

# Conditional probabilities and van Lambalgen theorem revisited\*

Bruno Bauwens<sup>†</sup>    Alexander Shen<sup>‡</sup>    Hayato Takahashi<sup>§</sup>

## Abstract

The definition of conditional probability in case of continuous distributions was an important step in the development of mathematical theory of probabilities. How can we define this notion in algorithmic probability theory? In this survey we discuss the developments in this direction trying to explain what are the difficulties and what can be done to avoid them. Almost all the results discussed in this paper have been published (and we provide the references), but we tried to put them into perspective and to explain the proofs in a more intuitive way. We assume that the reader is familiar with basic notions of measure theory and algorithmic randomness (see, e.g., [8] or [7] for a short introduction).

## 1 Conditional probability

Let  $P$  be a computable measure on the product of the two copies of the Cantor space  $\Omega_1 \times \Omega_2$ , and let  $P_1$  be its marginal distribution (=projection of  $P$  onto  $\Omega_1$ ). Consider some  $\alpha_1 \in \Omega_1$ . We want to define the conditional distribution on  $\Omega_2$  with the condition “the first coordinate is equal to  $\alpha_1$ ”. For that we consider a prefix  $a_1$  of  $\alpha_1$  and the conditional distribution on  $\Omega_2$  with the condition “the first coordinate starts with  $a_1$ ”. (For this we need that  $P_1(a_1)$ , the probability of the interval  $[a_1]$  of all extensions of  $a_1$ , to be positive.) In this way we get a family of measures  $P_{a_1}$  on  $\Omega_2$ :

$$P_{a_1}(a_2) = \frac{P(a_1, a_2)}{P_1(a_1)}$$

Here the numerator is the  $P$ -measure of the product  $[a_1] \times [a_2]$ , and the denominator is the  $P$ -measure of  $[a_1] \times \Omega_2$ . Then, for given  $a_2$ , we consider the limit of probability  $P_{a_1}(a_2)$  as the length of prefix  $a_1$  (of  $\alpha_1$ ) tends to infinity.

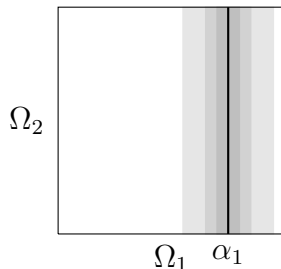
---

\*The paper records the discussion of conditional randomness results by the participants of the Heidelberg Focus Semester on algorithmic randomness, 2015 (Laurent Bienvenu, Mathieu Hoyrup, Rupert Hölzl, Wolfgang Merkle, Jason Rute and others)

<sup>†</sup>National Research University Higher School of Economics (HSE), Faculty of Computer Science, Kochnovskiy Proezd 3, Moscow, 125319 Russia, [bbauwens@hse.ru](mailto:bbauwens@hse.ru)

<sup>‡</sup>Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier, CNRS, Université de Montpellier, [alexander.shen@lirmm.fr](mailto:alexander.shen@lirmm.fr). Supported by ANR RaCAF grant.

<sup>§</sup>1-1 Yanagido, Gifu City 501-1193, Japan. Organization for Promotion of Higher Education and Student Support, Gifu University. [hayato.takahashi@ieee.org](mailto:hayato.takahashi@ieee.org). Supported by JSPS KAKENHI grant number 24540153.



**Theorem 1** ([9]). *If  $\alpha_1$  is Martin-Löf random with respect to  $P_1$ , then this limit is well defined and determines a measure on  $\Omega_2$ .*

*Proof.* For a fixed  $a_2$  the function  $m: a_1 \mapsto P_{a_1}(a_2)$  is a computable martingale on  $\Omega_1$  with respect to  $P_1$  (being a ratio of some measure and  $P_1$ ; note that the denominator  $P_1(a_1)$  is not zero since  $\alpha_1$  is random with respect to  $P_1$ ), so the limit exists due to effective martingale convergence theorem. Here is its proof. If  $m$  is a computable non-negative martingale, then for every rational  $c$  the set of sequences along which  $m$  exceeds  $c$  somewhere, is effectively open, and has measure at most  $1/c$ , if the initial capital is 1. So  $m$  is bounded along every Martin-Löf random sequence. Also, for every pair of positive rationals  $u < v$ , the set of sequences where the martingale is infinitely often less than  $u$  and infinitely often greater than  $v$ , is an effectively null set. Indeed, the set of sequences where there are at least  $N$  changes across  $(u, v)$ , is effectively open, and its measure is small due to “buy low — sell high” argument. Formally, we consider another martingale that follows  $m$ ’s bets, starting when  $m$  becomes less than  $u$ , until  $m$  becomes greater than  $v$ , and then waits for the next time when  $m$  becomes less than  $u$ , starts following  $m$  again, etc. So for Martin-Löf sequences convergence is guaranteed.<sup>1</sup>

To see that we indeed get a measure on  $\Omega_2$  in this way, we have to check the (finite) additivity which is obvious (limit of the sum is the sum of limits).  $\square$

This measure can be denoted by  $P(\cdot|\alpha_1)$ . Of course, if we start with a product measure  $P = P_1 \times P_2$ , the conditional probability is the same for all  $\alpha_1$ , and equals  $P_2$ .

*Remark.* The definition of a conditional measure in the classical probability theory is usually given using Radon–Nikodym theorem; then the Lebesgue differentiation theorem can be used to show that the conditional measure defined in this way coincides almost everywhere with the limit we considered.

## 2 Non-computable conditional probability

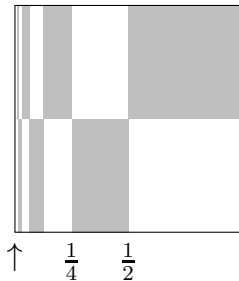
Let us note first that the limit in the definition of conditional probably may not exist for some conditions (though these conditions form a null set, as we have seen). This is shown by the following

*Example 1.* Consider the following distribution on  $\Omega_1 \times \Omega_2$ . We identify this product with the square  $[0, 1] \times [0, 1]$ ; the binary-rational points have two representations as sequences, but

---

<sup>1</sup>Technically the martingale  $m$  may be infinite if  $P_1(a_1) = 0$  for some strings  $a_1$ , but this does not matter for the argument: these intervals are covered by an effectively null set we construct.

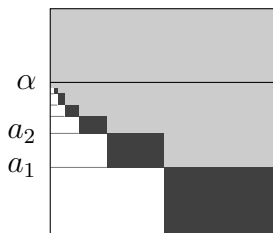
this does not matter much. In our distribution the grey areas have double density compared with the uniform distribution on the square, while the white areas have zero density.



Then for the leftmost point (shown by an arrow) the limit distribution does not exist. Indeed, the conditional probabilities of two halves oscillate between  $1/3$  and  $2/3$  depending on the length of  $a_1$ , as one may easily check.

Our next example shows that the conditional probability for a computable distribution on pairs might exist but be non-computable. A first example of this type was constructed in [1].<sup>2</sup>

*Example 2.* Let  $a_1, a_2, \dots$  be an increasing computable sequence of rational numbers whose limit  $\alpha$  is non-computable. Consider the following distribution:



Vertical lines are drawn at points  $1/2, 1/4, \dots$ ; in the grey zone the density is the same as for the uniform distribution; in the black zone the density is twice bigger, and in the white zone the density is zero. Since the widths of the black and white stripes on every horizontal line are the same, the total amount of mass does not change; we just move all the mass horizontally from the white part to the black part.

Note that the distribution on the square is computable even though  $\alpha$  is not computable. Indeed, if we are interested in the mass of some rectangle  $R$  (the product of two binary intervals), the mass transfers in the small rectangles (thinner than  $R$ ) do not matter, and we may look only on finitely many  $a_i$  (they can be computed).

