Von Mises' notion of random sequence reconsidered

van Lambalgen, M.

DOI
10.1017/S0022481200029728

Publication date
1987

Published in
Journal of Symbolic Logic

THE JOURNAL OF SYMBOLIC LOGIC
Volume 52, Number 3, Sept. 1987

VON MISES' DEFINITION OF RANDOM SEQUENCES RECONSIDERED

MICHIEL VAN LAMBALGEN

Abstract. We review briefly the attempts to define random sequences (§0). These attempts suggest two theorems: one concerning the number of subsequence selection procedures that transform a random sequence into a random sequence (§§1–3 and 5); the other concerning the relationship between definitions of randomness based on subsequence selection and those based on statistical tests (§4).

Acknowledgement. We thank Johan van Benthem for his deconstruction of an earlier version of this manuscript, and also Roger Cooke and Mike Keane for helpful conversations.

Contents

§0. Introduction.

0.1. Random sequences were introduced by Richard von Mises ([1], [2], [3]), under the name of "collectives", as a foundation for probability theory. The crucial features characterising collectives are, on the one hand, the existence of limiting relative frequencies within the sequence and, on the other hand, the invariance of the limiting relative frequencies under the operation of "admissible place selection". An admissible place selection is a procedure for selecting a subsequence of the given sequence x in such a way that the decision to select a term x_n does not depend on the value of x_n. This attractive and seemingly clearcut notion has defied adequate formalisation up till now.

We believe that, in the last analysis, no such formalisation is possible in classical, Platonistic, mathematics. Some reasons for this conviction, together with a bird's eye view of the history of the problem, are given in §0.

Received March 11, 1986; revised September 19, 1986.
The author's research was supported by the Dutch Foundation for Scientific Research ZWO.

0022-4812/87/5203-0011/$04.10

In this article, we nevertheless investigate to what extent von Mises' intuitions can be salvaged using only classical mathematics. Perhaps somewhat surprisingly, the framework for this investigation is provided, not by the work of the direct heirs of von Mises, Wald and Church, but by the modern definitions of randomness using sequential tests, a development initiated by Martin-Löf [1] (see also Schnorr [1]).

The main result of the paper is that, if a sequence is random in the sense of Martin-Löf or Schnorr, then so are "almost all" of its subsequences. In §1 this rather vague statement is made precise. In §2 we state and prove an important technical lemma (in several versions, corresponding to different notions of randomness), which may be viewed as an effective analogue of Fubini's theorem for product measures. In §3, the main result is proved.

§4 is devoted to a comparison of notions of randomness based on place selections (in the spirit of von Mises, Wald, and Church) and notions of randomness based on statistical tests (following Martin-Löf). An old theorem of Ville says that the latter notions are more restrictive. We prove this theorem anew, using different techniques and obtaining a more informative result, which can be stated roughly as follows: statistical tests based on the law of large numbers and place selections cannot discriminate between processes converging to the same probabilities, whereas statistical tests in the sense of Martin-Löf or Schnorr can discriminate between processes which converge "slow" or "fast" to the same probabilities. A precise statement will of course be given in §4.

In §5, we interpret the results found in §§2 and 3 in terms of Kolmogorov complexity. An appendix, §6, collects notation and background material.

0.2. Von Mises' conception of probability theory differs widely from the current conception. Following Kolmogorov's ***Grundbegriffe der Wahrscheinlichkeitsrechnung***, we are used to thinking of probability theory as a branch of pure mathematics; but von Mises explicitly considered the laws of probability to be laws of nature, i.e. laws reflecting the statistical behaviour of particular large aggregates of physical objects. To quote from von Mises [2a]:

> The calculus of probability, i.e. the theory of probabilities, in so far as they are numerically representable, is the theory of definite observable phenomena, repetitive or mass events. Examples are found in games of chance, population statistics, Brownian motion etc. The word "theory" is used here in the same way as when we call hydrodynamics, the "theory of the flow of fluids", thermodynamics, the "theory of heat phenomena", or geometry, the "theory of space phenomena".
>
> Each theory of this kind starts with a number of so called axioms. In these axioms, use is made of general experience: they do not, however, state directly observable facts (102–103).

Von Mises' axioms try first and foremost to describe in (semi-) mathematical terms essential properties of the "repetitive or mass events" in whose statistical behaviour he is interested. These properties can be described as follows (in the version of [1]) (for unexplained notation we refer to the Appendix, §6).

Let M (for "Merkmalraum") be the set of possible outcomes (of an experiment) or "labels" (of an object).

A *collective* is a sequence $x \in M^{\omega}$ such that

(1) for all $A \subseteq M$, $P(A) := \lim_{n\to\infty}(1/n)\sum_{k=1}^{n} 1_A(x_k)$ exists, and

(2) if Φ is an *admissible place selection*, i.e. a selection of a subsequence of x which proceeds as follows: "Aus der unendliche Folge [x wird] eine unendliche Teilfolge dadurch ausgewählt, dass über die Indizes der auszuwählenden Elemente ohne Benützung der Merkmalunterschiede verfügt wird" (57); then also, for all $A \subseteq M$,

$$P(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1_A(\Phi x)_k.$$

The function P defined in (1) is called the *probability distribution determined by the collective* x. We see that in this set-up, probability is defined explicitly, as limiting relative frequency in a collective, rather than implicitly.

On this foundation, von Mises erected a probability calculus that overlaps with, but is not identical to, modern measure theoretic probability theory. It is not our purpose here to analyse in full detail in which mathematical aspects the two approaches differ; we refer the interested reader to von Mises' discussion of the strong law of large numbers in [2, p. 152].

We now return to the somewhat elusive notion of an admissible place selection, which forms the subject of this article. Von Mises attempts to make clear what he means by giving various examples. It suffices to consider the simplest cases of collectives, where M consists of just two elements, 0 and 1.

Here are some examples of admissible place selections:

—choose those x_n for which n is prime,

—choose those x_n which follow the word 010,

—toss a (different) coin; choose x_n if the nth toss yields heads.

The first two selection procedures may be called lawlike, the third random. It is more or less obvious that all of these procedures are admissible: the value of x_n is not used in determining whether to choose x_n.

0.3. Although von Mises' efforts met with sympathy, doubts were raised concerning the soundness of the foundations. The following comment is typical (E. Tornier [1, p. 320]).

> Ich glaube nicht, dass Versuche, die v. Misessche Theorie rein mathematisch zu fassen, zum Erfolg führen können, und glaube auch nicht, dass solche Versuche dieser Theorie zum Nutzen gereichen. Es liegt hier offensichtlich der sehr interessante Fall vor, dass ein praktisch durchaus sinnvoller Begriff—Auswahl ohne Berücksichtigung der Merkmalunterschiede—prinzipiell jede rein mathematische, auch axiomatische Festlegung ausschliesst. Wohl aber wäre es wünschenswert, dass sich diesem Sachverhalt, der vielleicht von grundlegender Bedeutung ist, das Interesse weiter mathematischen Kreise zuwendet.

As we see from the quotation just given, especially the second condition on collectives poses obstacles to mathematical treatment. Tornier does not deny that it has strong empirical credentials, but he envisages insuperable difficulties in stating what it means consistently.

The difficulties all come down to this (see e.g. Kamke's report for the Deutsche Mathematiker Verein [1, p. 23]).

Suppose, so the argument runs, that $x \in 2^{\omega}$ is a collective which induces the distribution $P(0) = P(1) = \frac{1}{2}$. Consider the set of all strictly increasing sequences of integers. This set can be formed independently of x; but, among its elements, we have the strictly increasing sequence $\{n \mid x_n = 1\}$ and this sequence defines an admissible place selection which selects the subsequence 111111... from x. Hence x is not a collective after all. To remedy this situation, Kamke proposes to restrict the class of increasing sequences of integers which may represent admissible place selections, to those sequences which are lawlike.

This argument calls for two remarks.

The reader may well feel uncomfortable with the mathematical structure of the argument. Kamke claims to have shown that for every putative collective x there exists an admissible place selection Φ such that Φx is not a collective.

The use of the existential quantifier here is classical, Platonistic, mathematics at its most extreme. Indeed, it seems impossible to exhibit *explicitly* a procedure which satisfies von Mises' criterion for admissibility and at the same time selects the sequence 1 1 1 1 1 1 ... from x. The author is convinced that a satisfactory treatment of random sequences is possible only in set theories lacking the power set axiom, in which random sequences "are not already there". In this article, however, we stay entirely within the classical framework.

Even if we uncritically accept classical mathematics, Kamke's argument is somewhat beside the mark in that it fails to appreciate the purpose of von Mises' axiomatisation. It refers to what *could* happen, whereas von Mises' axioms are rooted in experience and refer to what *does* happen. An analogy, which turns out to have heuristic value, may be helpful here. In various places von Mises likens condition (2) to the first law of thermodynamics. Both are statements of impossibility: condition (2) is the "principle of the excluded gambling strategy", while the first law (conservation of energy) is equivalent to the impossibility of *perpetuum mobile* of the first kind.

It may be even more appropriate to compare condition (2) to the second law of thermodynamics, the law of increase of entropy or the impossibility of a *perpetuum mobile* of the second kind, especially in view of Kamke's criticism. Indeed, Kamke's objection is reminiscent of Maxwell's celebrated demon, that "very observant and neat-fingered being", invented to show that entropy decreasing evolutions may occur. Maxwell's argument of course in no way detracts from the validity of the second law, but serves to highlight the fact that statistical mechanics cannot provide an absolute foundation for entropy increase, since it does not talk about what *actually* happens. This analogy is not far-fetched; indeed, it will be argued that our main result is related to von Mises' original proposal as statistical mechanics is related to thermodynamics. We shall come back to this point in §3.

Summarising, we can say that the intent of von Mises' axioms is not affected by Kamke's criticism; the question how to develop an adequate formalisation is still open.

0.4. The early attempts (before World War II) to formalise the theory went along the lines laid out by Kamke: start with a class of lawlike place selections and then construct a set of collectives with respect to that class.

Various authors (e.g. Popper, Reichenbach, Copeland) independently arrived at the so-called Bernoulli selections, which can be described as follows (for collectives x in 2^ω): fix a binary word w and choose all x_n such that w is a final segment of $x(n-1)$.

Note that this selection chooses an infinite subsequence from x iff x contains infinitely many occurrences of w. The *domain* of a place selection Φ will be the set of those x such that Φ, operating on x, yields an infinite subsequence of x.

We formalise these remarks in the following definition.

