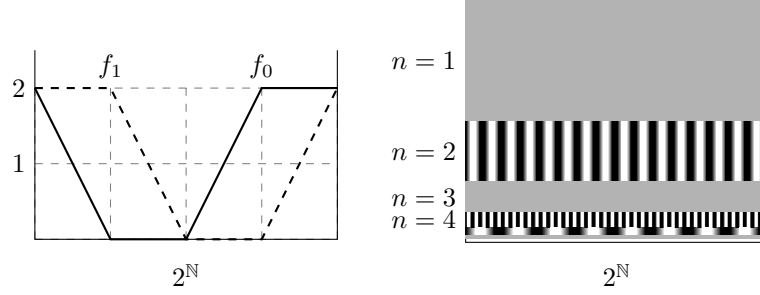
Figure 3: Left: measures f_0 and f_1 over $2^{\mathbb{N}}$. Right: measure P over $\mathbb{N} \times 2^{\mathbb{N}}$.

continuous functions. Also, f_M is bounded from below by a positive constant (if $m \notin A$, then $f_M \geq 2^{-m}$). Hence, the conditional measure is continuous on singleton sets:

$$P(n|\beta) = \frac{f(n, \beta)}{\sum_{i \in \mathbb{N}} f(i, \beta)}.$$

For $S \subseteq \mathbb{N}$, $P(S|\beta)$ is a uniformly convergent sum of continuous functions, and hence also continuous.

By Lemma 5, it remains to show for each β that $P(\cdot|\beta)$ computes A, (i.e., A is computed by a machine that has oracle access to approximations of $P(n|\beta)$ of any precision). For each fixed β, the values of $P(n|\beta)$ for all $n \notin A$ are the same. Unfortunately, there can be many $n \in A$ for which $P(n|\beta)$ is close to this value. Hence, A might not be computable from $P(\cdot|\beta)$.

We adapt the construction of P by encoding membership of n in A using two values of the conditional measure: $P(2n|\beta)$ and $P(2n+1|\beta)$. Note that for each $\beta \in 2^{\mathbb{N}}$ at least one of the values $f_0(\beta), f_1(\beta)$ is either 0 or 2. Hence, for $b \in \{0, 1\}$, we define P using

$$f(2n + b, \beta) = \begin{cases} 2^{-2n-b} f_b(\beta_{t_n+1} \beta_{t_n+2} \dots) & \text{if } n \in A, \\ 2^{-2n-b} & \text{otherwise.} \end{cases}$$

For the same reasons as before, $P(n|\cdot)$ is continuous. If $n \in A$ and $m \notin A$, at least one of the values $P(2n|\beta)$, $P(2n+1|\beta)$ is zero or $2P(2m|\beta) > 0$. Hence, $P(\cdot|\beta)$ computes A.

$\square$

Appendix: Two definitions of conditional measure coincide

In probability theory, conditional measures are defined implicitly using the Radon-Nikodym theorem. Any measure that satisfies the conditions of this theorem can be used as a conditional measure. The following lemma states that such measures are almost everywhere equal to the conditional measure P_C defined above.

Lemma 7 (Folklore). *Let 2^* be the set of strings. For every measure P on $2^{\mathbb{N}} \times 2^{\mathbb{N}}$ and for every function $f : 2^* \times 2^{\mathbb{N}}$ such that*

$$P([x], [y]) = \int_{[y]} f(x, \beta) P_M(d\beta).$$

we have that $f(\cdot, \beta) = P_C([\cdot]|\beta)$ for all β in a set of measure one.

In the proof we use the Lebesgue differentiation theorem for Cantor space. The proof of this version follows the original proof for Real numbers.

6

Theorem 8 (Lebesgue differentiation theorem for Cantor space). *Let Q be a measure on $2^{\mathbb{N}}$. For every Q-integrable function $g : 2^{\mathbb{N}} \to \mathbb{R}$ we have that*

$$\lim_n \frac{\int_{[\beta_1,\ldots,\beta_n]} g(\gamma)Q(d\gamma)}{Q([\beta_1,\ldots,\beta_n])} = g(\beta)$$

for Q-almost all β.

of Lemma 7. For a fixed x, apply the Lebesgue differentiation theorem with $g(\cdot) = f(x, \cdot)$ and $Q = P_M$. By assumption on f, the nominator simplifies to $P([x], [\beta_1 \ldots \beta_n])$. It follows that $f(x, \beta)$ differs from $P_C([x]|\beta)$ in at most a set of β with P_M-measure zero. Because there are countably many strings x, it follows that $f(\cdot, \beta)$ and $P_C([\cdot], \beta)$ differ in at most a set of P_M-measure zero. □

Acknowledgements

References

[1] N.L. Ackerman, C.E. Freer, and D.M. Roy. Noncomputable conditional distributions. In *Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 107–116. IEEE, 2011.

[2] L. Bienvenu and C. Porter. Strong reductions in effective randomness. *Theoretical Computer Science*, 459:55–68, 2012.

[3] G.J. Chaitin. A theory of program size formally identical to information theory. *J. Assoc. Comput. Mach.*, 22(3):329–340, 1975.

[4] Karel De Leeuw, Edward F Moore, Claude E Shannon, and Norman Shapiro. Computability by probabilistic machines. *Automata studies*, 34:183–198, 1955.

[5] B. Kjos-Hanssen. The probability distribution as a computational resource for randomness testing. *Journal of Logic and Analysis*, 2, 2010.

[6] M. Van Lambalgen. The axiomatization of randomness. *Journal of Symbolic Logic*, pages 1143–1167, 1990.

[7] J. Rute. Personal communication. 2014.

[8] Jason Rute. When does randomness come from randomness? *arXiv preprint arXiv:1508.05082*, 2015.

[9] A. Shen et al. Conditional probabilities and van Lambalgen theorem revisited. In preparation, 2015.

[10] H. Takahashi. On a definition of random sequences with respect to conditional probability. *Information and Computation*, 206(12):1375–1382, 2008.

[11] H. Takahashi. Algorithmic randomness and monotone complexity on product space. *Information and Computation*, 209(2):183–197, 2011.

[12] H. Takahashi. Generalization of van Lambalgen's theorem and blind randomness for conditional probability. *Preprint arXiv:1310.0709*, 2013. Presented in sept 2013 at CCR in Moscow.