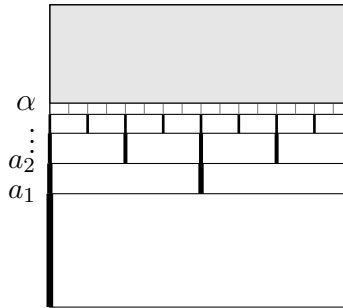
It is easy to see that the limit distribution (at the leftmost point) is the uniform distribution on  $[\alpha, 1]$ , and it is not computable, since the density  $1/(1 - \alpha)$  is not computable.

---

<sup>2</sup>In fact, the example in [1] has an additional property: the set of  $\alpha_1$  for which  $P(\cdot|\alpha_1)$  is not computable with oracle  $\alpha_1$ , has positive  $P_1$ -measure. The measure constructed in Example 3 (see below) also has this property, see [2, Corollary 2]. On the other hand, the example in [1] has a conditional measure  $P(\cdot|\alpha_1)$  that is continuous in  $\alpha_1$ , unlike Example 3.

In this example the conditional distribution is non-computable only at one point. However, the example can be easily changed so that the conditional distribution is the same non-computable distribution for all  $\alpha_1$  except for binary-rational numbers (sequences with finitely many zeros or ones).

*Example 3.* Consider again the increasing computable sequence  $a_1, a_2, \dots$  of rational numbers that converges to a non-computable real  $\alpha$ .



In the grey area above the horizontal line with coordinate  $\alpha$  we still keep the density the same as in the uniform distribution; however, below  $\alpha$  all the mass is concentrated on black vertical segments. For example, the mass  $a_1$  is concentrated on the segment  $\{0\} \times [0, a_1]$ , and is distributed uniformly there (so the mass transfer happens only in the horizontal direction). The mass  $a_2 - a_1$  is then split evenly between two vertical segments shown (at horizontal coordinates 0 and  $1/2$ ), etc. One can say that each vertical segment “horizontally grabs” all the mass of the white rectangle on the right of it (so the latter has zero density except for its left side).

As before, it is easy to see that the resulting distribution is computable: to find the mass of a binary rectangle of width  $2^{-n}$ , it is enough to take into account only  $a_1, \dots, a_n$  and use the uniform distribution above  $a_n$ .

It is also easy to see that for every  $\alpha_1$  that is not binary rational (in other words, for vertical lines that do not cross black segments), the conditional probability  $P(\cdot | \alpha_1)$  is uniformly distributed on  $[\alpha, 1]$ .

We will reuse this example in Section 5.

### 3 van Lambalgen theorem

Now we consider the relation between the randomness of the pair and its components. The basic result in this field goes back to Michiel van Lambalgen (see [6, Theorem 5.10], where this result is stated in an implicit way). It considers the case of the product  $P$  of two computable measures  $P_1$  on  $\Omega_1$  and  $P_2$  on  $\Omega_2$ , and says that the pair  $(\alpha_1, \alpha_2)$  is Martin-Löf random with respect to  $P$  if and only if two conditions are satisfied:

- $\alpha_1$  is random with respect to the measure  $P_1$ ;
- $\alpha_2$  is random with respect to the measure  $P_2$  with oracle  $\alpha_1$ .

(See, for example, [8, chapter 5] for the proof.) Note that these conditions are not symmetric; of course, one can exchange the coordinates and conclude that  $\alpha_1$  is also random with respect to  $P_1$  with oracle  $\alpha_2$ .

It is natural to look for some version of van Lambalgen theorem generalized to non-product measures  $P$ . Informally speaking, such a version should say that  $(\alpha_1, \alpha_2)$  is  $P$ -random if and only if

- $\alpha_1$  is  $P_1$ -random (where  $P_1$  is the projection of  $P$ );
- $\alpha_2$  is random with respect to the conditional probability measure  $P(\cdot|\alpha_1)$  with oracle  $\alpha_1$ .

But some precautions are needed. The problem is that Martin-Löf randomness is usually defined for computable measures, while the conditional measure is defined as a limit. As we have seen, it may not be computable even with oracle  $\alpha_1$  for  $P_1$ -random  $\alpha_1$ . Indeed, in our example the conditional distribution was  $[\alpha, 1]$  for every irrational condition  $\alpha_1$ , and among them there are  $P_1$ -random conditions that do not compute  $\alpha$ . To prove this, let us note that uniformly random reals are all  $P_1$ -random and some of them do not compute  $\alpha_1$ . Indeed, if the uniform measure of conditions  $\alpha_1$  that compute  $\alpha$  were positive, then the same would be true for some fixed oracle machine due to countable additivity of the uniform measure. Then Lebesgue density theorem says that there is some interval where most of the oracles compute  $\alpha$ , so  $\alpha$  can be computed without oracle by majority voting — but  $\alpha$  is not computable. (The last argument is known as de Leeuw – Moore – Shannon – Shapiro theorem [5].)

Still we can make several observations.

## 4 Image randomness and beyond

If  $(\alpha_1, \alpha_2)$  is (Martin-Löf) random with respect to  $P$ , then  $\alpha_1$  is (Martin-Löf) random with respect to the marginal distribution  $P_1$ . This is obvious; every cover of  $\alpha_1$  with small  $P_1$ -measure gives a cover of  $(\alpha_1, \alpha_2)$  with the same  $P$ -measure, being multiplied by  $\Omega_2$ .

This result can be considered also as a special case of the image randomness theorem (see the section about image randomness in [8]) applied to the projection mapping. Moreover, the reverse direction of image randomness theorem (“no randomness from nothing”) guarantees that every  $P_1$ -random  $\alpha_1$  is a first component of some  $P$ -random pair  $(\alpha_1, \alpha_2)$ .

So we know that for every  $P_1$ -random  $\alpha_1$  there exists at least one  $\alpha_2$  that makes the pair  $(\alpha_1, \alpha_2)$   $P$ -random. It is natural to expect that there are much more. Indeed this is the case, as the following result from [9] shows.

**Theorem 2.** *Let  $\alpha_1$  be  $P_1$ -random. Then the set of  $\alpha_2$  such that  $(\alpha_1, \alpha_2)$  is  $P$ -random, has probability 1 according to the conditional probability distribution  $P(\cdot|\alpha_1)$ .*

*Proof.* Consider a universal Martin-Löf test on the product space. Let  $U_n$  be the effectively open set of measure at most  $2^{-2n}$  provided by this test. In the case of product measure  $P_1 \times P_2$  we would consider the set  $V_n$  of all  $\alpha_1$  such that the  $\alpha_1$ -section of  $U_n$  has  $P_2$ -measure greater than  $2^{-n}$ , note that  $V_n$  is an effectively open set of measure at most  $2^{-n}$ , and conclude that random  $\alpha_1$  do not belong to  $V_n$  for all sufficiently large  $n$ . (If a point

$x$  is covered by infinitely many sets  $V_n$ , it is covered by all sets  $\overline{V}_n = \bigcup_{k>n} V_k$ , so  $x$  is not random. This is often called the *Solovay randomness criterion*.) So for every random  $\alpha_1$  and for sufficiently large values of  $n$  the  $\alpha_1$ -section of  $U_n$  has measure at most  $2^{-n}$ , so almost every  $\alpha_2$  lies outside the  $\alpha_1$ -section of  $U_n$  for all large  $n$ , which gives the desired result. Moreover, every  $\alpha_2$  that is  $P_2$ -random with oracle  $\alpha_1$  works, since the open cover for  $\alpha_2$  provided by the  $\alpha_1$ -section of  $U_n$  is  $\alpha_1$ -enumerable. (This is how the van Lambalgen theorem is proven.)

For the general case of non-product measure  $P$  we should be more careful since the conditional probability is defined only in the limit. Instead of  $V_n$ , we consider all  $n$ -heavy intervals  $I$  in  $\Omega_1$ , i.e., all intervals  $I$  such that  $U_n$  occupies more than  $2^{-n}$ -fraction in  $I \times \Omega_2$  measured according to  $P$ , in other words, all intervals  $I$  such that  $P(U_n \cap (I \times \Omega_2)) > 2^{-n}P_1(I)$ . This is an enumerable family of intervals since  $P$  and  $P_1$  are computable.