DEFINITION 0.4.1. Let $w \in 2^\omega$. We define a place selection $\Phi_w: 2^\omega \to 2^\omega$ in three steps. $\phi_w: 2^{<\omega} \to \{0,1\}$ is given by

$$\phi_w(u) = \begin{cases} 1 & \text{if } w \text{ is a final segment of } u, \\ 0 & \text{if not.} \end{cases}$$

$\bar{\Phi}_w: 2^{<\omega} \to 2^{<\omega}$ is given by

$$\bar{\Phi}_w(u\alpha) = \begin{cases} \bar{\Phi}_w(u)\alpha, & \phi_w(u) = 1, \\ \bar{\Phi}_w(u), & \phi_w(u) = 0. \end{cases}$$

Lastly, a partial function $\Phi_w: 2^\omega \to 2^\omega$ is defined by

(a) dom $\Phi_w = \{x \in 2^\omega \mid \forall n \exists k \geq n (\phi_w(x(k)) = 1)\}$, and

(b) $x \in$ dom Φ_w implies $\Phi_w(x) := \bigcap_n [\bar{\Phi}_w(x(n))]$.

DEFINITION 0.4.2. A *Bernoulli sequence* (with respect to a probability p) is a sequence $x \in 2^\omega$ such that for all words w:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} (\Phi_w x)_k = p.$$

Although von Mises admitted that his treatment of collectives left something to be desired (e.g. in [2], [3]), he did not consider Bernoulli sequences to be a satisfactory formalisation of the notion of a collective. He repeatedly emphasised that no *fixed* set of place selections suffices to treat all problems of probability theory. He thus abandoned the program of providing an "absolute" characterisation of collectives and retreated to an operationalist position:

> Wir verabreden, dass, wenn in einer konkreten Aufgabe ein Kollektiv einer bestimmten Stellenauswahl unterworfen wird, wir annehmen wollen, diese Stellenauswahl ändere nichts an den Grenzwerten der relativen Häufigkeiten. Nichts darüber hinaus enthält mein Regellosigkeits-axiom [2, p. 119].

A theorem of Wald showed the existence of collectives under conditions more general than those of Definition 0.4.2.

Wald first defined place selections without reference to admissibility as follows.

DEFINITION 0.4.3. Let $\phi: 2^{<\omega} \to \{0,1\}$ be any function. Then copy Definition 0.4.1 with ϕ instead of ϕ_w.

DEFINITION 0.4.4. Let $\mathfrak{A}$ be a countable set of place selections, $p \in (0,1)$. Then

$$C(\mathfrak{A}, p) := \left\{x \in 2^\omega \mid \forall \Phi \in \mathfrak{A} \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} (\Phi x)_k = p\right\}.$$

Wald's theorem can be put succintly as follows:

THEOREM 0.4.1. *For any* $\mathfrak{A}$, *any* $p \in (0,1)$, $C(\mathfrak{A}, p)$ *has the cardinality of the continuum.*

Von Mises was perfectly satisfied with this result; for him it showed that his probability theory is consistent:

> ...from what we know so far, it is certain that the probability calculus, founded on the notion of the collective, will not lead to logical inconsistencies in any of the applications of the theory known today [2a, p. 93].

0.5. We have seen that von Mises did not wish to fix once and for all a set of place selections and, hence, a set of collectives. Other researchers, however, tried to isolate a class of sequences which satisfy all "intuitively required" properties of randomness. E.g. Alonzo Church [1] suggested considering $C(\mathfrak{A}, p)$, where $\mathfrak{A}$ consists of all those place selections Φ such that the generating function $\phi: 2^{<\omega} \to \{0,1\}$ is (total) recursive.

We may now ask whether Church's proposal indeed yields an intuitively acceptable notion of collective. Although the set of collectives has been constructed using only countably many place selections, one feels that each individual collective should be much more random than that: even if it is impossible that every place selection selects a collective from a given collective (vide Kamke's objection), it should nevertheless be the case that "almost every" (in a sense to be made precise) does so. For the moment, we do not consider the difficult question how to obtain a satisfactory definition of *admissible* place selection; we just ask for a strong characterisation of randomness in classical mathematics. This problem will be taken up in §3, after preliminary work in §§1 and 2. The main result will be that, for the modern definitions of randomness to be introduced in 0.7, we have the following "principle of homogeneity" (so called after the homogeneous ensembles of quantum mechanics): If x is random, so is Φx for "almost all" place selection Φ.

0.6. We now return to the historical development. Although Wald's reformulation of von Mises' ideas solved the problem of consistency in a way, it lead to an objection of an entirely different kind, based on a theorem of Ville [1, p. 55]. For any countable set of place selections $\{\Phi_n\}$, Ville could construct $x \in 2^\omega$ such that

(a) for all n, $\lim_{m \to \infty}(1/m)\sum_{k=1}^{m}(\Phi_n x)_k = \frac{1}{2}$, and

(b) for all m, $(1/m)\sum_{k=1}^{m} x_k \geq \frac{1}{2}$.

Such an x is a collective with respect to the Φ_n, but seems to be far too regular to be called random. Formally, x's with property (b) form a set of Lebesgue measure 0, since Lévy had previously shown that (λ denotes Lebesgue measure)

$$\lambda\{x \in 2^\omega \mid \text{for infinitely many } m, (\tfrac{1}{2} - (1/m)\textstyle\sum_{k=1}^{m} x_k) > 1/\sqrt{m}\} = 1.$$

Lévy's law is a special case of the law of the iterated logarithm.

Ville and Fréchet used this construction to argue that collectives in the sense of von Mises and Wald do not necessarily satisfy all intuitively required properties of randomness. Von Mises remained unmoved, confining himself to the laconic remark: "J'accepte ce théorème, mais je n'y vois pas une objection" (in a letter to the Geneva conference on probability theory (1936) which he was unable to attend). As a partial explanation of von Mises' cavalier dismissal of this objection, we may note that, contrary to the first impression, the law of the iterated logarithm *is* derivable in von Mises' system, as pointed out by Wald at the Geneva conference.

In §4, we give a new proof of Ville's result, strengthening it in some respects and, in a way, explaining it. We cannot state our result precisely, however, until we have introduced the modern definitions of random sequences.

0.7. Ville introduced a new way of characterising random sequences, based on the following idea: a random sequence should satisfy all properties of probability one. Strictly speaking, this is of course impossible: we have to choose (countably many) from among those properties.

We note that, despite some superficial similarities, this idea is really completely foreign to von Mises. In the first place the relation between collective and probability expressed in the slogan "Erst das Kollektiv, dann die Wahrscheinlichkeit" is reversed by Ville. Moreover, for von Mises, a collective $x \in 2^\omega$ induces a probability on $\{0, 1\}$, not on 2^ω, hence there is no connection at all between properties of probability one in 2^ω and properties of individual collectives. Needless to say, from Ville's perspective the attempt to characterise random sequences loses much of its foundational importance. Ville even considered that the arbitrariness in choosing a countable set of properties of probability one made the notion of randomness irredeemably relative. But, interestingly, in the sixties a more or less canonical choice for these properties emerged, based on the fact that a correspondence can be established between certain types of sets of measure zero and statistical tests.

We recall some ideas from statistics. As always, we consider only experiments with outcomes 0 and 1. Let the measure μ on 2^ω be our hypothesis (usually, μ is of the form $(1 - p, p)^\omega$).

Fixed sample size tests for significance consist essentially in a partition of 2^n (where n is the sample size) in sets of sequences S_0, S_1; observation of an outcome sequence in S_0 leads to a rejection of the hypothesis μ at significance level $\alpha = \sum_{w \in S_0} \mu(w)$.

S_0 is also called a *critical region*. In practice, of course, we fix α and choose n accordingly.

We obtain a *sequential significance test* when the sample size is no longer fixed beforehand. It now suffices to specify a test statistic t and a number a such that, if w is the observed sequence of outcomes and $t(w) > a$, then the hypothesis μ is rejected at significance level

$$\alpha = \mu\{x \mid \exists n (t(x(n)) > a \wedge \forall m < n\, t(x(m)) \leq a)\}$$

Again, a is adjusted so as to obtain a prescribed value of α.

Now in practice $t: 2^{<\omega} \to \mathbf{R}$ will be a computable function. It follows that for a suitable computable real a, the critical region $\{x \mid \exists n (t(x(n)) > a \wedge \forall m < n\, t(x(m)) \leq a)\}$ is Σ_1.

This observation led Martin-Löf [1] to the following definition:

DEFINITION 0.7.1. Let μ be computable measure on 2^ω. A subset N of 2^ω is a *recursive sequential test with respect to* μ if a) N is a Π_2 set $\bigcap_n O_n$, O_n in Σ_1, b) $O_{n+1} \subseteq O_n$, and c) $\mu O_n \leq 2^{-n}$.

We may view O_n as the critical region corresponding to a significance level less than or equal to 2^{-n}.

DEFINITION 0.7.2. A sequence x is *random with respect to* μ ($x \in R(\mu)$) if x passes all recursive sequential tests with respect to μ, i.e. if for every recursive sequential test N with respect to μ, $x \notin N$.

In other words, the idea behind this definition is that a sequence is random with respect to a probability measure μ if it does not cause μ to be rejected at arbitrarily small levels of significance.

(Sequential tests were originally devised by Wald for the purpose of hypothesis testing rather than significance testing, e.g. to decide between two different probabilistic hypotheses on 2^ω. In this case, t can be taken to be the likelihood ratio of the two hypotheses.)

Lemma 4.2.2. will show us that there always exists a recursive sequential test which separates two computable, nonequivalent hypotheses. Note that, while a recursive sequential test is a Π_2 subset of 2^ω, a recursive fixed sample size test corresponds to a Π_1 set in 2^ω. Hence the following question is interesting, both in recursion theory and statistics: is it possible to separate two computable, nonequivalent measures by a recursive fixed sample size test? We know the answer only in the special case that the hypotheses are singular (Lemma 4.2.2(b)).

Subsequently, several definitions in the same vein have been proposed. Schnorr [1] introduced the *total recursive sequential tests* characterised by

DEFINITION 0.7.3. Let μ be a computable measure on 2^ω. A recursive sequential test N with respect to μ is called *total* if, for O_n as in Definition 0.7.1, the function $n \mapsto \mu O_n$ is computable. A sequence x is (*Schnorr*) *random* with respect to μ if for no total recursive sequential test N (with respect to μ), $x \in N$.

