Conditional Probabilities and van Lambalgen’s Theorem Revisited

Bruno Bauwens¹ · Alexander Shen² · Hayato Takahashi³,⁴

Published online: 17 June 2017
© Springer Science+Business Media, LLC 2017

Abstract The definition of conditional probability in the case of continuous distributions (for almost all conditions) was an important step in the development of mathematical theory of probabilities. Can we define this notion in algorithmic probability theory for individual random conditions? Can we define randomness with respect to the conditional probability distributions? Can van Lambalgen’s theorem (relating randomness of a pair and its elements) be generalized to conditional probabilities? We discuss the

This article is part of the Topical Collection on Special Issue on Computability, Complexity and Randomness (CCR 2015)

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 Höllzl, Wolfgang Merkle, Jason Rute and others).

Alexander Shen
alexander.shen@lirmm.fr
Bruno Bauwens
bbauwens@hse.ru
Hayato Takahashi
hayato.takahashi@ieee.org

¹ National Research University Higher School of Economics (HSE), Faculty of Computer Science, Kochnovskiy Proezd 3, Moscow, 125319, Russia
² Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier, CNRS, Université de Montpellier, Montpellier, France
³ Organization for Promotion of Higher Education and Student Support, Gifu University, 1-1 Yanagido, Gifu City 501-1193, Japan
⁴ Present address: Random Data Laboratory, Tokyo 121-0062, Japan
developments in this direction. We present almost no new results trying to put known results into perspective and explain their 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., Shen et al. 2013 or Shen 2015 for a short introduction).

**Keywords** Conditional probability · Algorithmic randomness · van Lambalgen’s theorem

### 1 Conditional Probability

Let $P$ be a computable distribution on the product of two copies of Cantor space $\Omega_1 \times \Omega_2$, and let $P_1$ be its marginal distribution (= the 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 $a_1$, is 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 a 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.

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

**Proof** The proof goes as follows: 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 $P_1$, and an effective version of the martingale convergence theorem can be applied. Let us now elaborate the details.

We assume for now that $P_1(a_1) \neq 0$ for all $a_1$, so the ratio is well defined. This assumption is not always true, but at least $P_1(a_1) \neq 0$ for strings that are prefixes of random sequences. We return to the general case later. Under this assumption, the
function $m$ is well defined (for every fixed $a_2$) and indeed is a $P_1$-martingale; this means that

$$m(x) = \frac{P_1(x|0)}{P_1(x)}m(x|0) + \frac{P_1(x|1)}{P_1(x)}m(x|1)$$

and is true because $P_1(x)m(x)$ is a measure.

The classical martingale convergence theorem says that (with $P_1$-probability 1 over $\alpha_1$) the values $m(\alpha_1 \upharpoonright n)$ have finite limit as $n \to \infty$ (where $\alpha_1 \upharpoonright n$ is the $n$-bit prefix of $\alpha_1$). We need the effective version of this statement saying that this convergence happens for every $P_1$-random $\alpha_1$.

For the proof we use the effective version of the martingale convergence theorem. To see why this theorem can be effectivized, let us recall its proof. First we show that every martingale is bounded on almost all sequences (the martingale $m$ is always bounded by 1, but this argument will be used later for a different martingale). For that, we note that for every $c > 0$ and random $\alpha_1$, the probability of the event “the martingale exceeds $c$ at some prefix of $\alpha_1$” is at most $1/c$, and sequences where the martingale is unbounded belong to this event for every $c$.

To prove the convergence, it remains to show that for every pair of positive ratios $u < v$ the set of sequences where the martingale $m$ is infinitely often less than $u$ and infinitely often greater than $v$, is a null set. Indeed, the set of sequences where there are at least $N$ changes across $(u, v)$ has small measure due to a ‘buy low–sell high” argument. Formally, we apply the argument above to another martingale $m'$ that follows $m$’s bets, starting when $m$ becomes less than $u$ and doing this until $m$ becomes greater than $v$; then $m'$ waits (keeping the capital unchanged) for the next time when $m$ becomes less than $u$, then starts following $m$ again, etc.

An effective version of this argument says that if $P_1$ and $m$ are computable, then the $P_1$-null set of sequences that have no limit is an effectively $P_1$-null set, and therefore it does not contain $P_1$-random sequences. Indeed, the set of prefixes where the martingale is large, as well as the set of sequences where at least $N$ low–high oscillations happen, are effectively open, and have small measure (as the classical proof shows). Therefore, the convergence happens for all $P_1$-random $\alpha_1$.

To finish the proof, we need to consider the case when some $P_1(\alpha_1)$ are zeros. Still we may consider the set of prefixes where $P_1$ does not vanish and more than $N$ oscillations happen. This set is effectively open and has small measure as the same martingale argument shows.