**Lemma 1.** *The  $P_1$ -measure of the union of all  $n$ -heavy intervals is at most  $2^{-n}$ .*

*Proof.* To prove that the union of  $n$ -heavy intervals has measure at most  $2^{-n}$ , it is enough to prove this for every finite union of  $n$ -heavy intervals. Without loss of generality we may assume that intervals in this union are disjoint (consider only maximal intervals). For every  $n$ -heavy interval  $I$  the fraction of  $U_n$  in the stripe  $I \times \Omega$  exceeds  $2^{-n}$ , so the total  $P_1$ -measure of disjoint  $n$ -heavy intervals cannot exceed  $2^{-n}$ , otherwise  $U_n$  would be too big (its measure would be greater than  $2^{-2n}$ ).  $\square$

In other words, the function

$$I \mapsto \text{fraction of } U_n \text{ in } I \times \Omega_2$$

is a (lower semicomputable) martingale with initial value  $2^{-2n}$ . So due to the martingale inequality the union of intervals where the martingale exceeds  $2^{-n}$  is at most  $2^{-n}$ .

**Lemma 2.** *If  $\alpha_1$  is outside any  $n$ -heavy interval, then the  $\alpha_1$ -section of  $U_n$  has measure at most  $2^{-n}$  according to the conditional probability with condition  $\alpha_1$ .*

*Proof.* If the conditional measure of the  $\alpha_1$ -section of  $U_n$  exceeds  $2^{-n}$ , then there exists a finite set of disjoint vertical intervals  $J_1, \dots, J_k$  that have total conditional measure more than  $2^{-n}$  and all belong to the  $\alpha_1$ -section of  $U_n$ . Since  $U_n$  is open, the compactness argument shows that for sufficiently small intervals  $I$  containing  $\alpha_1$  we have

$$I \times J_1, \dots, I \times J_k \subset U_n.$$

By assumption, the conditional measure of  $J_1 \cup \dots \cup J_k$  exceeds  $2^{-n}$ , and the conditional probability is defined as the limit of conditional probabilities with condition  $I$  when intervals  $I$  containing  $\alpha_1$  decrease. So for all sufficiently small  $I$  the conditional measure of  $J_1 \cup \dots \cup J_k$  with condition  $I$  exceeds  $2^{-n}$ , but this means that  $I$  is  $n$ -heavy and  $\alpha_1$  is covered by  $I$ , contrary to our assumption.  $\square$

Now consider (for some fixed random  $\alpha_1$ ) all  $\alpha_2$  such that  $(\alpha_1, \alpha_2)$  is non-random. Being non-random, these pairs belongs to all  $U_n$ , so for such a pair  $\alpha_2$  is inside  $\alpha_1$ -sections of  $U_n$  for all  $n$ . Since  $\alpha_1$  is random, it is not covered by  $n$ -heavy intervals  $I$  for all sufficiently large  $n$  and by Lemma 2 all bad  $\alpha_2$  are covered by a set of conditional measure  $2^{-n}$  at most  $n$  (for all large  $n$ ). So the  $\alpha_1$ -conditional measure of the set of bad  $\alpha_2$  is equal to 0.  $\square$

In fact, we have proven the following result from [10, 11] (one direction of van Lambalgen theorem).

**Theorem 3.** *If  $\alpha_1$  is  $P_1$ -random and  $\alpha_2$  is blind (Hippocratic) random with respect to the conditional probability  $P(\cdot|\alpha_1)$ , then the pair  $(\alpha_1, \alpha_2)$  is  $P$ -random.*

By *blind (Hippocratic)* randomness we mean a version of Martin-Löf definition of randomness with respect to noncomputable measure. In this version (studied by Kjos-Hansen [4]) uniformly effectively open tests are considered and the random sequence is required to pass all of them (if the measure is non-computable, there may be no universal test). It is opposed to *uniform* randomness where the test is effectively open with respect to the measure (see [3] for the details).

*Proof of Theorem 3.* Indeed, in the construction above we get a cover for bad  $\alpha_2$  that is enumerable with oracle  $\alpha_1$ . □

## 5 A counterexample

The following counterexample from [2] shows that the statement of Theorem 3 cannot be reversed.

**Theorem 4.** *There exists a computable measure  $P$  on  $\Omega_1 \times \Omega_2$ , for which conditional measure  $P(\cdot|\alpha_1)$  is defined for all  $\alpha_1$ , and a  $P$ -random pair  $(\alpha_1, \alpha_2)$  such that  $\alpha_2$  is not blind random with oracle  $\alpha_1$  with respect to conditional distribution  $P(\cdot|\alpha_1)$  on  $\Omega_2$ .*

*Proof.* We use the measure from Example 3. The second component  $\alpha_2$  of the pair is now a lower semicomputable random real  $\alpha$  that is the limit of a computable increasing sequence of binary fractions  $a_i$  (Chaitin's  $\Omega$ -number). We start with the following observation: for this  $\alpha_2$  the pair  $(\alpha_1, \alpha_2)$  is random if and only if this pair is random with respect to the uniform measure. Indeed, if we have some enumerable set of rectangles that covers  $(\alpha_1, \alpha_2)$ , we can safely discard parts of the rectangles that are below some  $a_i$ , since this does not change anything for  $(\alpha_1, \alpha_2)$ . In this way we may ensure that the  $P$ -measure of these rectangles equals their uniform measure (for a thin rectangle we need to discard more of it), so a  $P$ -test can be transformed into a uniform test and vice versa if we are interested only in points with second coordinate  $\alpha_2 = \alpha$ .

Now we can find a random point  $(\alpha_1, \alpha_2)$  with second coordinate  $\alpha_2 = \alpha$  (according to classical van Lambalgen theorem it is enough to take  $\alpha_1$  that is random with respect to uniform measure with oracle  $\alpha$ ). Since  $\alpha_1$  is random and not binary-rational, the conditional probability  $P(\cdot|\alpha_1)$  is uniformly distributed on  $[\alpha, 1]$ . It remains to show that the *lower semicomputable real  $\alpha$  is not blind random with respect to the uniform distribution on  $[\alpha, 1]$* . Indeed, for every rational  $\varepsilon > 0$  the interval  $(0, \alpha + \varepsilon)$  is effectively open, since it can be represented as the union of  $(0, a_i + \varepsilon)$ , and its measure with respect to the uniform measure on  $[\alpha, 1]$  is proportional to  $\varepsilon$  (so it is small for small  $\varepsilon$ ). □

## 6 The case of computable conditional measure

Still the van Lambalgen result can be generalized to non-product measure with an additional computability assumption. As before, we consider a computable measure  $P$  on  $\Omega_1 \times \Omega_2$  and its projection  $P_1$  on  $\Omega_1$  (the marginal distribution). The following result was proven by Hayato Takahashi [9, 10]:

**Theorem 5.** *If a pair  $(\alpha_1, \alpha_2)$  is  $P$ -random, and the conditional distribution  $P(\cdot|\alpha_1)$  is computable with oracle  $\alpha_1$ , then  $\alpha_2$  is Martin-Löf random with oracle  $\alpha_1$  with respect to this conditional distribution.*

Before proving this theorem, let us make several remarks about its statement:

- Under our assumption Martin-Löf randomness is well defined (the distribution is computable with oracle  $\alpha_1$ ).
- We already know that  $\alpha_1$  is  $P$ -random, so we get a randomness criterion for pairs (assuming the conditional distribution is computable given the condition).
- We assume the computability of conditional distribution *only for condition*  $\alpha_1$ ; for other random elements of  $\Omega_1$  (used as conditions) the conditional distribution may not be computable.