In other words, for a total recursive sequential test we have uniformly computable levels of significance, whereas for recursive sequential tests we only have computable upper bounds. Sometimes the computability condition in Definition 0.7.3 is hard to verify: this explains why theorems concerning total recursive sequential tests are often much harder to prove than the corresponding theorems for recursive sequential tests; cf. the effective Fubini theorems in §2.

The modification introduced by Gaifman and Snir [1] goes the other way; they abstract from levels of significance and concentrate on the arithmetical structure of statistical tests. Statistical tests are now just Π_n nullsets, for some n. We can introduce corresponding notions of randomness as follows.

DEFINITION 0.7.4. $x \in 2^\omega$ is *n-random* with respect to a computable measure μ if $x \notin N$ for all Π_n sets N with $\mu N = 0$.

The analogues of the main theorems for Π_n tests are easy to prove; cf. Theorem 2.1.1.

In view of the rather abstract nature of these definitions, the following questions immediately present themselves:

Do random sequences in the new sense satisfy familiar statistical laws such as the strong law of large numbers or the law of the iterated logarithm?

What is the relation between von Mises' proposal (in the brushed-up version of Wald and Church) and the new definition? We have already noted that the spirit behind the two definitions is totally different; we now ask how these definitions are related extensionally.

The first question is easily answered. Schnorr [1, Satz 10.1 and Satz 8.10] shows that the set

$$\left\{x \in 2^{\omega} \,\middle|\, \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k < \frac{1}{2} \text{ or } \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k > \frac{1}{2}\right\}$$

is contained in a total recursive sequential test with respect to λ (of course, an analogous result holds for other computable reals in (0.1)). It is typical for the degree of abstractness of the subject that the validity of the law of the iterated logarithm for the random sequences is not verified. This is easily remedied, however, using an effective version of the Borel-Cantelli lemma (see §4) and Feller's proof of the law of the iterated logarithm (Feller [1, Chapter VIII]). These two results of course only slightly alleviate the task of giving a concrete picture of $\aleph_0$ (total) recursive sequential tests. A few more concrete examples are provided in this essay.

As to the second question, if $\mathfrak{A}$ is the set of recursive place selections, and $R(\lambda)$ Martin-Löf's notion of randomness, it is not difficult to see that $R(\lambda) \subseteq C(\mathfrak{A}, \frac{1}{2})$. (For Schnorr's definition, there are some complications involving place selections whose domain does not have full measure.)

Ville's result, combined with the observation that the Martin-Löf random sequences satisfy the law of the iterated logarithm, shows that the inclusion is strict. In §4, we show that $C(\mathfrak{A}, \frac{1}{2}) \cap R(\lambda)^c$ is in fact rather large: we construct a nonatomic computable measure μ such that $\mu C(\mathfrak{A}, \frac{1}{2}) \cap R(\lambda)^c = 1$.

0.8. At the referee's suggestion, we add some remarks on Kolmogorov complexity and randomness. Undoubtedly the notion of complexity, with its attendant complexity-based definition of randomness, is the most important development issuing from von Mises' attempts to define Kollektivs. In fact, it is this definition of randomness which has found its way to the physics community, mainly in connection with research on (chaotic) dynamical systems. A typical problem in this area is the relation between traditional measures of chaotic behaviour such as entropy and measures of randomness for sequences. Measures of complexity have been most helpful here; we shall come to speak about this problem later (Theorem 5.2). References which treat the connection between chaotic behaviour and randomness include Alekseev and Yakobson [1], Lichtenberg and Lieberman [1] and Ford [1]; the latter two references present a somewhat garbled account of the subject.

In this subsection we introduce the pertinent definitions concerning complexity and state, without proof, a theorem on the relation between Martin-Löf randomness and complexity for infinite sequences (until the end of this subsection, randomness will always stand for randomness in the sense of Martin-Löf). In §5 this theorem is proved, together with a theorem on the relation between complexity and (metric) entropy, and we use these theorems to reflect upon results obtained in §§2 and 3.

Kolmogorov originally introduced the concept of complexity both to obtain a nonprobabilistic notion of "amount of information" and to characterise randomness for finite sequences (see [2] and the references given therein). Such a characterisation seemed to be called for in view of the foundations of probability;

the frequency interpretation, even if not explicitly present in the mathematical formulation of probability theory, necessitates a concept of random sequence. Infinite sequences having no practical utility whatsoever [1, p. 369], the problem arises to define randomness for finite sequences.

In later articles, Kolmogorov refined this program to study the conditions under which probability theory is applicable to the real world. In particular he was interested in the precise nature of the transition from chaotic sequences, which have no apparent regularities, to random sequences, which exhibit statistical regularities and form the domain of probability theory proper. He concluded that "practical deductions of probability theory can be justified as consequences about the limiting complexity, under given restrictions, of the phenomena in question" [2, p. 34]. The last statement will reappear near the end of this subsection, but now as a theorem on the relation between randomness and complexity of infinite sequences.

We now turn to the definition of complexity for finite (binary) sequences. The intuition behind the definition can be stated in various ways. One might say that if a sequence exhibits a regularity, it can be written as the output of a rule applied to a (simple) input. Another way to express this idea is that a sequence exhibiting a regularity can be coded efficiently, using the rule to produce the sequence from its code. Taking rules to be partial recursive functions from the set of finite binary words to itself, we may define the complexity of a word w with respect to a rule A as the length of the shortest input (also to be called program) p with $A(p) = w$. Sequences with low complexity (with respect to A) are then supposed to be fairly regular (with respect to A). In order to take account of all rules (i.e. partial recursive functions) we use a universal machine. Formally:

DEFINITION 0.8.1. Let $f: 2^{<\omega} \to 2^{<\omega}$ be a partial recursive function, with Gödelnumber $\ulcorner f \urcorner$. The complexity $K_f(w)$ of w with respect to f is defined to be

$$K_f(w) = \begin{cases} \infty & \text{if there is no } p \text{ such that } f(p) = w, \\ |p| & \text{if } p \text{ is the shortest input such that } f(p) = w. \end{cases}$$

An *asymptotically optimal* universal machine U is specified by the requirement that on inputs of the form $q = 0^{\ulcorner f \urcorner}1p$ (i.e. a sequence of $\ulcorner f \urcorner$ zeros followed by 1 followed by a string p), U simulates the action of f on p. Put $K(w) := K_U(w)$. Clearly (*) $K(w) \leq K_f(w) + \ulcorner f \urcorner + 1$.

$K(w)$ is called the *Kolmogorov complexity*, or just *complexity*, of w. Obviously the definition of complexity is open to the charge of arbitrariness on various accounts. For one thing, we might have chosen a different Gödelnumbering, or a different universal machine. This changes the value of complexity by an additive constant, and in particular the asymptotic results derived later are not affected by such manoeuvres. More serious is perhaps the decision to restrict the concept of a rule to partial recursive functions, especially when we bear in mind the intended application to the definition of randomness. Of course there do exist convincing arguments which show that mathematical rules must be equated with partial recursive functions; but a random sequence is a physical concept and the "absolute irregularity" commonly attributed to such sequences might signify the absence of more than just the computable regularities. We shall have more to say on this topic in §5.

We now embark upon the promised definition of finite random sequences. A note on terminology first: since the term "random" has been used up till now to designate sequences which exhibit only statistical regularities, we shall employ the word "irregular" for sequences which have as defining characteristic the absence of (simple) regularities. Note that (*) in Definition 0.8.1 implies that for some M, for all w, $K(w) \leq |w| + M$. We wish to say that a sequence is irregular if it has maximum complexity. Formally:

DEFINITION 0.8.2. Fix some natural number c. A binary word w is called *irregular* if $|w| > c$ and $K(w) > |w| - c$.

A simple counting argument shows that infinitely many irregular sequences exist.

The definition implies that for irregular w, the expression $K(w)/|w|$, which might be termed the *compression coefficient* of w, is close to unity; the condition in Definition 0.8.2 is however slightly stronger (cf. the difference between Theorem 5.1 and 5.2 below).

The peculiar logical properties of the set of irregular sequences (this set is immune, i.e. infinite, but without infinite r.e. subsets) will not concern us here. Rather, we are interested in the question of whether irregularity, i.e. absence of simple regularities, has anything to do with the statistical notion of randomness. As mentioned before, Kolmogorov devised complexity to obtain finite approximations to Kollektivs. He was able to show that an irregular sequence has relative frequency of 1 close to $\frac{1}{2}$, and that this relative frequency is not significantly changed if we select a subsequence from w using a "simple" algorithm (see [1]).

But our concept of randomness (i.e. Definition 0.7.2) is formulated for infinite sequences and, moreover, with respect to arbitrary (albeit computable) measures; whereas Definition 0.8.2 speaks of finite sequences and makes no mention of a probability distribution. How can we express in complexity-theoretic terms "irregularity with respect to a measure μ"? Before we tackle this problem, we consider irregularity for infinite sequences.

The extension of the definition of irregularity from finite to infinite sequences seems straightforward: we wish to say that an infinite sequence is irregular if all its initial segments are (except perhaps finitely many). Unfortunately this attempt comes to grief; *no* infinite sequence is irregular in this sense, as Martin-Löf has shown. The trouble is that, for infinitely many n, $K(x(n))$ is of the order of $n - \log_2 n$. One can show that for random x, $\liminf(n - K(x(n))) < \infty$, but the converse fails, so no straightforward characterisation of randomness seems possible using K. (Once we have the complexity function I, to be introduced below, at our disposal, we can use the relation between K and I to give such a characterisation; but it is pretty horrible.)

Satisfactory definitions of irregularity (i.e. yielding equivalence with randomness) have been obtained only by modifying the definition of K, typically by restricting the class of algorithms. A definition which stays as close as possible to the original intuition, that a sequence is irregular if (almost) all its initial segments are, has been given independently by Schnorr [2] and Levin [1]. Instead of arbitrary partial recursive functions, they use *monotone* partial recursive functions f, which have the property that if v is an initial segment of w, $f(v)$ is an initial segment of $f(w)$. This definition has some technical disadvantages (see Gacs [1]), which fact leads us to prefer the approach developed by Chaitin [1, 2] and, again, Levin [2], involving *prefixalgorithms.*

To explain this concept, we note that our definition of the complexity K has been slightly dishonest. The intuition behind the definition allegedly was that if p is a minimal program (on U) for w, then p contains all information necessary to reproduce w on U. Now this may well be false: U might begin its operation by scanning all of p to determine its length, only then to read the contents of p bit for bit. In this way, the information p is really worth $|p| + \log|p|$ bits, so it is clear that we have been cheating in calling $|p|$ the complexity of w. We may circumvent this problem if the reading heads of our machines are constrained to read the input in one direction only and if blanks are not allowed as endmarkers. We say that a machine A (of this type) performs a successful computation on input p if A halts when the reading head is scanning the last bit of p. The fact that we defined a successful computation using the last bit of p and not the first blank following p means that p must itself indicate how long it is: it must be a self-delimiting program. Formally, this means that the domain of A, that is, the set of p such that $A(p)$ halts, is *prefixfree*: if p, q are both in the domain of A, neither is an initial segment of the other. We may now introduce

DEFINITION 0.8.3. A *prefixalgorithm* is a partial recursive function $f\colon 2^{<\omega} \to 2^{<\omega}$ whose domain is prefixfree.