Now we know that the limit exists (and is finite) for every $P_1$-random sequence $\alpha_1$. To see that we indeed get a measure on $\Omega_2$, it remains to check finite additivity which is obvious: a limit of a sum is the sum of limits. The measure of $\Omega_2$ is 1, so we get a probability distribution.

Remark 1 The definition of a conditional distribution in classical probability theory is usually given using the Radon–Nikodym theorem; the Lebesgue differentiation
theorem can be used to show that the conditional distribution 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 probability may not exist for some conditions (though these conditions form a null set, as we have seen). This is shown by the following example. In this section we identify \( \Omega_2 \) with the real interval \([0, 1]\); the dyadic rationals have two representations as sequences, but this does not matter much.

**Example 1** Consider the following distribution on \( \Omega_1 \times [0, 1] \). In our distribution the grey areas have double density compared with the uniform distribution on the square, while the white areas have zero density.

![Diagram](image)

It is easy to show that for the leftmost point (shown by an arrow) the limit that defines the conditional distribution does not exist. Indeed, the conditional probabilities of the two halves \([0, 1/2]\) and \([1/2, 1]\) oscillate between 1/3 and 2/3 depending on the length of \(a_1\), as one may easily check.

Example 2 below shows that the conditional probability for a computable distribution on pairs might exist but at the same time might be non-computable. An example of this type was first constructed in [1] and was more complicated.\(^1\)

**Example 2** Let \(r_1, r_2, \ldots\) be an increasing computable sequence of rational numbers whose limit \(\rho\) is non-computable. Consider the following distribution:

---

\(^1\)In fact, the example in [1] has an additional property that our example does not have: the set of \(a_1\) for which \(P(\cdot|a_1)\) is not computable with oracle \(a_1\), has positive \(P_1\)-measure. The distribution constructed in Example 3 (see below) also has this property, see also [2, Corollary 1]. On the other hand, the example in [1] has a conditional distribution \(P(\cdot|a_1)\) that is continuous in \(a_1\), unlike Example 3. A simplified construction of a continuous (in \(a_1\)) distribution can be found in [2, Theorem 5].
Vertical lines are drawn at right endpoints of the intervals \([0], [00], \ldots\); 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 this distribution (on the square) is computable even though \(\rho\) is not computable. Indeed, if we are interested in the mass of some rectangle \(R\) (the product of two dyadic intervals), the mass transfers in small rectangles (thinner than \(R\)) do not matter, and we may look only on finitely many \(r_i\) (they can be computed).

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

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 conditions \(\alpha_1\) except for sequences that contain finitely many ones. Here is the corresponding construction.

**Example 3** Consider again the increasing computable sequence \(r_1, r_2, \ldots\) of rational numbers that converges to a non-computable real \(\rho\).

In the grey area above the horizontal line with coordinate \(\rho\) we keep the same density as in the uniform distribution. Below \(\rho\) all the mass is concentrated on black vertical segments. For example, the mass \(r_1\) is concentrated on the segment \(\{000\ldots\} \times [0, r_1]\), and is distributed uniformly there (so the mass transfer happens only in the horizontal direction). The mass \(r_2 - r_1\) is then split evenly between two vertical segments shown (at horizontal coordinates 0000\ldots and 1000\ldots), etc. One
can say that each vertical segment “horizontally grabs” all the mass of the white rectangle on the right of it (so this white rectangle has zero density except for its left side, a sequence with finitely many ones).

As before, it is easy to see that the resulting distribution is computable: to find the mass of a dyadic rectangle of width $2^{-n}$, it is enough to take into account only $r_1, \ldots, r_n$ and use the uniform distribution above $r_n$.

It is also easy to see that for every $\alpha_1$ with infinitely many ones (in other words, for vertical lines that do not contain black segments), the conditional probability $P(\cdot | \alpha_1)$ is uniformly distributed on $[\rho, 1]$.

The conditional probability exists also for $\alpha_1$ with finitely many ones, i.e., for the case when $\alpha_1 = x000\ldots$ and in this case it is even computable: if the non-zero prefix of $\alpha_1$ consists of $k$ bits, the density of the conditional distribution is zero below $r_k$, the density is constant on $[r_k, r_{k+1}]$, twice smaller on $[r_{k+1}, r_{k+2}]$, again twice smaller on $[r_{k+2}, r_{k+3}]$, etc., and is zero above $\rho$. The density is computable due to fast convergence of the series $\sum (r_{k+1} - r_k) 2^{-k}$.

We will reuse this example in Section 5.