*Proof of Theorem 5.* Let us first recall the proof for the case of a product measure  $P_1 \times P_2$ . Assume that  $\alpha_2$  is *not* random. Then there is a set  $Z \subset \Omega_2$  of arbitrarily small  $P_2$ -measure that covers  $\alpha_2$  and is *effectively open with oracle*  $\alpha_1$ . The latter statement means that  $Z$  is a section of some effectively open set of pairs  $U \subset \Omega_1 \times \Omega_2$  obtained by fixing the first coordinate equal to  $\alpha_1$ . This set  $U$  covers  $(\alpha_1, \alpha_2)$  by construction. The problem is that only the  $\alpha_1$ -section of  $U$  is guaranteed to be small while other sections may be large, and we need a bound for the total measure of  $U$  to show the non-randomness of  $(\alpha_1, \alpha_2)$ .

The solution is that we “trim”  $U$  making all its sections small. Enumerating the rectangles in  $U$ , we look at the  $P_2$ -size of all sections. When some section attempts to become too big, we prevent this and stop increasing that section. In this way we miss nothing in the  $\alpha_1$ -section of  $U$  since it was small in the first place.

This argument works for the case of product distributions. How can we do similar things in the general case of arbitrary computable measures on  $\Omega_1 \times \Omega_2$ ? Again we start with a set  $Z$  of small conditional measure containing  $\alpha_2$  and represent  $Z$  as  $\alpha_1$ -section of some effectively open  $U \subset \Omega_1 \times \Omega_2$ . But trimming  $U$  is now not so easy. To understand the problem better, let us first consider two simple approaches that do not work.

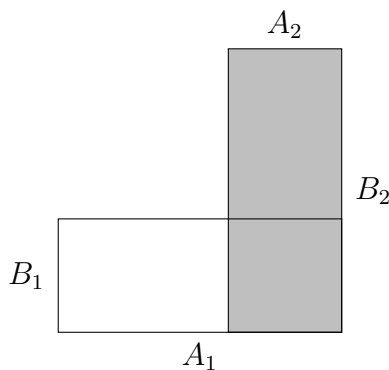
*First non-working approach.* The problem for the general case is that we have no “etalon” measure on sections that can be used for trimming. It is natural to use the conditional measure  $P(\cdot|\alpha_1)$ , and by assumption, there is an algorithm  $\Gamma$  that computes it using  $\alpha_1$  as oracle. Given some rectangle, we can split this rectangle horizontally (i.e., fix more and more bits of  $\alpha_1$ ) and use  $\Gamma$  to compute the conditional measure with more and more precision for all the parts, letting through only the rectangles where the vertical side is guaranteed to have small ( $\Gamma$ -computed) measure for the values of  $\alpha_1$  that belong to the horizontal side.



The problem with this approach is that the conditional measure is computable given  $\alpha_1$  *only on the  $\alpha_1$ -section* but not elsewhere. So the algorithm  $\Gamma$  may have no relation to the measure  $P$  outside this section. In this case small values produced by  $\Gamma$  do not guarantee anything about the measure of the rectangles that are let through, and we are in trouble.

*Second non-working approach.* Instead of computing the conditional probability, we may use the actual conditional probability when the condition is an interval; unlike the limit probability, it is computable. Imagine that we have some rectangle  $A \times B$ . Then we can compute the conditional probability of  $B$  with condition “the first coordinate is in  $A$ ”, i.e., the ratio  $P(A \times B)/P_1(A)$ , and let the rectangle through if this ratio is small (we assume that there are no earlier rectangles in the same vertical stripe). This guarantees that the  $P$ -measure of the rectangle is small; if we have several allowed rectangles with disjoint horizontal footprints, and for each of them this conditional probability is at most  $\varepsilon$ , then the  $P$ -measure of their union is also at most  $\varepsilon$ , since for each of them the  $P$ -measure is bounded by  $\varepsilon$  times the horizontal size of the rectangle, and the sum of horizontal sizes is at most 1.

What is the problem with this approach? (There should be a problem, since in this argument we do not use  $\Gamma$ , and this is unavoidable, as the counterexample above shows.) The problem becomes clear if we consider the case of overlapping rectangles.



For example, imagine that the set we want to trim contains some rectangle  $A_1 \times B_1$ . We compute the conditional probability  $P(B_1|A_1) = P(A_1 \times B_1)/P_1(A_1)$ . (Note that  $P_1(A_1) = P(A_1 \times \Omega_2)$ , so this conditional probability is the density of the rectangle in the vertical  $A_1$ -stripe, measured according to  $P$ .) We find that this conditional probability is slightly less than the threshold  $\varepsilon$ , so we let this rectangle ( $A_1 \times B_1$ ) through. Then we discover another rectangle  $A_2 \times B_2$  where  $A_2$  is a part of  $A_1$ , but  $B_2$  is bigger than  $B_1$  (as shown in the picture), and again find that  $P(B_2|A_2)$  is slightly less than  $\varepsilon$ . But if we let the second rectangle through, the average vertical measure of the resulting union may exceed  $\varepsilon$ . This could happen, for example, if all the mass in  $A_1 \times B_1$  is concentrated outside  $A_2 \times B_1$ ; then the conditional measure of  $B_1$  exceeds  $\varepsilon$  outside  $A_2$  and is zero inside  $A_2$ , thus leaving space for additional measure from  $B_2$ .

So the second approach also does not work. How can we deal with this problem?

*Main idea:* We combine the two approaches and always check (before adding something) that the actual conditional probability (with the interval as the condition) is close to the tentative conditional probability computed by  $\Gamma$ . The latter will remain almost the same for

smaller intervals (a valid computation remains valid when more information about the oracle is known), so the errors related to the change will be bounded.

*Details.* First we need to introduce some terminology and notation. We consider *basic* (=clopen) sets in  $\Omega_1 \times \Omega_2$ , i.e., finite unions of products of intervals. Every effectively open set is a union of a computable increasing sequence of basic sets. A basic set is a *rectangle* if it is the product of two clopen sets in  $\Omega$  (not necessarily intervals). By a *vertical stripe* we mean a rectangle  $S = I \times \Omega_2$ , where  $I$  is some interval in  $\Omega_1$  (i.e.,  $I$  consists of all extensions of some finite string). A basic set  $U$  is *stable in the stripe*  $S = I \times \Omega_2$  if  $U \cap S = I \times V$  for some  $V \subset \Omega_2$ . This means that all the vertical sections of  $U$  inside  $S$  are the same; we denote these sections by  $U|_S$ .

The *horizontal size*  $h(S)$  of a stripe  $S = I \times \Omega$  is defined as  $P_1(I)$  (and is equal to the  $P$ -measure of this stripe). If a set  $U$  is stable in the stripe  $S$ , its *vertical size in  $S$*  is defined as  $P(U \cap S)/P(S)$ , i.e., the conditional probability of  $U|_S$  with condition  $I$ . We denote the vertical size by  $v(U|S)$ . Note that the vertical size can increase or decrease if we replace  $S$  by a smaller stripe  $S'$  in  $S$  (and if it increases, say, for the left half of  $S$ , then it decreases for the right half); so “average vertical size” would be a better name for  $v(U|S)$ .

We want to trim an effectively open set  $U$  that is the union of a computable increasing sequence of basic sets

$$U_1 \subset U_2 \subset U_3 \subset \dots$$

Let us explain first which parts of  $U_1$  will be let through. We divide  $\Omega_1$  into two stripes, then divide each stripe into two halves, and so on. We use the algorithm  $\Gamma$  to get the approximations for the tentative conditional probabilities for all stripes. Let us agree, for example, that for stripes  $S$  of level  $n$  (with footprints of length  $2^{-n}$ ) we always make  $n$  steps of the  $\Gamma$ -computation, using  $n$  first bits of the oracle  $\alpha_1$  (i.e., the bits that are fixed for a stripe  $S$ ) and produce some lower and upper bounds  $\underline{P}^S(V)$  and  $\overline{P}^S(V)$  for the tentative conditional probability of all intervals  $V \subset \Omega_2$ . Note that for a given  $V$  the interval  $[\underline{P}^S(V), \overline{P}^S(V)]$  can only decrease as  $S$  becomes smaller. We know that these intervals should converge to  $P(V|\alpha_1)$  if we decrease the size of intervals  $V$  containing  $\alpha_1$ ; for other points the convergence is not guaranteed.