It is easily checked that there exists an enumeration of the set of prefixalgorithms, so defining U as in Definition 0.8.1, we see that U itself must be a prefixalgorithm. Analogous to Definition 0.8.2 we put

DEFINITION 0.8.4 (CHAITIN). $I(w) = \min\{|p| \mid U(p) = w\}$.

For I, we shall indiscriminatingly employ the terms "complexity" and "information". The latter name derives from the main strength of I: it has more or less the formal properties associated with an information function (in contradistinction to K). Interestingly, although complexity was designed to be a nonprobabilistic measure for information content, I can be written in a form which closely resembles the usual probabilistic definition of the information content of an event E as $-\log_2 \mu E$, where μ is some probability measure. We put

DEFINITION 0.8.5. $P(w) := \sum_{U(p)=w} 2^{-|p|}$.

This gives us

THEOREM 0.8.1. (a) $\sum_w P(w) \leq 1$.

(b) *For some c, for all w*: $|I(w) - [-\log_2 P(w)]| \leq c$.

P may be called the a priori probability on 2^ω (although it is not properly speaking a measure). We shall meet P again when we discuss the definition of irregularity with respect to a measure μ.

As was to be expected from our informal discussion of the meaning of I, the upper bound on I is now larger: for some c, for all w: $I(w) \leq |w| + I(|w|) + c$.

We may again define irregularity for finite sequences by fixing a c and calling w irregular if $I(w) > w + I(|w|) - c$. This stipulation suggests a definition of irregularity for infinite sequences of the form: $\exists m \forall n\, I(x(n)) > n + I(n) - m$. However, equivalence with randomness (for Lebesgue measure) is obtained only if we set

DEFINITION 0.8.6. x is *irregular* iff $\exists m \forall n\, I(x(n)) > n - m$.

In fact, this definition of irregularity has all the vices of its virtues: precisely because we have a perfect correspondence between the irregularity and randomness, irregular sequences satisfy the usual laws of probability, one of which is that for all ε, for almost all x, x has infinitely many initial segments of the form

$$x(n + [(1-\varepsilon)\log n]) = x(n)1^{[(1-\varepsilon)\log n]}.$$

Clearly the value of $I(x(n + [(1-\varepsilon)\log n]))$ lies about $[(1-\varepsilon)\log n]$ below its upper bound. (See Feller [1, p. 210] for the above-mentioned law of probability.)

Although we had to adjust our intuitions on irregularity a bit, still the fundamental idea survived, that irregularity is a relation between $I(x(n))$ and n. To incorporate the effect of different distributions, one might try to modify this relation. In this vein, Kolmogorov suggested [2] that a sequence is random with respect to $\mu_p = (1-p, p)^\omega$ if $I(x(n))$ is sufficiently close to $n \cdot H(\mu_p)$, where $H(\mu_p)$, the entropy of μ_p, is defined by $-p \cdot \log p - (1-p) \cdot \log(1-p)$ (for the definition of (metric) entropy for arbitrary measures on 2^ω see the Appendix). The correct definition of irregularity with respect to a computable distribution μ involves an entirely new idea, however, to wit a comparison between $I(x(n))$, the (algorithmic) information in $x(n)$, and $[-\log_2 \mu[x(n)]]$, the probabilistic information in $x(n)$. Precisely

DEFINITION 0.8.7. x is *irregular with respect to* μ if

$$\exists m \forall n\, I(x(n)) > [-\log_2 \mu[x(n)]] - m.$$

Irregularity in the sense of Definition 0.8.6 is easily seen to be the special case where $\mu = \lambda$, Definition 0.8.7 is justified by the following theorem, to be proved in 5.1:

THEOREM. *Let μ be a computable measure on 2^ω. Then x is in $R(\mu)$ iff x is irregular with respect to μ.*

One way to view this result is by means of a generalised likelihood ratio: if P is the a priori probability (on $2^{<\omega}$) defined above, we may write, using Theorem 0.8.1,

$$x \notin R(\mu) \quad \text{iff} \quad \forall m \exists n\, P(x(n))/\mu x(n) > 2^{m+c}.$$

Since P does not define a nice measure on 2^ω, this interpretation is not very appealing. In 5.2 we shall attempt to give a more plausible interpretation. One last remark: it will be noticed that length does not occur explicitly in the right-hand side of Definition 0.8.7. This might be a cause for surprise, since we set out with a definition of irregularity involving a comparison between complexity and length. The reason for the discrepancy is, that the intuition which lead to Definitions 0.8.2–0.8.6 really corresponds (formally) to the topological entropy of the system $(2^\omega, T)$, whereas the condition in Definition 0.8.7. is tailored to fit the metric entropy of the system $(2^\omega, T, \mu)$, as Theorem 5.2 will make plain. That Definition 0.8.6 is a special case of 0.8.7 reflects the fact that the system $(2^\omega, T, \mu)$ admits a maximum entropy measure (namely, $\mu = \lambda$), whose metric entropy equals the topological entropy of the system $(2^\omega, T)$. One may develop a theory of irregularity in the topological context, but a further discussion of this subject would take us too far afield.

0.9. We can now state the two main questions to be treated in this article. The first one is: make precise the following intuition (called the principle of homogeneity in 0.5): "If x is random (in the sense of Martin-Löf or Schnorr), so is Φx for "almost every" place selection Φ," and prove it. The precise version will be given in §1. §2 is devoted to a technical lemma (an effective form of Fubini's theorem). In §3 the precise version is proven, and both there and in §5 we comment on the result.

The second problem is: explain the difference between notions of randomness based on place selections and those on statistical tests. This will be done in §4.

Definitions of unexplained concepts and symbols can be found in §6.

§1. The principle of homogeneity. Stated informally, the principle of homogeneity says that if x is random, so is Φx for "almost all" place selections Φ. In this section we give a precise formulation of this statement and prove some preliminary lemmas.

1.1. To formulate the principle of homogeneity, we must put some sort of a measure on the set of place selections. This can be achieved by identifying this set with 2^ω. Let $/: 2^\omega \times 2^\omega \to 2^\omega$ be the (partial) function defined by $(x/y)_n = x_m$ iff m is the index of the nth 1 in y. The operation $/$ is defined for all pairs $\langle x, y \rangle$ such that y contains infinitely many ones. In order to avoid cumbersome statements of theorems, we introduce once and for all the following *convention*: *the expression "for all measures* ν *on* $2^\omega \cdots$" *is short for "for all probability measures* ν *on* 2^ω *such that* $\nu\{y \mid y$ *contains finitely many ones*$\} = 0 \cdots$".

Let Φ be a place selection; then for all x we can find y such that $\Phi x = x/y$. This observation suggests that we might profitably study place selections via the operation $/$.

We fix some notation. For $p \in (0,1)$, μ_p denotes the measure $(1-p, p)^\omega$; and LLN(p) denotes the set $\{x \mid \lim_{n\to\infty}(1/n)\sum_{k=1}^n x_k = p\}$.

Using standard probabilistic techniques, one can show

THEOREM 1.1.1. $x \in \text{LLN}(p)$ *iff for all* $q \in (0,1)$

$$\mu_q\{y \mid x/y \in \text{LLN}(p)\} = 1.$$

This theorem will be useful in §2, but it is defective as a formulation of the principle of homogeneity, since there is no mention of other randomness properties beside the law of large numbers. For example, using Wald's definition of randomness, one would like to have ($\mathfrak{A}$ is a countable set of place selections)

$$x \in C(\mathfrak{A}, p) \quad \text{implies} \quad \mu_p\{y \mid x/y \in C(\mathfrak{A}, p)\} = 1.$$

Such a result does not seem to be provable using the methods which suffice to establish Theorem 1.1.1. We therefore use a totally different method, inspired by Fubini's theorem, which, at least for notions of randomness based on statistical tests, leads to the desired result.

Principle of homogeneity (precise version). If $x \in R(\mu_p)$ (p a computable real, $R(\mu_p)$ the notion of randomness of Martin-Löf or Schnorr), then for all computable measures ν, $\nu\{y \mid x/y \in R(\mu_p)\} = 1$.

For the notions of Gaifman and Snir the principle can be formulated using a slightly stronger concept of computability for measures (see Theorem 3.1).

1.2. The method to be used in the proofs is based on the following two observations.

LEMMA 1.2.1 (DOOB). *If* Φ *is a place selection,* A *a Borel set in* 2^ω, *then* $\mu_p\{x \mid \Phi x \in A\} \leq \mu_p A$; *if* $\mu_p(\text{dom }\Phi) = 1$, *then we have in fact equality for all* A.

For a proof, see Schnorr [1, Lemma 3.3].

LEMMA 1.2.2 *For all* $p \in (0,1)$, *all measures* ν *on* 2^ω, *and all Borel sets* A *in* 2^ω:

$$(\mu_p \times \nu)\{\langle x, y \rangle \mid x/y \in A\} = \mu_p A.$$

PROOF. For fixed y, $/y\colon 2^\omega \to 2^\omega$ is a place selection which is total (we may assume that y contains infinitely many ones). By Fubini's theorem

$$(\mu_p \times \nu)\{\langle x, y\rangle \mid x/y \in A\} = \int 1_{\{\langle x,y\rangle \mid x/y \in A\}} d(\mu_p \times \nu)$$
$$= \int \mu_p\{x \mid x/y \in A\} d\nu(y) = \mu_p A. \qquad \blacksquare$$

Let $R(\lambda)$ be e.g. Schnorr's notion of randomness. Lemma 1.2.2 implies:

$$\text{for } \lambda\text{-a.a. } x \in R(\lambda)\text{:} \quad \lambda\{y \mid x/y \in R(\lambda)\} = 1.$$

To obtain the stronger result

$$x \in R(\lambda) \quad \text{implies} \quad \lambda\{y \mid x/y \in R(\lambda)\} = 1,$$

and eventually the principle of homogeneity itself, we must state and prove some effective refinements of Fubini's theorem. This is done in §2.