3 van Lambalgen’s Theorem

Now we consider the relation between randomness of a pair and its components. The basic result 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 distributions $P_1$ 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 distribution $P_1$;
- $\alpha_2$ is random with respect to the conditional probability distribution $P(\cdot | \alpha_1)$ 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’s theorem that can be applied to non-product distributions $P$. 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 distribution $P(\cdot | \alpha_1)$ with oracle $\alpha_1$.

But some precautions are needed. The problem is that Martin-Löf randomness is usually defined for computable distributions, while the conditional distribution may not be computable. Moreover, it may not be computable even with oracle $\alpha_1$ for $P_1$-random $\alpha_1$. Indeed, in Example 3 the conditional distribution was $[\rho, 1]$ for every irrational condition $\alpha_1$, and among them there are $P_1$-random conditions that do not compute $\rho$. To prove their existence, let us note that random reals with respect to the uniform Lebesgue measure are all $P_1$-random and some of them do not compute $\rho$. Indeed, if the uniform Lebesgue measure of conditions $\alpha_1$ that compute $\rho$ were
positive, then the same would be true for some fixed oracle machine due to countable additivity. Then the Lebesgue density theorem says that in this case there exists some interval where most of the oracles compute \( \rho \), so \( \rho \) can be computed without oracle by majority voting — but \( \rho \) 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.

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 the 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 most \( \alpha_2 \) should have this property. Indeed this is the case, as the following result from [9] shows.

**Theorem 2** Assume that \( \alpha_1 \) is \( P_1 \)-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. \( \square \)

First, let us consider the case where \( P \) is a product \( P_1 \times P_2 \) of two computable distributions. In this case the proof goes as follows:

- 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 \( 2^{-n} \).
- Conclude that a random \( \alpha_1 \) does not belong to \( V_n \) for all sufficiently large \( n \). Indeed, if a point \( \alpha_1 \) is covered by infinitely many sets \( V_n \), it is covered by all sets \( V_n' = \bigcup_{k>n} V_k \), so \( \alpha_1 \) is not random. (This is often called the Solovay randomness criterion.)
- Now we know that for every random \( \alpha_1 \) and for sufficiently large values of \( n \) the \( \alpha_1 \)-section of \( U_n \) has measure at most \( 2^{-n} \). If \((\alpha_1, \alpha_2)\) is not random, then this pair is contained in all \( U_n \), so \( \alpha_2 \) is covered by the \( \alpha_1 \)-section of all \( U_n \).
- Therefore, the set of \( \alpha_2 \) such that \((\alpha_1, \alpha_2)\) is not random, is a null set.

Moreover, for every \( \alpha_2 \) that is \( P_2 \)-random with oracle \( \alpha_1 \) the pair \((\alpha_1, \alpha_2)\) is random, since the open cover for \( \alpha_2 \) provided by the \( \alpha_1 \)-section of \( U_n \) is \( \alpha_1 \)-enumerable. (This is how one direction of van Lambalgen’s theorem is proven.)
For the general case of a non-product distribution $P$ we should be more careful since the conditional probability is defined as a limit, so we cannot use it directly to construct effectively open sets. Instead of $V_n$, we consider all $n$-heavy intervals $I \subset \Omega_1$, 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, an interval $I$ is $n$-heavy if $P(U_n \cap (I \times \Omega_2)) > 2^{-n}P_1(I)$. 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 all $n$-heavy intervals has measure at most $2^{-n}$, it is enough to prove this for an arbitrary finite union of $n$-heavy intervals. Without loss of generality we may assume that the intervals in this union are disjoint (consider only maximal intervals). For every $n$-heavy interval $I$ the fraction of $U_n$ in the stripe $I \times \Omega_2$ 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 $2^{-2n}$).

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 not covered by any $n$-heavy interval, then the $\alpha_1$-section of $U_n$ has measure at most $2^{-n}$ according to the conditional probability distribution with condition $\alpha_1$ (assuming this probability is defined for 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 $J_1, \ldots, J_k$ 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, a compactness argument shows that for sufficiently small intervals $I$ containing $\alpha_1$ we have

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

By assumption, the conditional measure of $J_1 \cup \ldots \cup J_k$ 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 $J_1 \cup \ldots \cup J_k$ 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.

Now consider (for some fixed random $\alpha_1$) all “bad” $\alpha_2$, i.e., all $\alpha_2$ such that $(\alpha_1, \alpha_2)$ is non-random. Being non-random, these pairs belong to all $U_n$, so every bad $\alpha_2$ 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 $n$. By Lemma 2, for those $n$ the $\alpha_1$-section of $U_n$ covers $\alpha_2$ and is a set of conditional measure at most $2^{-n}$. So the $\alpha_1$-conditional measure of the set of all bad $\alpha_2$ is equal to 0.
In fact, we have proven the following result (see Theorem 3 below) from [10, 11] (one direction of van Lambalgen’s theorem extended to non-product distributions). It uses the notion of blind (Hippocratic) randomness, a version of Martin-Löf definition of randomness extended to noncomputable distributions. 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 distribution 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 distribution (see [3] for the details); uniform randomness implies blind randomness (since more tests are allowed in the uniform version), so we may require uniform randomness in Theorem 3.