As soon as a stripe becomes small enough to make  $U_1$  stable in this stripe, we compute the lower and upper bounds  $\underline{v}(U_1|S)$  and  $\overline{v}(U_1|S)$  for the vertical size  $v(U_1|S)$  such that the difference between the lower and upper bounds is at most  $2^{-n}$  for stripes of level  $n$ .<sup>3</sup> Unlike for  $\underline{P}$  and  $\overline{P}$ , the interval  $[\underline{v}(U_1|S), \overline{v}(U_1|S)]$  does not necessarily decrease as  $S$  becomes smaller. Still they converge to the conditional probability of the  $\alpha_1$ -section of  $U_1$  if  $S$  are decreasing stripes around  $\alpha_1$  (since conditional probability is well defined for every  $P_1$ -random point in  $\Omega_1$ ).

If for some stripe  $S$  all the four numbers  $\underline{P}^S(U_1|S)$ ,  $\overline{P}^S(U_1|S)$ ,  $\underline{v}(U_1|S)$ , and  $\overline{v}(U_1|S)$  are close to each other, more precisely, if all four can be covered by some interval of size  $\delta_1$  (where  $\delta_1$  is a small number, see below), and at the same time the upper bound  $\overline{v}(U_1|S)$  is

---

<sup>3</sup>Since  $P$  is computable, we can compute  $v(U_1|S)$  for each  $U_1$  and  $S$  with arbitrary precision. The only exception is the case when  $P(S) = 0$ ; to avoid it, let us agree that we start processing stripe  $S$  only after we discover that  $P(S) > 0$ . In this way we lose all stripes with  $P(S) = 0$  but this does not matter since these stripes do not contain random pairs  $(\alpha_1, \alpha_2)$ . (Recall that our goal was to prove that  $(\alpha_1, \alpha_2)$  is not random contrary to the assumption.)

less than the threshold  $\varepsilon$  selected for trimming, we say that  $S$  is  $U_1$ -good and let  $U_1$  through inside  $S$ . Note that smaller stripes may be  $U_1$ -good or not, but this does not matter at this stage, since  $U_1$  is already let through inside  $S$ .

In this way we get a trimmed version  $\hat{U}_1 \subset U_1$ . The set  $\hat{U}_1$  may not be a basic set, but it is effectively open. Before going further, let us prove some properties of this construction:

1. Assume that some pair  $(\beta_1, \beta_2)$  is covered by  $U_1$ , the conditional probability  $P(\cdot|\beta_1)$  is well defined and is computed by  $\Gamma$  with oracle  $\beta_1$ , and the  $\beta_1$ -section of  $U_1$  has conditional measure (with condition  $\beta_1$ ) less than  $\varepsilon$ . Then  $(\beta_1, \beta_2)$  is covered by  $\hat{U}_1$ .
2. The  $P$ -measure of the trimmed set  $\hat{U}_1$  is at most  $\varepsilon$ .

Proof of 1: Indeed, look at the smaller and smaller stripes that contain  $\beta_1$ . Starting from some point,  $U_1$  is stable in these stripes, and the vertical size and tentative probabilities converge to some number smaller than  $\varepsilon$ . So they finally get into  $\delta_1$ -interval and are all less than  $\varepsilon$ . Therefore all small enough stripes containing  $\beta_1$  are  $U_1$ -good, and the  $\beta_1$ -section of  $U_1$  is not trimmed.

Proof of 2: In every  $U_1$ -good stripe  $S$  the vertical size  $v(U_1|S)$  is less than  $\varepsilon$ , so the measure of  $U_1$  inside this stripe is at most  $\varepsilon h(S)$ . We may consider only maximal  $U_1$ -good stripes, and the sum of their horizontal sizes is bounded by 1.

Now we switch to the next set  $U_2$  (we should decide which part of it should remain). We start to consider  $U_2$  only inside maximal  $U_1$ -good stripes selected at the first stage.<sup>4</sup> Let  $S$  be one of them. We start dividing  $S$  into smaller stripes; at some point they are small enough to make both  $U_1$  and  $U_2$  stable. Then we start checking if they are both  $U_1$ -good (according to our definitions) and  $U_2$ -good. The latter means that they satisfy the similar requirement for  $U_2$  with smaller error tolerance  $\delta_2$  (the four numbers for  $U_2$  are in some  $\delta_2$ -interval and the upper bound for the vertical size of  $U_2$  in the stripe is less than  $\varepsilon$ ). If we find a stripe  $S'$  inside  $S$  that is both  $U_1$ -good and  $U_2$ -good, then the set  $U_2$  is let through inside  $S'$ . So finally we have (inside  $S$ ) the set

$$(S \cap U_1) \cup \bigcup_{S'} (S' \cap U_2)$$

where the union is taken over stripes  $S' \subset S$  that are both  $U_1$ - and  $U_2$ -good. Doing this for all maximal  $U_1$ -good stripes  $S$ , we get the trimmed version  $\hat{U}_2$  of  $U_2$ . By construction  $\hat{U}_1 \subset \hat{U}_2 \subset U_2$ .

Now the key estimate for the size of  $\hat{U}_2$  inside a maximal  $U_1$ -good stripe  $S$ :

**Lemma 3.** *The  $P$ -measure of  $\hat{U}_2 \cap S$  is bounded by  $(\varepsilon + 2\delta_1)h(S)$ .*

Adding these inequalities for all maximal  $U_1$ -good stripes  $S$  (they are disjoint), we see that the total measure of  $\hat{U}_2$  is bounded by  $\varepsilon + 2\delta_1$ . (Note that  $\hat{U}_2$ , as well as  $\hat{U}_1$ , is contained in the union of maximal  $U_1$ -good stripes.)

---

<sup>4</sup>Note that  $U_1$ -good stripe may have empty intersection with  $U_1$ , so this does not prevent us from adding some stripes that intersect  $U_2$  but not  $U_1$ .

*Proof of Lemma 3.* The measure in question can be rewritten as

$$P(S \cap U_1) + \sum_{S'} P(S' \cap (U_2 \setminus U_1))$$

(we separate points added on the first and second stages). This sum can be rewritten as

$$h(S)v(U_1|S) + \sum_{S'} h(S')[v((U_2 \setminus U_1)|S')]$$

or

$$h(S)v(U_1|S) + \sum_{S'} h(S')v(U_2|S') - \sum_{S'} h(S')v(U_1|S')$$

Imagine for the moment that in the last term the condition is  $S$ , not  $S'$ . Then we could combine the first and last term and get

$$\left( h(S) - \sum_{S'} h(S') \right) v(U_1|S) + \sum_{S'} h(S')v(U_2|S') \quad (*)$$

The factors  $v(U_1|S)$  and  $v(U_2|S')$  are bounded by  $\varepsilon$  (for all  $S'$  where  $U_2$  is let through), and the sum of horizontal sizes is just  $h(S)$ , so the lemma is proven without  $2\delta_1$ -term. This term comes because of the replacement we made: the difference between  $v(U_1|S)$  and  $v(U_1|S')$  is bounded by  $2\delta_1$ , and the sum of all  $h(S')$  is at most  $h(S)$ . Indeed, the interval between lower and upper approximations  $\underline{P}, \overline{P}$  only decreases, and both sizes  $v(\cdot|S)$  and  $v(\cdot|S')$  are in  $\delta_1$ -neighborhood of every point in the smaller interval (that corresponds to  $S'$ ).  $\square$

To simplify the accounting in the future, we can rewrite the bound we have proved. The second term in (\*) is the size of  $U_2$  inside  $U_1$ - $U_2$ -good stripes  $S'$ , while the first term plus the error term bounded by  $2\delta_1 h(S)$  is the bound for the size of  $U_1$  inside  $U_1$ -stripe  $S$  minus  $U_1$ - $U_2$ -good stripes.