1.3. We collect some results that will be useful later.

LEMMA 1.3.1 *Let $\{N_n\}$ be a recursively enumerable union of total recursive sequential tests (with respect to μ). Then $\bigcup_n N_n$ is contained in a total recursive sequential test with respect to μ.*

This is Satz 8.10 in Schnorr [1]. Using this lemma, one can show (ibidem, Satz 10.1):

LEMMA 1.3.2. *Let p be a computable real. Then $\mathrm{LLN}(p)^c$ is contained in a total recursive sequential test with respect to μ_p.*

The proof of the following lemma is straightforward.

LEMMA 1.3.3. *If $O \subseteq 2^\omega$ is Σ_1, then the set $\{\langle x, y\rangle \mid x/y \in O\}$ is Σ_1, with Gödelnumber primitive recursive in a Gödelnumber for O.*

§2. Effective Fubini theorems. Let μ, ν be computable measures on 2^ω. It is obvious how to generalise the notion of (total) recursive sequential test (with respect to $\mu \times \nu$) to $2^\omega \times 2^\omega$.

Now let $N \subseteq 2^\omega \times 2^\omega$ be a total recursive sequential test with respect to $\mu \times \nu$; so $\mu \times \nu N = 0$. It follows from Fubini's theorem that

$$\mu\{x \mid \nu N_x > 0\} = 0.$$

This section addresses the following question: is it possible to construct a total recursive sequential test M (with respect to μ) such that $\{x \mid \nu N_x > 0\} \subseteq M$? The answer is yes, but the construction is somewhat complicated (2.3).

We also treat (more briefly) the analogous question for Π_n $(\mu \times \nu)$-nullsets (2.1) and recursive sequential tests with respect to $\mu \times \nu$ (2.2).

2.1. The definitions of *computable* and *strongly computable* measures can be found in §6.

The following lemma is in essence due to Sacks (see Sacks [1] or Kechris [1]).

LEMMA 2.1.1. (i) *Let ν be a computable measure on 2^ω, $A \subseteq 2^\omega \times 2^\omega$, A in Σ_0. The function $x \mapsto \nu A_x$ is of the form $\sum_{k=1}^m c_k 1_{C_k}(x)$, where $C_k \subseteq 2^\omega$, $C_k \in \Sigma_0$, c_k a computable real. In addition, if ν is strongly computable, then the sets $\{a \in \mathbf{Q} \mid c_k < a\}$ and $\{a \in \mathbf{Q} \mid c_k > a\}$ are Δ_1.*

(ii) *Let* ν *be a computable measure on* 2^ω, $A \subseteq 2^\omega \times 2^\omega$, A *in* Σ_1. *Then the set* $\{\langle a, x\rangle \mid a \in \mathbf{Q}, x \in 2^\omega, \nu A_x > a\}$ *is* Σ_1.

(iii) *Let* ν *be a strongly computable measure on* 2^ω. *If* $A \subseteq 2^\omega \times 2^\omega$ *is* Σ_n, *then the set* $\{\langle a, x\rangle \mid a \in \mathbf{Q}, x \in 2^\omega, \nu A_x > a\}$ *is* Σ_n; *if* $A \subseteq 2^\omega \times 2^\omega$ *is* Π_n, *then* $\{\langle a, x\rangle \mid a \in \mathbf{Q}, x \in 2^\omega, \nu A_x > a\}$ *is* Σ_{n+1}.

PROOF. (i) Using, if necessary, a suitable tiling of A, we can write A in the form $\bigcup_{i=1}^m [w^i] \times [v^i]$, where the $[v^i]$ are either disjoint or identical. Define $C_n := \bigcup\{[w^j] \mid \text{for some } i, [w^j] \times [v^n] = [w^i] \times [v^i]\}$. Eliminate the redundancies from the $[v^n]$; then A can be written as $\sum_{k=1}^n 1_{C_k}(x) \cdot 1_{[v^k]}(y)$ and the function $x \mapsto \nu A_x$ as $\sum_{k=1}^n 1_{C_k}(x) \cdot \nu[x^k]$. If ν is computable, $\nu[v^k]$ is a computable real. If ν is strongly computable, the sets $\{a \in \mathbf{Q} \mid a < \nu[v^k]\}$, $\{a \in \mathbf{Q} \mid a > \nu[v^k]\}$ are Δ_1.

(ii) Let $A = \{\langle x, y\rangle \mid \exists n R(n, x, y)\}$, R recursive. Write $A^m := \{\langle x, y\rangle \mid \exists n \leq m\ R(n, x, y)\}$; then A^m is Σ_0. We have

$$\{\langle a, x\rangle \mid \nu A_x > a\} = \{\langle a, x\rangle \mid \exists m (\nu A_x^m > a)\}$$

and the result follows by (i).

(iii) If A is Π_1, then $A = \{\langle x, y\rangle \mid \forall n R(n, x, y)\}$, for some recursive R. If $A^m := \{\langle x, y\rangle \mid \forall n \leq m\ R(n, x, y)\}$, then A^m is Σ_0 and we may write

$$\{\langle a, x\rangle \mid \nu A_x > a\} = \{\langle a, x\rangle \mid \exists \delta > 0 \forall m (\nu A_x^m > a + \delta)\},$$

which set is Σ_2 by (i), if ν is strongly computable. The result now follows by induction on n. ∎

THEOREM 2.1.1. *Let* μ, ν *be strongly computable measures on* 2^ω, $N \subseteq 2^\omega \times 2^\omega$ *a* Π_n $\mu \times \nu$-*nullset. Then* $\{x \mid \nu N_x > 0\}$ *is a* Σ_{n+1} μ-*nullset.*

PROOF. $\{x \mid \nu N_x > 0\} = \{x \mid \exists a \in \mathbf{Q}^+(\nu N_x > a)\}$ is Σ_{n+1} and a μ-nullset by Fubini's theorem.

Theorem 2.1.1 is slightly unsatisfactory, in that one would like to have "Π_n" in the conclusion of the theorem. We do not know whether the above estimate is exact. We can show however, that Σ_{n+1} cannot be replaced by Σ_n.

We construct a Π_2 $\lambda \times \lambda$-nullset in $2^\omega \times 2^\omega$ such that $\{x \mid \lambda N_x > 0\}$ is not contained in a Σ_2 λ-nullset. Let M be a total recursive sequential test which contains $\mathrm{LLN}(\frac{1}{2})^c$ (Lemma 1.3.2). Consider $N := \{\langle x, y\rangle \mid x/y \in M\}$.

By Lemma 1.3.3, N is Π_2; by Lemma 1.2.2, $\lambda N = 0$. Suppose $\{x \mid \lambda N_x > 0\}$ were contained in a Σ_2 B with $\lambda B = 0$. If x is an element of $\mathrm{LLN}(\frac{1}{2})^c$, then by Theorem 1.1.1, $\lambda\{y \mid x/y \in \mathrm{LLN}(\frac{1}{2})^c\} = 1$; hence $\lambda N_x = 1$ and thus $x \in B$. Therefore $B^c \subseteq \mathrm{LLN}(\frac{1}{2})$. But this is impossible, since $\mathrm{LLN}(\frac{1}{2})$ is first category while B^c is residual: the first is obvious and the second statement follows since B^c is a G_δ set which is dense by $\lambda B^c = 1$.

2.2. The following theorem can be proved by formalising Theorem 14.1 in Oxtoby [1].

THEOREM 2.2.1. *Let* μ, ν *be computable measures on* 2^ω, $N \subseteq 2^\omega \times 2^\omega$ *a recursive sequential test with respect to* $\mu \times \nu$. *Then* $\{x \mid \nu N_x > 0\}$ *is contained in a recursive sequential test with respect to* μ.

Since the formalisation is pretty much routine, we shall not bother to write it down here. We do wish, however, to draw attention to a consequence of Theorem 2.2.1 which serves to show that the analogue of Theorem 2.2.1 for *total* recursive sequential tests cannot be derived from it.

Martin-Löf proves (in [1]) the existence of a *universal* recursive sequential test, i.e. a recursive sequential test with respect to a measure μ such that every sequential test with respect to μ is contained in it.

COROLLARY 2.2.1. *Let $U \subseteq 2^\omega \times 2^\omega$ be the universal recursive sequential test with respect to $\lambda \times \lambda$ and $U' \subseteq 2^\omega$ the universal recursive sequential test with respect to λ. Then $U' = \{x \mid \lambda U_x > 0\}$.*

PROOF. By Theorem 2.2.1, $\{x \mid \lambda U_x > 0\} \subseteq U'$. On the other hand, $U' \times 2^\omega \subseteq U$. ∎

Consequently, if N is a recursive sequential test, $\{x \mid \lambda N_x > 0\}$ need not be contained in a *total* recursive sequential test; the latter cannot be universal (cf. Schnorr [1, p. 67]).

2.3. The purpose of this section is to prove

THEOREM 2.3.1. *Let μ, ν be computable measures on 2^ω. Let $N \subseteq 2^\omega \times 2^\omega$ be a total recursive sequential test with respect to $\mu \times \nu$. Then $\{x \mid \nu N_x > 0\}$ is contained in a total recursive sequential test with respect to μ.*

This theorem can presumably be proved by formalising proofs of Fubini's theorem from constructive analysis. However, since we allowed ourselves to use classical logic and mathematics, a more direct approach is possible. The key of the proof consists in the following observation:

If $O \subseteq 2^\omega \times 2^\omega$ is a Σ_1 set such that $\mu \times \nu O$ is computable, and if the distribution function F is defined by

$$F(s) = \mu\{x \mid \nu O_x \leqq s\} \qquad (s \in [0, 1]),$$

then the set of points of continuity of F has a Π_2 definition. Moreover, since this set is dense, it follows from an effective version of the Baire category theorem that F has a recursively enumerable dense set of computable points of continuity. From then on, the going is easy.

The density of the set of points of continuity of F is classically a trivial consequence of the fact that the set of discontinuities of a distribution function is countable, a conclusion that is not available constructively. But although we do not reason constructively, some concepts of constructive measure theory (cf. Bishop and Cheng [1]) have been helpful, in particular the definitions of the functions g_{uv} (cf. Definition 2.3.1) and of a function being "integrable" (cf. Lemma 2.3.1).