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

**Proof of Theorem 3** Indeed, the construction above gives a cover for bad $\alpha_2$ that is effectively open relative to oracle $\alpha_1$ (namely, the $\alpha_1$-section of $U_n$ for large $n$).

5 A Counterexample

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

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

This result shows that the implication stated by Theorem 3 cannot be reversed, and even a bit more: the reverse implication would state that $\alpha_2$ is blind random with oracle $\alpha_1$, but in our example $\alpha_2$ is not blind random even without oracle.

**Proof** We use the distribution from Example 3. The second component $\alpha_2$ of the pair is now a lower semicomputable random real $\rho$ that is the limit of a computable increasing sequence of dyadic rationals $r_i$ (e.g., Chaitin’s $\Omega$-number). We start with the following observation: for this $\alpha_2$ and for arbitrary $\alpha_1$ the pair $(\alpha_1, \alpha_2)$ is random if and only if this pair is random with respect to the uniform distribution. Indeed, non-randomness of a point means that we can effectively cover the point by an enumerable set of rectangles with arbitrarily small total measure. Now we have two distributions (the uniform one and the distribution from Example 3); let us show that the difference between them does not matter for covering points with second coordinate $\alpha_2 = \rho$. 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 $r_i$, 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 $\alpha_2 = \rho$.

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

6 The Case of Computable Conditional Distribution

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

**Theorem 5** If a pair $(\alpha_1, \alpha_2)$ is $P$-random, and for this $\alpha_1$ 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:

- If $(\alpha_1, \alpha_2)$ is $P$-random, then $\alpha_1$ is $P_1$-random, and by Theorem 1, $P(\cdot | \alpha_1)$ is well defined.
- Under our assumption Martin-Löf randomness with respect to this conditional distribution and with oracle $\alpha_1$ is well defined, because this distribution is computable with oracle $\alpha_1$.
- We assume the computability of the conditional distribution only for one given condition: $P(\cdot | \beta)$ is computable with oracle $\beta$ for $\beta = \alpha_1$, but for other $\beta$ it is not required (or guaranteed) even if $\beta$ is $P_1$-random.

Recalling Theorem 3 we get a randomness criterion for pairs:

**Corollary 1** For all $\alpha_1$ such that $P(\cdot | \alpha_1)$ is defined and computable with oracle $\alpha_1$, the following statements are equivalent:

- $(\alpha_1, \alpha_2)$ is Martin-Löf random with respect to distribution $P$,
- $\alpha_1$ is Martin-Löf random with respect to $P_1$ and $\alpha_2$ is Martin-Löf random with respect to $P(\cdot | \alpha_1)$ with oracle $\alpha_1$.

**Proof of Theorem 5** Let us first recall the proof for the case of a product distribution $P_1 \times P_2$. Assume that $\alpha_2$ is not random with oracle $\alpha_1$. 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 $\alpha_1$. Springer
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 distributions on \( \Omega_1 \times \Omega_2 \)? Again we start with a set \( Z \) of small conditional measure containing \( \alpha_2 \) and represent \( Z \) as the \( \alpha_1 \)-section of some effectively open \( U \subset \Omega_1 \times \Omega_2 \). But trimming \( U \) is now not so easy; we need to trim enough to obtain a set of small \( P \)-measure, but we should keep the \( \alpha_1 \)-section untouched. 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” distribution on sections that can be used for trimming. It is natural to use the conditional distribution \( P(\cdot | \alpha_1) \), 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 distribution 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 distribution is computable given \( \alpha_1 \) only on the \( \alpha_1 \)-section but not elsewhere. So the algorithm \( \Gamma \) may be unrelated to the distribution \( P \) 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 using the tentative conditional probability provided by \( \Gamma \), 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 \( \Gamma \), but we have to use it, as the counterexample above shows.) The problem is that the argument does not work if the horizontal footprints are not disjoint, as the following example shows.
For example, imagine that the set we want to trim contains some rectangle $A_1 \times B_1$. We compute the conditional probability $P(B_1 \mid A_1) = \frac{P(A_1 \times B_1)}{P_1(A_1)}$. (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 outside $B_1$ (as shown in the picture), and again find that $P(B_1 \cup B_2 \mid A_2)$ 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 $\Gamma_1$. The latter will remain almost the same for smaller intervals (a valid computation remains valid when more information about the oracle is available), 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 an increasing computable sequence of basic sets. A basic set is a rectangle if it is the product of two clopen sets in $\Omega_1$ (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 $U \mid S$. For example, the set $U = (A_1 \times B_1) \cup (A_2 \times B_2)$ in our example is stable in stripes $S = I \times \Omega_2$ if $I \subset A_2$; we have $U \mid S = B_1 \cup B_2$ for those stripes. The set $U$ is also stable in stripes $S = I \times \Omega_2$ for $I$ that are contained in $A_1 \setminus A_2$ (and $U \mid S = B_1$), and for $I$ that do not intersect $A_1$ (then $U \mid S = \emptyset$).
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 an increasing computable sequence of basic sets