We can add these bounds for all maximal  $U_1$ -good stripes. Let  $G_1$  be their union, and let  $G_2$  be the union of maximal  $U_1$ - $U_2$ -good stripes (so  $G_2 \subset G_1$ ). Then  $\hat{U}_2$  is empty outside  $G_1$ , coincides with  $U_1$  inside  $G_1 \setminus G_2$ , and coincides with  $U_2$  inside  $G_2$ . The bounds for  $\hat{U}_2$  in the last two cases are  $\varepsilon P(G_1 \setminus G_2) + 2\delta_1 h(G_2)$  and  $\varepsilon P(G_2)$  respectively.

Another thing we need to check is the following. Assume that (1)  $(\beta_1, \beta_2)$  is covered by  $U_2$ ; (2) conditional probability  $P(\cdot|\beta_1)$  with condition  $\beta_1$  is well defined and  $\beta_1$ -conditional size of the  $\beta_1$ -section of  $U_2$  is computed by  $\Gamma$  with oracle  $\beta_1$ ; (3) this size is less than  $\varepsilon$ . Then  $(\beta_1, \beta_2)$  is covered by the trimmed set. Indeed, consider smaller and smaller stripes containing  $\beta_1$ . Starting from some point, both sets  $U_1$  and  $U_2$  are stable in those stripes, all the approximations converge (both for  $U_1$  and  $U_2$ ), and the limits are less than  $\varepsilon$ . So at some stage the stripes become both  $U_1$ - and  $U_2$ -good, and at this moment  $(\beta_1, \beta_2)$  is covered (unless this happened earlier).

Now we consider the next set  $U_3$ . The same construction is used: consider maximal  $U_1$ - $U_2$ -good stripes selected at the second stage. For each of them we look for stripes inside that are both  $U_2$ -good and  $U_3$ -good (the latter means that the set  $U_3$  is stable, all four parameters are  $\delta_3$ -close, and the upper bound for the vertical size is less than  $\varepsilon$ ). Then we

do the same thing as before, but not for  $U_1$ - $U_2$ -good stripes inside some  $U_1$ -good one, but for  $U_2$ - $U_3$ -good stripes inside some  $U_2$ -good one. The same approach is used for  $U_4$ ,  $U_5$ , etc.

In this way we get the set  $G_3$  that is the union of maximal  $U_2$ - $U_3$  good stripes. In this set  $U_3$  is let through to be included into  $\hat{U}_3$ . Then we get  $G_4$  where  $U_4$  is let through to be included into  $\hat{U}_4$ , etc. The same reasoning as in the proof of Lemma 3 gives us the following bounds:

**Lemma 4.**

- $P(U_{i-1} \cap (G_{i-1} \setminus G_i)) \leq \varepsilon P(G_{i-1} \setminus G_i) + 2\delta_{i-1}P(G_i)$ ;
- $P(U_i \cap G_i) \leq \varepsilon P(G_i)$ .

What have we achieved? We explained how to trim the set  $U_k$  for each  $k$  and get  $\hat{U}_k \subset U_k$ . The union  $\hat{U} = \bigcup_k \hat{U}_k$  is the trimmed version of the effectively open set  $U$  we started with. In other words,  $\hat{U}$  coincides with  $U_{i-1}$  inside  $G_{i-1} \setminus G_i$  and coincides with  $U$  in  $\cap_i G_i$ . What are the properties of this  $\hat{U}$ ?

- The trimming procedure is effective: the set  $\hat{U}$  is effectively open uniformly in  $U$ . This is guaranteed by the construction.
- The  $P$ -measure of  $\hat{U}$  is small. Indeed, for each  $k$  the measure of  $\hat{U}_k$  is bounded by  $\varepsilon + 2\sum_i \delta_i$  (sum of the first bounds from Lemma 4 for  $i = 2, \dots, k$  and the second bound for  $i = k$ ). Then we note that computable  $\delta_i$  can be chosen in such a way that  $\sum \delta_i < \varepsilon$ , and we achieve  $P(\hat{U}) \leq 3\varepsilon$ , since  $P(G_i) \leq 1$  for all  $i$ .
- Assume that conditional probability is well defined for some condition  $\beta_1$  and is computed by  $\Gamma$  with oracle  $\beta_1$ . Assume also that the  $\beta_1$ -section of  $U$  has conditional measure less than  $\varepsilon$  and contains some  $\beta_2$ . Then  $(\beta_1, \beta_2) \in \hat{U}$ . Indeed,  $(\beta_1, \beta_2)$  belongs to some  $U_i$  and (under the conditions mentioned) belongs to  $\hat{U}_i$  as explained above.

Then the proof ends in the same way as in the standard van Lambalgen theorem: since  $(\alpha_1, \alpha_2)$  is  $P$ -random, the first coordinate  $\alpha_1$  is  $P_1$ -random, the conditional probability  $P(\cdot|\alpha_1)$  is well defined, and our assumption says that it is computed by  $\Gamma$  with oracle  $\alpha_1$ . If  $\alpha_2$  is not (Martin-Löf) random with respect to the conditional probability, we consider a Martin-Löf test with oracle  $\alpha_1$  for measure  $P(\cdot|\alpha_1)$  rejecting  $\alpha_2$ , represent its elements as  $\alpha_1$ -sections of a sequence of uniformly effectively open sets  $U_i \subset \Omega_1 \times \Omega_2$ , and trim these sets. This gives us a Martin-Löf test with respect to measure  $P$  that rejects  $(\alpha_1, \alpha_2)$ , which contradicts the assumption.  $\square$

## 7 Quantitative version for uniformly computable conditional probabilities

In the previous section we started with a computable probability distribution  $P$  on  $\Omega_1 \times \Omega_2$ , and then defined the conditional distributions on  $\Omega_2$ . However, in many cases the

natural order could be different: we first generate a sequence  $\omega$  randomly according to some distribution  $P_1$  on  $\Omega_1$ , and then generate  $\omega'$  randomly according to some distribution  $P^\omega$  on  $\Omega_2$  that depends on  $\omega$ . If the dependence of  $P^\omega$  on  $\omega$  is computable, then we get some computable distribution  $P$  on  $\Omega_1 \times \Omega_2$ . It is easy to check that for  $P$  the conditional probabilities indeed coincide with  $P^\omega$ , so we can apply Takahashi's results from the preceding section. But in this special case the argument could be easier, and a stronger quantitative version could be obtained (as shown by Vovk and Vyugin in [12, Theorem 1, page 261], though in somehow obscure notation).

To state this quantitative version, we need to use the notion of randomness deficiency. More precisely, we use expectation-bounded randomness deficiency (see [3] for the details). In other words, we consider the maximal (up to  $O(1)$ -factor) lower semicomputable function  $t$  on  $\Omega_1$  with non-negative real values (including  $+\infty$ ) such that

$$\int_{\Omega_1} t(\omega) dP_1(\omega) \leq 1.$$

One can prove (see [3]) that such a function exists. We denote this maximal function by  $\mathbf{t}_{P_1}(\omega)$ ; the value  $\mathbf{t}_{P_1}(\omega)$  is finite for  $P_1$ -random  $\omega$  and infinite for non-random ones. Then we switch to the logarithmic scale and define deficiency as  $\mathbf{d}_{P_1}(\omega) = \log \mathbf{t}_{P_1}(\omega)$ .

In a similar way one can define randomness deficiency for pairs with respect to  $P$ : it is the logarithm of the maximal lower semicomputable function  $t(\omega, \omega')$  on  $\Omega_1 \times \Omega_2$  such that

$$\iint_{\Omega_1 \times \Omega_2} t(\omega, \omega') dP(\omega, \omega') = \int_{\omega} \int_{\omega'} t(\omega, \omega') dP^\omega(\omega') dP_1(\omega) \leq 1.$$

We denote this maximal function by  $\mathbf{t}_P(\omega, \omega')$  and its logarithm by  $\mathbf{d}_P(\omega, \omega')$ .