We now begin the proof of Theorem 2.3.1. Write $N = \bigcap_n O^n$, $O^n \in \Sigma_1$, $\mu \times \nu O^n$ computable and $< 2^{-n}$; also, $O^{n+1} \subseteq O^n$. Since everything we do is uniform in n, we suppress the superscript n whenever possible. So let O be one of the O^n's. Define on $[0, 1]$ the image measure θ as follows:

$$\theta[0, a] := \mu\{x \mid \nu O_x \leq a\}, \qquad a \in [0, 1].$$

θ need not be a computable measure, but nevertheless some integrals with respect to θ are computable; we shall use this fact to compute $\theta[0, a]$ for a recursively enumerable dense set of computable reals a.

DEFINITION 2.3.1. For rationals u, v in the unit interval, $u < v$, we determine a function g_{uv} as follows:

$$g_{uv}(t) = \begin{cases} 1, & t < u, \\ (v - t)/(v - u), & u \leq t \leq v, \\ 0, & v < t. \end{cases}$$

Let $u_0 < v_0 < u_1 < v_1$ be rationals. The functions $f_{u_0 v_0 u_1 v_1}$ are defined by

$$f_{u_0 v_0 u_1 v_1}(t) = \min\{1 - g_{u_0 v_0}, g_{u_1 v_1}\}.$$

Before we can motivate the introduction of these auxiliary functions, we need a lemma.

LEMMA 2.3.1. *The integrals* $\int g_{uv}\, d\theta$ *and* $\int f_{u_0 v_0 u_1 v_1}\, d\theta$ *are computable, uniformly in the parameters* n, u, v (n, u_0, v_0, u_1, v_1).

PROOF (SKETCH). For the duration of the proof, call an arithmetical function $f\colon 2^\omega \to \mathbf{R}^+$ *integrable* with respect to ρ (abbreviated: integrable (ρ)), if there exists a recursively enumerable sequence of arithmetical functions $f_m\colon 2^\omega \to \mathbf{R}^+$ such that $\sum_m f_m(x) = f(x)$ ρ-a.e. and $\sum_m \int f_m\, d\rho$ is finite and computable. By the monotone convergence theorem, $\int f\, d\rho$ is computable. The point of the definition of integrability is not only to ensure that $\int f\, d\rho$ is computable, but also that we have a canonical algorithm to compute this integral. The following facts are easy to verify.

a) If c is a computable real, f integrable (ρ), then cf is integrable (ρ).

b) If f, g are integrable (ρ), so is $f + g$.

c) If f, g are integrable (ρ) with canonical sequences $\{f_n\}$, $\{g_n\}$ and if for each n, $\min(\sum_{k=1}^n f_k, \sum_{k=1}^n g_k)$ is integrable (ρ), then so is $\min(f, g)$.

It is straightforward to show that the functions g_{uv} and $f_{u_0 v_0 u_1 v_1}$ are integrable (θ), uniformly in the parameters, since these functions can be written by means of the operations mentioned above. For example,

$$\int g_{uv}\, d\theta = \int g_{uv} \circ vO_x\, d\mu(x)$$

and

$$g_{uv} \circ vO_x = \min\{1, (v - \min(vO_x, v))/(v - u)\},$$

where $\circ$ denotes functional composition. ■

Now consider a computable real a and rationals u_0, v_0, u_1, v_1 such that $u_0 < v_0 < a < u_1 < v_1$. Obviously,

$$\int g_{u_0 v_0}\, d\theta \le \theta[0, a] \le \int g_{u_1 v_1}\, d\theta$$

and by the preceding lemma the left and right terms are computable. What remains to be done is to find a computable estimate of the difference

$$\int g_{u_1 v_1}\, d\theta - \int g_{u_0 v_0}\, d\theta.$$

For special a, this can be achieved using the function f_{----}.

DEFINITION 2.3.2. $a \in [0, 1]$ is an *atom of* θ iff $\theta\{a\} > 0$.

$a \in [0, 1]$ is a *point of continuity of* θ (abbreviated: p.c. of θ) iff $\theta\{a\} = 0$.

The key of the proof of Theorem 2.3.1 is that the set of p.c.'s of all θ_n has Π_2 definition (where θ_n equals θ with O replaced by O^n).

LEMMA 2.3.2. *a is p.c. of all θ_n iff*

$$(*)\quad \forall n \forall \varepsilon > 0 \exists \delta > 0 \exists u_0 v_0 u_1 v_1\ (v_0 < a - \delta < a + \delta < u_1\ \&\ \int f_{u_0 v_0 u_1 v_1}\, d\theta_n < \varepsilon),$$

where the quantifiers "$\forall\varepsilon$" and "$\exists\delta$" range over the rationals. Moreover, $(*)$ *is a* Π_2 *statement.*

PROOF. The first statement is obvious as soon as we realise that the condition "$v_0 < a - \delta < a + \delta < u_1$" in (*) means that $f_{u_0 v_0 u_0 v_1}$ equals 1 on $(a - \delta, a + \delta)$. The second statement follows from Lemma 2.3.1. ■

The Π_2 definition of "a is p.c. of all θ_n" enables us to apply the following effective version of the Baire category theorem:

LEMMA 2.3.3. *Let G be a Π_2 subset of* $[0, 1]$, $G = \bigcap_n A_n$, A_n *in* Σ_1 *and dense in* $[0, 1]$. *Then G has a recursively enumerable dense subset of computable reals.*

PROOF. Formalise a proof of the Baire category theorem (e.g. Oxtoby [1, p. 2]). ■

Combining these lemmas, we get

LEMMA 2.3.4. *There exists a recursively enumerable dense set $\underline{D}$ of computable points of continuity of all θ_n.*

PROOF. By Lemma 2.3.2 the set of p.c.'s of all θ_n is Π_2. This set is dense in $[0, 1]$, since the set of atoms for some θ_n is countable (this reasoning is nonconstructive). Now apply Lemma 2.3.3. ■

We are now almost done.

LEMMA 2.3.5. *Let $a \in [0, 1]$ be a computable p.c. of θ. Then $\theta[0, a]$ is computable.*

PROOF. Choose $\varepsilon > 0$. We must effectively determine $u < v < \underline{u} < \underline{v}$ such that

$$(1) \qquad \int g_{uv}\, d\theta \le \theta[0, a] \le \int g_{\underline{u}\underline{v}}\, d\theta,$$

$$(2) \qquad \int g_{\underline{u}\underline{v}}\, d\theta - \int g_{uv}\, d\theta < \varepsilon.$$

Choose recursively enumerable sequences of rationals $\{b^k\}$, $\{c^k\}$ such that $b^k < a < c^k$ and $c^k - b^k < 2^{-k}$. By Lemma 2.3.2, there exist $\delta > 0$ and $u_0 < v_0 < u_1 < v_1$ such that $v_0 < a - \delta < a + \delta < u_1$ and $\int f_{u_0 v_0 u_1 v_1}\, d\theta < \varepsilon$.

Choose k large enough so that $a - b^k < \delta/4$, $c^k - a < \delta/4$. Define $u := b^k - \delta/4$, $v := b^k$, $\underline{u} := c^k$ and $\underline{v} := c^k + \delta/4$. Then $v_0 < u < v < a < \underline{u} < \underline{v} < u_1$; hence

$$\int g_{uv}\, d\theta \le \theta[0, a] \le \int g_{\underline{u}\underline{v}}\, d\theta,$$
$$\int g_{\underline{u}\underline{v}}\, d\theta - \int g_{uv}\, d\theta < \int f_{u_0 v_0 u_1 v_1}\, d\theta < \varepsilon. \quad ■$$

Let $\underline{D}$ be the set constructed in Lemma 2.3.4. Theorem 2.3.1 follows if we can show that the set

$$\bigcup_{a \in \underline{D}} \bigcap_n \{x \mid \nu O_x^n > a\}$$

is contained in a total recursive sequential test with respect to μ. Now for computable ν, the function $x \mapsto \nu O_x^n$ satisfies (cf. Lemma 2.1.1)

$$\{\langle x, b\rangle \mid \nu O_x^n > b\} \in \Sigma_1;$$

hence in particular for $a \in \underline{D}$,

$$\{x \mid \nu O_x^n > a\} \in \Sigma_1.$$

Moreover, since the $\mu\{x \mid \nu O_x^n > a\}$ are computable, uniformly in n, and $\mu \bigcap_n \{x \mid \nu O_x^n > a\} = 0$ by Fubini's theorem, we can determine a recursively enumerable infinite sequence $\{n_k\}$ of natural numbers such that, for all k,

$$\mu\{x \mid \nu O_x^{n_k} > a\} < 2^{-k}.$$

Because $O^{n+1} \subseteq O^n$,

$$\bigcap_n \{x \mid \nu O_x^n > a\} = \bigcap_k \{x \mid \nu O_x^{n_k} > a\};$$

and $\bigcap_k \{x \mid \nu O_x^{n_k} > a\}$ is a total recursive sequential test with respect to μ.

By Lemma 1.3.1, the union of these tests over a in $\underline{D}$ is contained in a total recursive sequential test with respect to μ. But this union equals $\{x \mid \nu N_x > 0\}$. This concludes the proof of Theorem 2.3.1. ■

§3. Main results. A set E in 2^ω has *absolute measure zero* if for every finite nonatomic measure μ on 2^ω we can find a Borel set A such that $E \subseteq A$ and $\mu A = 0$. (Hausdorff constructed a nontrivial example of such a set). This concept can be transferred to the constructive realm as follows: E is *constructively small* if for every *computable* finite nonatomic measure μ we can find a Borel set A such that $E \subseteq A$ and $\mu A = 0$. Theorems 3.2 and 3.3 will show that if $x \in R(\lambda)$ (either Martin-Löf's or Schnorr's notion), the set $\{y \mid x/y \notin R(\lambda)\}$ is constructively small. (In another sense, these sets are quite large, because they are residual.) For completeness' sake, we begin with n-randomness (cf. Definition 0.7.4).

We recall that for $p \in [0, 1]$, the measure μ_p is defined to be $(1 - p, p)^\omega$. A measure of this form is computable (strongly computable) if p is a computable (Δ_1 definable) real.

THEOREM 3.1. *Let ν be an arbitrary strongly computable measure, and p Δ_1 definable. If x is n-random ($n \geq 2$) with respect to μ_p, then $\nu\{y \mid x/y$ is n-random with respect to $\mu_p\} = 1$.*

PROOF. It suffices to show that for each Π_n μ_p-nullset N, the set $\{x \mid \nu\{y \mid x/y \in N\} > 0\}$ is contained in a Σ_{n+1} μ_p-nullset.