$$U_1 \subset U_2 \subset U_3 \subset \ldots$$

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 the first $n$ bits of the oracle $\alpha_1$ (i.e., the bits that are fixed for a stripe $S$) and produce some lower and upper bounds $P^S(V)$ 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 $[P^S(V), \overline{P}^S(V)]$ can only decrease as $S$ becomes smaller. We know that these intervals should converge to $P(V|\alpha_1)$ as the length of the intervals $S$ containing $\alpha_1$ approaches zero; for other points (not $\alpha_1$) 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 $v(U_1|S)$ 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 $n$.

Unlike for $P$ and $P$, the interval $[v(U_1|S), \overline{v}(U_1|S)]$ does not necessarily decrease as $S$ becomes smaller: for a smaller stripe we may get an interval that is not a subset of the interval for a bigger stripe. Still the endpoints of these intervals converge to the conditional probability of the $\alpha_1$-section of $U_1$ if decreasing stripes around $\alpha_1$ are used as $S$ (recall that conditional probability is well defined for every $P_1$-random point in $\Omega_1$).

If for some stripe $S$ all the four numbers $P^S(U_1|S)$, $\overline{P}^S(U_1|S)$, $v(U_1|S)$, 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 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 $S$. If $S$ is not $U_1$-good, we continue the process and consider two halves of $S$, checking

---

2Since $P$ 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 this problem, 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.)
whether they are good or not (making \(n + 1\) steps of the computation, using precision \(2^{-(n+1)}\), etc.

In this way we get a trimmed version \(\hat{U}_1 \subset U_1\). Formally speaking, \(\hat{U}_1\) is the disjoint union of \(U_1 \cap S\) for all maximal \(U_1\)-good stripes \(S\). (We can take the union over all \(U_1\)-good stripes, but only maximal ones really matter.) 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:

**Lemma 3**

(a) 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 \(\hat{U}_1\).

(b) The \(P\)-measure of the trimmed set \(\hat{U}_1\) is at most \(\varepsilon\).

**Proof** To prove (a), 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 a \(\delta_1\)-interval with both endpoints below \(\varepsilon\). 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 (b): 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 \(\varepsilon h(S)\). Since all the maximal \(U_1\)-good stripes are disjoint, 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 after trimming). We consider \(U_2\) only inside maximal \(U_1\)-good stripes selected at the first stage.\(^3\) 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 the definition above) and \(U_2\)-good. The latter means that they satisfy the similar requirement for \(U_2\) with some 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 a stripe \(S'\) inside \(S\) turns out to be both \(U_1\)-good and \(U_2\)-good, then the set \(U_2\) is let through inside \(S'\), otherwise we continue the process and consider the halves of \(S'\). 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 maximal 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 \(\hat{U}_1 \subset \hat{U}_2 \subset U_2\).

The key part of the proof is the following bound for the size of \(\hat{U}_2\) inside a maximal \(U_1\)-good stripe \(S\):

\(^3\)Note that a \(U_1\)-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 \(U_1\).
Lemma 4 The $P$-measure of $\hat{U}_2 \cap S$ is bounded by $(\varepsilon + 2\delta_1)h(S)$.

Then, summing up the inequalities provided by Lemma 4 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 $\hat{U}_2$ is contained in the union of maximal $U_1$-good stripes.)

Proof of Lemma 4 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 during 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

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

The factors $v(U_1|S)$ and $v(U_2|S')$ are bounded by $\varepsilon$ (for all $S'$ that are $U_1$-$U_2$-good), and the sum of horizontal sizes is just $h(S)$, so the statement of the lemma would be true even without $2\delta_1$-term. This term compensates for the error that we made. Indeed, 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 $h(S)$. To see why the difference is bounded, note that the interval between the lower and upper approximations $P, \overline{P}$ only decreases when we switch from $S$ to $S'$, and both sizes $v(\cdot|S)$ and $v(\cdot|S')$ are in a $\delta_1$-neighborhood of every point in the smaller interval (that corresponds to $S'$).