We need one more variant of randomness deficiency, and it is a bit more complicated. We want to measure the randomness deficiency of  $\omega'$  with respect to the measure  $P^\omega$  given some additional information as oracle. This additional information is  $\omega$  itself and some integer (its role will be explained later). We can use the general definition of uniform deficiency (as a function of a sequence and a measure, see [3]), but let us give an equivalent definition for this special case. A lower semicomputable function  $t(\omega', \omega, k)$  of three arguments ( $\omega'$  and  $\omega$  are sequences,  $k$  is an integer) is called a test, if

$$\int_{\omega'} t(\omega', \omega, k) dP^\omega(\omega') \leq 1$$

for every  $\omega$  and  $k$ . There exists a maximal test, as usual: we may trim all the lower semicomputable functions making them tests, and then take their sum with coefficients  $1/2^n$  (or other converging series). Trimming is easy since  $\omega$  is an argument and  $P^\omega$  is computable given  $\omega$  (uniformly for all  $\omega$ , according to our assumption). We denote the maximal test by  $\mathbf{t}_{P^\omega}(\omega'|\omega, k)$  and its logarithm by  $\mathbf{d}_{P^\omega}(\omega'|\omega, k)$ . (We use this notation since  $\omega$  is a parameter and  $k$  is an additional condition.)

Now we can state the Vovk–Vyugin result:

**Theorem 6.**

$$\mathbf{d}_P(\omega, \omega') = \mathbf{d}_{P_1}(\omega) + \mathbf{d}_{P^\omega}(\omega'|\omega, \mathbf{d}_{P_1}(\omega)) + O(1).$$

In this statement we assume that the value of  $\mathbf{d}_{P_1}(\omega)$  in the condition is rounded to an integer; the exact nature of rounding does not matter since it changes the deficiency only by  $O(1)$ .

Again, before proving this theorem, let us make some remarks:

- This result has high precision (up to  $O(1)$  additive term); if we were satisfied with logarithmic precision, we could omit  $\mathbf{d}_{P_1}(\omega)$  in the condition. Indeed, the standard argument shows that adding condition  $d$  could increase the deficiency at most by  $O(\log d)$  and decrease it at most by  $O(1)$ .
- It is easy to see that  $\mathbf{d}_{P^\omega}(\omega'|\omega, d)$  is finite if and only if  $\omega'$  is random with respect to measure  $P^\omega$  with oracle  $\omega$  (as we have mentioned, adding the condition  $d$  changes the deficiency at most by  $O(\log d)$ , so [in]finite values remain [in]finite). So we get a qualitative version:  $(\omega, \omega')$  is  $P$ -random if and only if  $\omega$  is  $P_1$ -random and  $\omega'$  is  $P^\omega$ -random with oracle  $\omega$ . (This statement generalizes the van Lambalgen theorem and is a special case of Takahashi's result considered in the previous section.)
- A special case of this statement, when  $P^\omega$  does not depend on  $\omega$  and is always equal to some computable measure  $P_2$ , gives a quantitative version of the van Lambalgen theorem for product measure  $P_1 \times P_2$ .
- One can consider a finite version of this theorem. If  $x$  is a constructive object, e.g., a string or a pair of strings, and  $A$  is a finite set containing  $x$ , we may define the randomness deficiency of  $x$  as an element of  $A$  in the following way:

$$d(x|A) = \log |A| - K(x|A),$$

where  $K(x|A)$  is the conditional prefix complexity of  $x$  given  $A$ . It is easy to check that  $d(x|A)$  is positive (up to  $O(1)$ -error) and that it can also be defined as a logarithm of maximal lower semicomputable function  $t(x, A)$  of two arguments ( $x$  is an object,  $A$  is a finite set) such that

$$\sum_{x \in A} t(x, A) \leq 1$$

for each finite set  $A$ . We can also define randomness deficiency with an additional condition as

$$d(x|A; y) = \log |A| - K(x|A, y).$$

Then we can state the following equality for the deficiency of a pair (with  $O(1)$ -precision):

$$d((x, y)|A \times B) = d(x|A; B) + d(y|B; x, A, d(x|A; B)).$$

It is just the Levin–Gacs formula for the complexity of pairs in disguise. Indeed, this statement can be rewritten (with  $O(1)$ -precision) as

$$\begin{aligned} \log |A \times B| - K(x, y|A, B) &= \\ &= \log |A| - K(x|A, B) + \log |B| - K(y|B, x, A, \log |A| - K(x|A, B)). \end{aligned}$$

The logarithms cancel each other and we have

$$K(x, y|A, B) = K(x|A, B) + K(y|B, x, A, \log |A| - K(x|A, B)).$$

which is just the Levin–Gacs theorem about prefix complexity of a pair (note that  $\log |A|$  in the condition does not matter since  $A$  is there anyway).

The proof below can be also adapted to the finite case.<sup>5</sup>

*Proof of Theorem 6.* We need to prove two inequalities. In each case, we construct some test and compare it with the maximal one.

We start with the  $\geq$ -direction, proving that  $d(\omega, \omega')$  is large enough. The function

$$T(\omega, \omega') = \mathbf{t}_{P_1}(\omega) \cdot \mathbf{t}_{P^\omega}(\omega'|\omega, \mathbf{d}_{P_1}(\omega))$$

(where  $\mathbf{d}_{P_1}(\omega)$  in the argument is rounded) has integral at most 1 with respect to measure  $P$ . Indeed,

$$\begin{aligned} \int_{\omega} \int_{\omega'} \mathbf{t}_{P_1}(\omega) \cdot \mathbf{t}_{P^\omega}(\omega'|\omega, \mathbf{d}_{P_1}(\omega)) dP^\omega(\omega') dP_1(\omega) &= \\ &= \int_{\omega} \mathbf{t}_{P_1}(\omega) \int_{\omega'} \mathbf{t}_{P^\omega}(\omega'|\omega, \mathbf{d}_{P_1}(\omega)) dP^\omega(\omega') dP_1(\omega) \leq \int_{\omega} \mathbf{t}_{P_1}(\omega) dP_1(\omega) \leq 1 \end{aligned}$$

(first we use that  $\mathbf{t}_{P^\omega}$  is a test, and then we use that  $\mathbf{t}_{P_1}$  is a test). One would like to say that this test  $T$  is bounded by the maximal test  $\mathbf{t}_P$ , but the problem is that the function  $T$  is not guaranteed to be a test: it may not be lower semicomputable, since it uses  $\mathbf{d}_{P_1}(\omega)$  as a condition. To avoid this problem, we consider a bigger function

$$T'(\omega, \omega') = \sum_{k < \mathbf{d}_{P_1}(\omega)} 2^k \mathbf{t}_{P^\omega}(\omega'|\omega, k).$$

It is indeed bigger (up to  $O(1)$ -factor) since the last term in the sum coincides with  $T$  up to  $O(1)$ -factor. This function is lower semicomputable since the property  $k < \mathbf{d}_{P_1}(\omega)$  is effectively open in the natural sense, and  $\mathbf{t}_{P^\omega}$  is lower semicomputable. And the integral is bounded not only for  $T$  but also for  $T'$ :

$$\begin{aligned} \iint \left[ \sum_{k < \mathbf{d}_{P_1}(\omega)} 2^k \mathbf{t}_{P^\omega}(\omega'|\omega, k) \right] dP^\omega(\omega') dP_1(\omega) &= \\ &= \int_{\omega} \left[ \sum_{k < \mathbf{d}_{P_1}(\omega)} 2^k \int_{\omega'} \mathbf{t}_{P^\omega}(\omega'|\omega, k) dP^\omega(\omega') \right] dP_1(\omega) \leq \int_{\omega} \left[ \sum_{k < \mathbf{d}_{P_1}(\omega)} 2^k \right] dP_1(\omega) \leq \\ &\leq O(1) \cdot \int_{\omega} 2^{\mathbf{d}_{P_1}(\omega)} dP_1(\omega) = O(1) \cdot \int_{\omega} \mathbf{t}_{P_1}(\omega) dP_1(\omega) = O(1). \end{aligned}$$

---

<sup>5</sup>It would be interesting to derive the statement of the infinite theorem using the formula for expectation-bounded deficiency in terms of prefix complexity and the formula for the complexity of pairs, but it is not clear how (and if) this can be done.