So let N be a Π_n μ_p-nullset. By Lemmas 1.2.2 and 1.3.3, $M := \{\langle x, y\rangle \mid x/y \in N\}$ is a Π_n subset of $2^\omega \times 2^\omega$ which has $\mu_p \times \nu$ measure 0. By Theorem 2.1.1, $\{x \mid \nu M_x > 0\}$ is contained in a Π_n μ_p-nullset. ■

THEOREM 3.2. *Let p be a computable real and ν a computable, but otherwise arbitrary, measure. Let $R(\mu_p)$ be Martin-Löf's definition of randomness. Then $x \in R(\mu_p)$ implies $\nu\{y \mid x/y \in R(\mu_p)\} = 1$.*

PROOF. Similar to that of Theorem 3.1. If N is a recursive sequential test, $N = \bigcap_n O_n$, $\mu_p O_n < 2^{-n}$, O_n in Σ_1, then, by Lemma 1.3.3, $\{\langle x, y\rangle \mid x/y \in O_n\}$ is Σ_1 and, by Lemma 1.2.2, $\mu_p \times \nu\{\langle x, y\rangle \mid x/y \in O_n\} < 2^{-n}$. Now apply Theorem 2.2.1. ■

THEOREM 3.3 *If p, ν are as in the preceding theorem and $R(\mu_p)$ is Schnorr's definition of randomness, then $x \in R(\mu_p)$ implies $\nu\{y \mid x/y \in R(\mu_p)\} = 1$.*

PROOF. If $\mu_p O_n$ is computable, so is (by Lemma 1.2.2) $\mu_p \times \nu\{\langle x, y\rangle \mid x/y \in O_n\}$. Now apply Theorem 2.3.1. ■

The operation / may be replaced by a composition of a recursive and a random selection: if Φ is a recursive place selection such that $\mu_p(\mathrm{dom}(\Phi)) = 1$, define ${}^\Phi/$ by $(x^\Phi/y)_k = x_m$ iff m is the index of the kth 1 in y and $\phi(x(m - 1)) = 1$.

This is so, since ${}^\Phi/ : 2^\omega \times 2^\omega \to 2^\omega$ also satisfies Lemmas 1.2.2 and 1.3.3.

In §0 we stated that our main result is related to von Mises' original proposal as statistical mechanics is to thermodynamics. We now explain what we mean by this statement.

We must emphasize that the theorems above do not distinguish between admissible and inadmissible place selections. Every strictly increasing sequence of natural numbers (every $y \in 2^{\omega}$) is taken to represent a place selection. Hence essential elements of the randomness axiom get lost, but given this proviso, we may sketch the interpretation that Theorems 3.1–3.3 give to the randomness axiom as follows: if x is random and if M is a probabilistic device that generates sequences y according to some computable probability distribution, then, for all practical purposes, x/y is random.

Loosely speaking, Theorems 3.1–3.3 thus say that the randomness axiom is true in a statistical sense, with probability one.

Such an interpretation of the randomness axiom is of course anathema to von Mises. Clearly the foundation of von Mises' theory via Theorems 3.1–3.3 and the foundation of thermodynamics in statistical mechanics suffer from the same defect in that the absolute character of the theories in question gets lost. In the case of von Mises' theory this problem is much more serious, since what is at stake is a foundation for probability itself.

In this respect, Theorems 3.1–3.3 do not bring us any nearer to a thoroughgoing frequency interpretation of probability; we are still forced to use a correspondence rule which closes the gap between the formal statement "A has probability one" and "for all practical purposes, A is certain to happen".

In §5 we shall see that a more satisfactory formulation of the randomness axiom is possible if we combine Theorems 3.1–3.3 with the ideas of Kolmogorov complexity.

§4. New proof of a theorem of Ville. In 4.1 we sketch a proof of the fact that sequences random in the sense of Schnorr satisfy the law of the iterated logarithm. We give an effective form of the first Borel-Cantelli lemma (Lemma 4.1.1); from then onwards, the proof of the law of the iterated logarithm in Feller [1, p. 205ff.] can be copied. The simple argument of Lemma 4.1.1 will be used again in 4.2. There we give a new proof of Ville's theorem, stated in §0, and comment on its significance.

4.1. To show that sequences random in the sense of Schnorr satisfy the law of the iterated logarithm (henceforth abbreviated LIL), it suffices to show that the sequences not satisfying LIL are contained in a total recursive sequential test.

Recall that LIL (for λ) consists of two parts:

(a) For $\alpha > 1$,

$$\lambda\left\{x \mid \forall k \exists n \geq k \left(\sum_{j=1}^{n} x_j - \frac{1}{2}n\right) > \alpha\sqrt{\tfrac{1}{2}n\operatorname{loglog} n}\right\} = 0.$$

(b) For $\alpha < 1$,

$$\lambda\left\{x \mid \exists k \forall n \geq k \left|\sum_{j=1}^{n} x_j - \frac{1}{2}n\right| \leq \alpha\sqrt{\tfrac{1}{2}n\operatorname{loglog} n}\right\} = 0$$

and

$$\lambda\left\{x \mid \exists k \forall n \geq k \left(\frac{1}{2}n - \sum_{j=1}^{n} x_j\right) \leq \alpha\sqrt{\tfrac{1}{2}n\operatorname{loglog} n}\right\} = 1.$$

Part (b) trivially corresponds to a total recursive sequential test, since it involves a Σ_2 λ-nullset, hence a recursively enumerable union of total recursive sequential tests.

For part (a), use the proof in Feller [1, p. 205] and the following effective version of the (first) Borel-Cantelli lemma:

LEMMA 4.1.1. *Let μ be a computable measure on 2^ω. Suppose there exists a recursively enumerable sequence of events A_k in Σ_0 and a recursively enumerable sequence of computable reals a_k such that $\mu A_k < a_k$ and $\sum_{k=1}^{\infty} a_k$ is finite and computable. Then $\bigcap_n \bigcup_{k \geq n} A_k$ is a total recursive sequential test with respect to μ.*

PROOF. Obviously, $\bigcap_n \bigcup_{k \geq n} A_k$ is Π_2 and $\lim_{n \to \infty} \mu \bigcup_{k \geq n} A_k = 0$. We have to show that $\mu \bigcup_{k \geq n} A_k$ is computable. It suffices to show that the sequence $\{\mu \bigcup_{k=n}^{m} A_k\}_{m=n+1}^{\infty}$ is recursively Cauchy, i.e. that there exists a recursive function g such that

$$\forall i \forall m, m' \geq g(i) \left| \mu \bigcup_{k=n}^{m'} A_k - \mu \bigcup_{k=n}^{m} A_k \right| < 2^{-i}.$$

Now if $m < m'$, then

$$\mu \bigcup_{k=n}^{m'} A_k - \mu \bigcup_{k=n}^{m} A_k = \mu \bigcup_{k=m}^{m'} A_k \leq \sum_{k=m}^{m'} \mu A_k \leq \sum_{k=m}^{m'} a_k;$$

choose recursive g such that for all i and all $m' \geq m \geq g(i)$, $\sum_{k=m}^{m'} a_k < 2^{-i}$.

The proof of the LIL in Feller then yields A_k, a_k which satisfy the condition of Lemma 4.1.1.

4.2. We prove Ville's theorem in the following form:

THEOREM 4.2. *Let $R(\lambda)$ and $C(\frac{1}{2})$ be Schnorr's and Church's notions of randomness, respectively. There exists a nonatomic computable measure μ such that $\mu(R(\lambda)^c \cap C(\frac{1}{2})) = 1$. A fortiori, $R(\lambda)^c \cap C(\frac{1}{2})$ has the cardinality of the continuum.*

The measure μ will be a computable product measure of the form $\Pi_n(1 - p_n, p_n)$, such that $\lim_{n \to \infty} p_n = \frac{1}{2}$ and $\mu \perp \lambda$. In fact, the proof shows that for *any* such measure μ, $\mu(R(\lambda)^c \cap C(\frac{1}{2})) = 1$.

LEMMA 4.2.1. *Let $\mu = \Pi_n(1 - p_n, p_n)$ be a computable product measure. Then $\mu C(\frac{1}{2}) = 1$ iff $\lim_{n \to \infty} p_n = \frac{1}{2}$.*

PROOF. $\rightarrow$. Suppose not. For some rational $\varepsilon > 0$, one of the sets $\{n \mid p_n > \frac{1}{2} + \varepsilon\}$ or $\{n \mid p_n < \frac{1}{2} + \varepsilon\}$ is infinite, say the first set. By the computability of μ, this set is recursively enumerable, hence contains an infinite recursive subset. Using this subset, we can define a recursive place selection Φ such that $\mu \Phi^{-1}(\text{LLN}(\frac{1}{2})) = 0$, a contradiction.

$\leftarrow$. For this direction, no assumption of recursiveness or computability is needed. So let Φ be a place selection, and μ a measure of the form $\Pi_n(1 - p_n, p_n)$, such that $\lim_{n \to \infty} p_n = \frac{1}{2}$. We show that $\mu \Phi^{-1}(\text{LLN}(\frac{1}{2})) = 1$, using the martingale convergence theorem (Feller [2, p. 242]).

Given Φ and its generating function ϕ (as in Definition 0.4.3), we define a function $\theta: 2^\omega \times \omega \to \omega$ as follows: $\theta(x,n) := 1 +$ smallest k such that $n = \sum_{j=1}^{k} \phi(x(j))$. Define random variables $Z_n: 2^\omega \to \mathbf{R}$ by

$$Z_n(x) = x_{\theta(x,n)} \cdot (p_{\theta(x,n)})^{-1} - 1.$$

If E_μ and $E_\mu(\cdot\cdot \mid \cdot\cdot)$ denote the expectation and conditional expectation with respect to μ, then we have $E_\mu(Z_n) = 0$ and $E_\mu(Z_n \mid Z_1, \ldots, Z_{n-1}) = 0$. It follows that the random variables $U_n := \sum_{k=1}^{n} Z_k/k$ form a martingale: $E_\mu(|U_n|)$ is bounded and

$$E_\mu(U_{n+1} \mid U_1, \ldots, U_n) = \sum_{k=1}^{n} \frac{1}{k} E_\mu(Z_k \mid U_1, \ldots, U_n) = \sum_{k=1}^{n} \frac{1}{k} Z_k = U_n.$$

Note that, since for some M, the $|Z_n|$ are uniformly bounded by M, we have $\sum_{k=1}^{n} E_\mu(Z_k^2)/k^2 < \infty$. Moreover, for each n, $E_\mu(U_n) \leq \sum_{k=1}^{n} E_\mu(Z_k^2)/k^2$, since all terms of U_n which contain a factor of the form $Z_i Z_j \, (i \neq j)$ have zero expectation. By the martingale convergence theorem, there exists a random variable U such that U_n converges to U μ-a.e.