Let us look again at this argument and introduce some notation that will be useful for the next stages. The second term in $(\ast)$ is the size of $U_2$ inside $U_1$-$U_2$-good stripes $S'$ (that are in a given maximal $U_1$-good stripe $S$), while the first term is an approximate (the difference is bounded by $2\delta_1 h(S)$) bound for the size of $U_1$ inside $S$ but outside maximal $U_1 - U_2$-good stripes $S' \subset S$.

We can sum up these bounds for all maximal $U_1$-good stripes $S$. Let $G_1$ be the union of these stripes, and let $G_2$ be the union of maximal $U_1$-$U_2$-good stripes (note that $G_2 \subset G_1$). 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 the size of $\hat{U}_2$ 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.

We have proved a statement similar to Lemma 3 (b), for the second step of the construction. Now we extend the statement (a) of that lemma in the same way. Assume that (1) $(\beta_1, \beta_2)$ is covered by $U_2$; (2) the conditional probability $P(\cdot|\beta_1)$ with condition $\beta_1$ is well defined and the $\beta_1$-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 we claim that $(\beta_1, \beta_2)$ is covered by the trimmed set $\hat{U}_2$. 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 the maximal $U_1$-$U_2$-good stripes (they are considered during 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 for a suitable $\delta_3$, see below, 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 maximal $U_1$-$U_2$-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 sub-stripes inside maximal $U_1$-$U_2$-good stripes considered earlier. Then $U_3 \cap G_3$ is added to $\hat{U}_2$ to obtain $\hat{U}_3$. In the next stage we define $G_4$ and let $\hat{U}_4 = (U_4 \cap G_4) \cup \hat{U}_3$, etc. The same reasoning as in the proof of Lemma 4 gives us the following bounds:

**Lemma 5**

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

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 $\hat{U}$ can be made arbitrarily small. Indeed, for each $k$ the measure of $\hat{U}_k$ is bounded by $\varepsilon + 2 \sum_{i \neq k} \delta_i$ (sum up the first bounds from Lemma 5 for $i = 2, \ldots, k$ and the second bound for $i = k$). Then we note that one can choose computable $\delta_i$ such that $\sum \delta_i < \varepsilon$; doing this, we conclude that $P(\hat{U}) \leq 3\varepsilon$, since $P(G_i) \leq 1$ for all $i$.
- Assume that the 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 of Theorem 5 ends in the same way as for the standard van Lambalgen’s theorem: if $(\alpha_1, \alpha_2)$ 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 oracle $\alpha_1$ with respect to the conditional probability, we consider a Martin-Löf test with oracle $\alpha_1$.
for the distribution $P(\cdot|\alpha_1)$ rejecting $\alpha_2$, represent its elements as $\alpha_1$-sections of a sequence of uniformly effectively open subsets of $\Omega_1 \times \Omega_2$ with small $\alpha_1$-sections, and trim each of these sets as described above. This gives us a Martin-Löf test with respect to distribution $P$ that rejects $(\alpha_1, \alpha_2)$, which contradicts the assumption.

7 A 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 a conditional distribution 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 $P_2^\omega$ on $\Omega_2$ that depends on $\omega$. If the dependence of $P_2^\omega$ on $\omega$ is computable (and therefore continuous), 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 $P_2^\omega$, so we can apply the results from the preceding section. But in this special case the argument could be easier, and a stronger quantitative version of Theorem 5 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 the notion of expectation-bounded randomness deficiency (see [3] for the details). In other words, to define the randomness deficiency for a computable distribution $P_1$ on $\Omega_1$ 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 $t_{P_1}(\omega)$; the value $t_{P_1}(\omega)$ is finite for $P_1$-random $\omega$ and infinite otherwise. Then we switch to logarithmic scale and define deficiency as $d_{P_1}(\omega) = \log t_{P_1}(\omega)$.

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

$$\int\int_{\Omega_1 \times \Omega_2} t(\omega, \omega') dP(\omega, \omega') = \int_{\omega} \int_{\omega'} t(\omega, \omega') dP_2^\omega(\omega') dP_1(\omega) \leq 1.$$ 

We denote this maximal function by $t_P(\omega, \omega')$ and its logarithm by $d_P(\omega, \omega')$.

We need one more variant of randomness deficiency which is a bit more complicated. We want to measure the randomness deficiency of $\omega'$ with respect to the distribution $P_2^\omega$ 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 distribution, 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 \( 1/2^n \) (other computably converging series can also be used). 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 \( t_{P^\omega}(\omega'|\omega, k) \) and its logarithm by \( d_{P^\omega}(\omega'|\omega, k) \).