Here we use that the sum of different powers of 2 coincides with its biggest term up to  $O(1)$ -factor. So we have found a function  $T'$  that is lower semicomputable and is a test, so  $T'$  and therefore  $T$  are bounded by the maximal  $P$ -test, which gives the required inequality.

Now we have to prove the reversed ( $\leq$ ) inequality. For that we consider the maximal  $P$ -test  $\mathbf{t}_P(\omega, \omega')$  and the maximal  $P_1$ -test  $\mathbf{t}_{P_1}(\omega)$ . The function

$$\omega \mapsto \int_{\omega'} \mathbf{t}_P(\omega, \omega') dP^\omega(\omega')$$

is lower semicomputable and its integral with respect to measure  $P_1$  is at most 1, therefore this function is bounded by  $O(1) \cdot \mathbf{t}_{P_1}(\omega)$ . So the ratio

$$t(\omega', \omega) = \mathbf{t}_P(\omega, \omega') / \mathbf{t}_{P_1}(\omega)$$

has bounded integral over  $\omega'$  (with respect to  $P^\omega$ ) for each  $\omega$  (and the bound does not depend on  $\omega$ ). If  $t$  were a test, we could compare  $t$  with the maximal test  $\mathbf{t}_{P^\omega}(\omega'|\omega)$  and get the desired inequality. But again the function  $t$  may not be lower semicomputable, since it has lower semicomputable function in the denominator.

This is why we need an additional argument  $d$  for the test function. Namely, we consider

$$t(\omega'|\omega, d) = [\mathbf{t}_P(\omega, \omega') / 2^{d+c}],$$

where the square brackets denote trimming that make this function a test (the integral over  $\omega'$  for each  $\omega$  and  $d$  should be bounded by 1, and the trimming should not change  $t(\omega'|\omega, d)$  if for this  $\omega$  the integral was already bounded by 1). The constant  $c$  should be chosen in such a way that for  $d = \mathbf{d}_{P_1}(\omega)$  trimming is not needed; this is possible due to the argument above.

It remains to compare this test with the maximal one and note that for this test the required inequality is true by construction.  $\square$

There is another generalization of the van Lambalgen theorem. In the previous results we considered only computable measures. However, one can define *uniform randomness test* as a lower semicomputable function  $t(\omega, P)$  of two arguments (where  $\omega$  is a sequence, and  $P$  is a probability distribution on the Cantor space) such that

$$\int_{\Omega} t(\omega, P) dP(\omega) \leq 1$$

for every  $P$ . Note that we should first define the notion of semicomputability for functions whose arguments are measures; this can be done in a natural way (even for arguments in an arbitrary constructive metric space, see [3] for the details). Also we can generalize this definition by allowing points in constructive metric spaces as additional conditions. After that one could prove that

$$\mathbf{d}((\omega_1, \omega_2) | P_1 \times P_2) = \mathbf{d}(\omega_1 | P_1; P_2) + \mathbf{d}(\omega_2 | P_2; P_1, \omega_1, \mathbf{d}(\omega_1 | P_1; P_2)).$$

if all these quantities are defined in a natural way. It would be interesting to combine this generalization with Vovk–Vyugin result (where  $P_2$  is not fixed, but depends on  $\omega_1$ ). One

possibility is to assume that  $P_1$  is a computable function of some parameter  $p$  (a point in a constructive metric space). For example, we may let  $P_1 = p$ , so  $P_1$  itself may be used as such a parameter. Then we assume that  $P_2^\omega(\cdot)$  is a computable function of  $p$  and  $\omega$ , so the distribution on pairs also becomes a computable function of  $p$ . In this case we get the equality

$$\mathbf{d}((\omega_1, \omega_2) | P_1; p) = \mathbf{d}(\omega_1 | P_1; p) + \mathbf{d}(\omega_2 | P_2^{\omega_1}; p, \omega_1, \mathbf{d}(\omega_1 | P_1; p)) + O(1),$$

where  $O(1)$ -constant does not depend on  $p$ .

## 8 Acknowledgements

We are grateful to the organizers of the “Focus Semester on Algorithmic Randomness” in June 2015: Klaus Ambos-Spies, Anja Kamp, Nadine Losert, Wolfgang Merkle, and Martin Monath. We thank the Heidelberg university and Templeton foundation for financial support.

Alexander Shen thanks Vitaly Arzumanyan, Alexey Chernov, Andrei Romashchenko, Nikolay Vereshchagin, and all members of Kolmogorov seminar group in Moscow and ES-CAPE team in Montpellier.

## References

- [1] Nathaniel L. Ackerman, Cameron E. Freer, Daniel M. Roy, Noncomputable conditional distributions, In: *26th Annual IEEE Symposium on Logic in Computer Science (LICS)*, June 2011, p. 107–116. See also: *On the computability of conditional probability*, <http://arxiv.org/pdf/1005.3014v2.pdf>.
- [2] Bruno Bauwens, *Conditional measure and the violation of Van Lambalgen’s theorem for Martin-Löf randomness*, 2015, <http://arxiv.org/abs/1509.02884>
- [3] Laurent Bienvenu, Peter Gács, Mathieu Hoyrup, Cristobal Rojas, Alexander Shen, Algorithmic tests and randomness with respect to a class of measures, *Proceedings of the Steklov Institute of Mathematics*, **274**(1), 34–89 (2011). DOI:10.1134/S0081543811060058. See also <http://arxiv.org/abs/1103.1529>.
- [4] Bjørn Kjos-Hanssen, The probability distribution as a computational resource for randomness testing, *Journal of Logic and Analysis*, **2**(1), 1–13 (2010), doi: [url10.4115/jla.2010.2.10](http://dx.doi.org/10.4115/jla.2010.2.10). See also <http://arxiv.org/abs/1408.2850>
- [5] K. de Leeuw, E.F. Moore, C.E. Shannon, and N. Shapiro. Computability by probabilistic machines. *Automata studies*, edited by C.E. Shannon and J. McCarthy, *Annals of Mathematics studies* no. 34, lithoprinted, Princeton University Press, Princeton 1956, pp. 183–212.
- [6] Michiel van Lambalgen, The axiomatization of randomness, *The Journal of Symbolic Logic*, **55**(3), Sept. 1990, p. 1143–1167, <http://www.jstor.org/stable/2274480>.

- [7] Alexander Shen, Around Kolmogorov complexity: basic notions and results, in *Measures of Complexity: Festschrift for Alexey Chervonenkis*, Springer, 2015, p. 75-116. See also: <http://arxiv.org/pdf/1504.04955>
- [8] Alexander Shen, Vladimir Uspensky, Nikolay Vereshchagin, *Kolmogorov complexity and algorithmic randomness*, Moscow, MCCME, 2013; English version: <http://www.lirmm.fr/~ashen/kolmbook-eng.pdf>.
- [9] H. Takahashi. On a definition of random sequences with respect to conditional probability. *Information and Computation*, **206**(12):1375–1382, 2008.
- [10] H. Takahashi. Algorithmic randomness and monotone complexity on product space. *Information and Computation*, **209**(2):183–197, 2011.
- [11] H. Takahashi. *Generalization of van Lambalgen’s theorem and blind randomness for conditional probability*, <http://arxiv.org/pdf/1310.0709>.
- [12] V. G. Vovk and V. V. V’yugin. On the empirical validity of the Bayesian method. *Journal of the Royal Statistical Society. Series B* (Methodological), **55**(1): 253–266, 1993.