Now we can state the Vovk–Vyugin result:

**Theorem 6**
\[
d_p(\omega, \omega') = d_{P_1}(\omega) + d_{P^\omega}(\omega'|\omega, d_{P_1}(\omega)) + O(1).
\]

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

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

- This result has high precision (it holds up to an \( O(1) \) additive term); if we were satisfied with logarithmic precision, we could omit \( d_{P_1}(\omega) \) in the condition. Indeed, the standard argument shows that adding a 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 \( d_{P^\omega}(\omega'|\omega, d) \) is finite if and only if \( \omega' \) is random with respect to distribution \( P^\omega \) with oracle \( \omega \): as we have mentioned, adding the condition \( d \) changes the deficiency at most by \( O(\log d) \), so \([\text{in}]\)finite values remain \([\text{in}]\)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 van Lambalgen’s theorem and is a special case of Theorem 5 from the previous section.)
- A special case of this statement, when \( P^\omega \) does not depend on \( \omega \) and is always equal to some computable distribution \( P_2 \), gives a quantitative version of van Lambalgen’s theorem for the product distribution \( P_1 \times P_2 \).
- 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 \( x \) (so \( A \) is also a constructive object), 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 \( A \). Note that we do not define \( d(x|A) \) for \( x \notin A \); one may also agree that \( d(x|A) = +\infty \) in this case. 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

\[
\log |A \times B| - K(x, y|A, B) = \log |A| - K(x|A, B) + \log |B| - K(y|B, x, A, \log |A| - K(x|A, B)).
\]

The logarithms cancel each other and we have (with \( O(1) \)-precision)

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

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 also be adapted to the finite case.\(^4\)

**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

\[
T(\omega, \omega') = t_{P_1}(\omega) \cdot t_{P^\omega}(\omega'|\omega, d_{P_1}(\omega))
\]

(where \( d_{P_1}(\omega) \) in the argument is rounded) has integral at most 1 with respect to distribution \( P \). Indeed,

\[
\int_{\omega} \int_{\omega'} t_{P_1}(\omega) \cdot t_{P^\omega}(\omega'|\omega, d_{P_1}(\omega)) \ dP^\omega(\omega') \ dP_1(\omega)
= \int_{\omega} t_{P_1}(\omega) \int_{\omega'} t_{P^\omega}(\omega'|\omega, d_{P_1}(\omega)) \ dP^\omega(\omega') \ dP_1(\omega) \leq \int_{\omega} t_{P_1}(\omega) \ dP_1(\omega) \leq 1
\]

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

\[
T'(\omega, \omega') = \sum_{k<d_{P_1}(\omega)} 2^k t_{P^\omega}(\omega'|\omega, k).
\]

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

\(^4\)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.
$k < d_{P_1}(\omega)$ is effectively open in the natural sense, and $t_{P_\omega}$ is lower semicomputable. The integral is bounded not only for $T$ but also for $T'$:

$$\int \int \left[ \sum_{k<d_{P_1}(\omega)} 2^k t_{P_\omega}(\omega'|\omega, k) \right] dP_\omega(\omega') dP_1(\omega)$$

$$= \int \omega \left[ \sum_{k<d_{P_1}(\omega)} 2^k \int \omega' t_{P_\omega}(\omega'|\omega, k) dP_\omega(\omega') \right] dP_1(\omega)$$

$$\leq \int \omega \left[ \sum_{k<d_{P_1}(\omega)} 2^k \right] dP_1(\omega) \leq O(1) \cdot \int \omega 2^{d_{P_1}(\omega)} dP_1(\omega)$$

$$= O(1) \cdot \int \omega t_{P_1}(\omega) dP_1(\omega) = O(1).$$

Here we use that the sum of different powers of 2 coincides with its biggest term up to an $O(1)$-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 $t_P(\omega, \omega')$ and the maximal $P_1$-test $t_{P_1}(\omega)$. The function

$$\omega \mapsto \int \omega' t_P(\omega, \omega') dP_\omega(\omega')$$

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

$$t(\omega', \omega) = t_P(\omega, \omega') / t_{P_1}(\omega)$$

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 $t_{P_\omega}(\omega'|\omega)$ and get the desired inequality. But again the function $t$ may not be lower semicomputable, since it has a lower semicomputable function in the denominator.

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

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

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 pair $(\omega, d)$ the integral was already bounded by 1). The constant $c$ should be chosen in such a way that for $d = d_{P_1}(\omega)$ 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.

There is another generalization of van Lambalgen’s theorem. In the previous results we considered only computable distributions. However, one can define a uniform randomness test as a lower semicomputable function $t(\omega, P)$ of two arguments.
\( \int_{\Omega} t(\omega, P) \, dP(\omega) \leq 1 \)

for every \( P \). Note that we should first define the notion of semicomputability for functions whose arguments are distributions; this can be done in a natural way (even for arguments in an arbitrary constructive metric space, see [3] for the details). There exists a universal uniform test \( t(\omega, P) \) and its logarithm \( d(\omega, P) = \log t(\omega, P) \) is called uniform randomness deficiency. For computable \( P \) the value of \( d(\omega, P) \) coincides with \( d_P(\omega) \) up to an additive constant that depends on \( P \). Also we can generalize this definition by allowing additional conditions (strings, oracles or even points in constructive metric spaces); we list these conditions after a semicolon, so \( d(\omega|P; \alpha, k) \) is the uniform randomness deficiency of \( \omega \) with respect to distribution \( P \) with additional conditions \( \alpha \) and \( k \).

One could prove that
\[
d((\omega_1, \omega_2)|P_1 \times P_2) = d(\omega_1|P_1; P_2) + d(\omega_2|P_2; P_1, \omega_1, d(\omega_1|P_1; P_2)) + O(1);
\]
it would be interesting to combine this generalization with the 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 \( P_2^{\omega_1}(\cdot) \) is a computable function of \( p \) and \( \omega_1 \), so the distribution on pairs also becomes a computable function of \( p \). In this case we get the equality
\[
d((\omega_1, \omega_2)|P_1; p) = d(\omega_1|P_1; p) + d(\omega_2|P_2^{\omega_1}; p, \omega_1, d(\omega_1|P_1; p)) + O(1),
\]
where the \( O(1) \)-constant does not depend on \( p \).

Acknowledgements  We are grateful to the organizers of the “Focus Semester on Algorithmic Randomness” (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. The visit of Hayato Takahashi to LIRMM was supported by NAFIT ANR-08-EMER-008-01 grant.

Alexander Shen thanks Vitaly Arzumanian, Alexey Chernov, Andrei Romashchenko, Nikolay Vereshchagin, and all members of Kolmogorov seminar group in Moscow and ESCAPE team in Montpellier.

Hayato Takahashi was supported by JSPS KAKENHi grant number 24540153.

Alexander Shen was supported by ANR-15-CE40-0016-01 RaCAF grant.

Last but not least, we are grateful to anonymous referees for very detailed reviews and many corrections and suggestions.

References

1. Ackerman, N.L., Freer, C.E., Roy, D.M.: Noncomputable conditional distributions. In: 26th Annual IEEE Symposium on Logic in Computer Science (LICS), pp. 107–116 (2011). See also: On the computability of conditional probability, arXiv:1005.3014v2.pdf

2. Bauwens, B.: Conditional measure and the violation of van Lambalgen’s theorem for Martin-Löf randomness. arXiv:1509.02884 (2015)

3. Bienvenu, L., Gács, P., Hoyrup, M., Rojas, C., Shen, A.: Algorithmic tests and randomness with respect to a class of measures. In: Proceedings of the Steklov Institute of Mathematics, vol. 274, pp. 34–89 (2011). doi:10.1134/S0081543811060058. See also arXiv:1103.1529
4. Kjos-Hanssen, B.: The probability distribution as a computational resource for randomness testing. J Log Anal 2(1), 1–13 (2010). doi:10.4115/jla.2010.2.10. See also arXiv:1408.2850

5. de Leeuw, K., Moore, E.F., Shannon, C.E., Shapiro, N.: Computability by probabilistic machines. Automata studies, edited by C.E. Shannon and J. McCarthy, Annals of Mathematics studies no. 34, lithoprinted, pp. 183–212. Princeton University Press, Princeton (1956)

6. van Lambalgen, M.: The axiomatization of randomness. J. Symb. Log. 55(3), 1143–1167 (1990). http://www.jstor.org/stable/2274480

7. Shen, A.: Around Kolmogorov complexity: basic notions and results. In: Measures of Complexity: Festschrift for Alexey Chervonenkis, pp. 75–116. Springer (2015). See also: arXiv:1504.04955

8. Shen, A., Uspensky, V., Vereshchagin, N.: Kolmogorov complexity and algorithmic randomness, Moscow, MCCME. English version: http://www.lirmm.fr/ashen/kolmbook-eng.pdf (2013)

9. Takahashi, H.: On a definition of random sequences with respect to conditional probability. Inf. Comput. 206(12), 1375–1382 (2008)

10. Takahashi, H.: Algorithmic randomness and monotone complexity on product space. Inf. Comput. 209(2), 183–197 (2011)

11. Takahashi, H.: Generalization of van Lambalgen’s theorem and blind randomness for conditional probability. arXiv:1310.0709

12. Vovk, V.G., V’yugin, V.V.: On the empirical validity of the Bayesian method. J. R. Stat. Soc. Ser. B Methodol. 55(1), 253–266 (1993)