By Kronecker's lemma (Feller [2, p. 239]), $(1/n)\sum_{k=1}^{n} Z_k$ converges to 0 μ-a.e. Hence for μ-a.a. x

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} x_{\theta(x,k)} \cdot (p_{\theta(x,k)})^{-1} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\Phi x)_k \cdot (p_{\theta(x,k)})^{-1} = 1.$$

The following observation then concludes the proof:

Let $\{a_n\}$, $\{b_n\}$ be sequences of positive reals such that

$$\lim_{n\to\infty} b_n = \frac{1}{2}, \qquad \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{a_k}{b_k} = 1.$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = \frac{1}{2}. \qquad \blacksquare$$

The "$\rightarrow$" direction of the preceding lemma is a special case of a more general theorem which reads as follows. Let μ be any measure on 2^ω. If for all place selections Φ such that ϕ recursive in μ and $\mu(\text{dom } \Phi) = 1$ we have $\mu\Phi^{-1}(\text{LLN}(\frac{1}{2})) = 1$, then μT^{-n} converges weakly to λ. We shall not prove this theorem here, but continue to work towards Theorem 4.2.

LEMMA 4.2.2. *Let μ, ν be computable measures. Then $\mu \perp \nu$ is equivalent to either one of the following statements:*

(i) *There exists a total recursive sequential test N such that $\nu N = 1$ and $\mu N = 0$.*

(ii) *For each rational $\varepsilon > 0$, there exists a Π_1 set A such that $\nu A > 1 - \varepsilon$ and $\mu A = 0$.*

PROOF. Trivially, (i), (ii) imply $\mu \perp \nu$. For $\mu \perp \nu$ implies (i), we use the following equivalence:

$$\mu \perp \nu \quad \text{iff} \quad \forall \varepsilon > 0 \exists C \in \Sigma_0 (\nu C > 1 - \varepsilon \;\&\; \mu C < \varepsilon),$$

and we take advantage of the Π_2 statement on the right-hand side.

Let $f\colon \mathbf{Q}^+ \to \Sigma_0$ be a recursive function which, for each ε in $\mathbf{Q}^+$, gives $f(\varepsilon)$ in Σ_0 such that $\nu(f(\varepsilon)) > 1 - \varepsilon$ and $\mu(f(\varepsilon)) < \varepsilon$. Let $N = \bigcap_n \bigcup_i f(2^{-i-n-1})$. Obviously N is Π_2. Since for each i and n, $\mu(f(2^{-i-n-1}) < 2^{-i-n-1}$, it follows that $\mu \bigcup_i f(2^{-i-n-1})$ is computable (see the proof of Lemma 4.1.1) and smaller than 2^{-n}. On the other hand, for each n and all i,

$$\nu \bigcup_i f(2^{-i-n-1}) \geq \nu(f(2^{-i-n-1})) \geq 1 - 2^{-i-n-1},$$

so $\nu \bigcup_i f(2^{-i-n-1}) = 1$.

For (i) implies (ii), reverse the roles of μ and ν in (i), obtaining $N = \bigcap_n O_n$ such that $\mu_n = 1$, $\nu_n = 0$ and each $O_n \in \Sigma_1$; then some O_n^c will do. ■

This lemma is interesting in its own right. It is a kind of adequacy condition for the formalisation of sequential tests, since it shows that singular computable statistical hypotheses can indeed be separated by (total) recursive sequential tests.

The following beautiful criterion for singularity of product measures is due to Kakutani [1].

LEMMA 4.2.3 *Let* $\mu = \prod_n(1 - p_n, p_n)$ *and* $\nu = \prod_n(1 - q_n, q_n)$ *be product measures on* 2^ω *such that for some* $\delta > 0$, *for all* n, $\delta < p_n, q_n < 1 - \delta$. *Then either* $\mu \perp \nu$ *or* $\mu \sim \nu$, *corresponding to the divergence or convergence of* $\sum_n(p_n - q_n)^2$.

It follows from the 0–1 law that product measures on 2^ω are either singular or equivalent, but Kakutani's theorem provides us with a criterion to distinguish these cases, and this is what we shall use to finish the proof of Theorem 4.2.

Let $p_n := \frac{1}{2}(1 + (n + 1)^{-1/2})$ and $\mu = \prod_n(1 - p_n, p_n)$; then μ is computable and, since $\sum_n(\frac{1}{2} - p_n)^2 = \frac{1}{4}\sum_n 1/n = \infty$, $\mu \perp \lambda$. Lemmas 4.2.1 and 4.2.2 now imply $\mu(R(\lambda)^c \cap C(\frac{1}{2})) = 1$. ■

This result allows us to view the relationship between $C(\frac{1}{2})$ and $R(\lambda)$ in a different light. We first state an analogue of Lemma 4.2.1 for $R(\lambda)$.

For the definitions of "$\mu \ll \lambda$", the mapping T and weak and strong convergence of measures, see the Appendix.

LEMMA 4.2.4. *Let* μ *be a computable measure, and* $R(\lambda)$ *Schnorr's notion of randomness. Then the following are equivalent*:

(i) $\mu R(\lambda) = 1$.

(ii) $\mu \ll \lambda$.

(iii) μT^{-n} *converges strongly to* λ.

PROOF (SKETCH). (i) implies (ii): it follows from the proof of Lemma 3.14 in Gaifman and Snir [1] that, if $\mu \not\ll \lambda$, for some total recursive sequential test N, $\mu N > 0$ and $\lambda N = 0$. The implication from (ii) to (i) is trivial. The equivalence of (ii) and (iii) follows from results in ergodic theory (e.g. the fact that T is strongly mixing).

Combining the preceding lemmas, we get ■

COROLLARY 4.2. *Let* $\mu = \prod_n(1 - p_n, p_n)$ *be a computable product measure.*

(i) $\mu C(\frac{1}{2}) = 1$ *iff* $\lim_{n\to\infty} p_n = \frac{1}{2}$ *iff* μT^{-n} *converges weakly to* λ.

(ii) $\mu R(\lambda) = 1$ *iff* $\sum_n(\frac{1}{2} - p_n)^2$ *converges iff* μT^{-n} *converges strongly to* λ.

(iii) $\mu(R(\lambda)^c \cap C(\frac{1}{2})) = 1$ *iff* $\lim_{n\to\infty} p_n = \frac{1}{2}$ *but* $\sum_n(\frac{1}{2} - p_n)^2$ *diverges*.

It follows that processes on 2^ω yielding with probability one outcome sequences in $C(\frac{1}{2})$ or in $R(\lambda)$ differ in the speed of convergence to $\frac{1}{2}$. A total recursive sequential test can detect whether $\lim_{n\to\infty} p_n = \frac{1}{2}$ but $\sum_n(\frac{1}{2} - p_n)^2 = \infty$; necessarily such a test cannot be of the form $\mu\Phi^{-1}(\text{LLN}(\frac{1}{2}))$ for some recursive place selection Φ. On the other hand, no total recursive sequential test can detect the difference between a $\mu = \prod_n(1 - p_n, p_n)$ such that $\sum_n(\frac{1}{2} - p_n)^2 < \infty$ and λ itself. Seen from this perspective, the notion of randomness $R(\lambda)$ is better than $C(\frac{1}{2})$, but only slightly so.

§5. Irregularity, independence, zero entropy. In this section we shall take a fresh look at results found in the previous sections, interpreting these results via complexity. Our main tool is the match between irregularity and randomness, announced in 0.8, a proof of which will be given presently (5.1). The connection between complexity (of a sequence) and entropy (of a measure) is derived as a consequence of Theorem 5.1 in 5.2. In 5.3 we look back upon von Mises' axioms and show that combining Theorem 2.2.1 with Theorem 5.1 goes a long way towards making sense of the enigmatic randomness condition. Lastly, we study sequences y with $\lim_{n\to\infty} I(y(n))/n = 0$, especially with regard to their virtues as place selections. We use Martin-Löf's definition of randomness throughout.

5.1. The object of this subsection is to prove

THEOREM 5.1. *Let μ be a computable measure on 2^ω. Then $x \in R(\mu)$ iff there exists m such that for all n*

$$I(x(n)) > [-\log_2 \mu[x(n)]] - m.$$

Here, we interpret $-\log_2 0$ as $+\infty$.

The proof is surprisingly simple and rests essentially upon two facts: the existence of a universal recursive sequential test, and the definition of I using prefixalgorithms. A technical advantage of prefixalgorithms resides in the circumstance that they are governed by the Kraft inequality. Note that if f is a prefixalgorithm, $\sum\{2^{-p} \mid f(p) \text{ defined}\} \leq 1$. Part (a) of the following lemma takes care of the converse.

LEMMA 5.1. (a) *Let $S \subseteq 2^{<\omega} \times \omega$ be an r.e. set such that $\sum\{2^{-m} \mid \langle w,m\rangle \in S\}$ is less than 1. Then there exists a prefixalgorithm f with the property*

$$\langle w,m\rangle \in S \quad \textit{iff} \quad \textit{for some } p, |p| = m \textit{ and } f(p) = w.$$

(b) *Simulating f on the universal machine U, we see that for all $\langle w,m\rangle \in S$, $I(w) \leq m + \ulcorner f \urcorner + 1$.*

For a proof, see Chaitin [1, p. 333]. We now proceed to the proof of Theorem 5.1.

"*Only if*". We note first that, since the universal machine U is a prefixalgorithm, $\sum_w 2^{-I(w)} \leq 1$. Now consider

$$\{x \mid \forall m \exists n\, I(x(n)) \leq [-\log_2 \mu[x(n)]] - m\};$$

we must show this set to be a recursive sequential test with respect to μ. Clearly it is Π_2; hence it suffices to show that

$$\mu\{x \mid \exists n\, I(x(n)) \leq [-\log_2 \mu[x(n)]] - m\} \leq 2^{-m}.$$

Evidently,

$$\begin{aligned}&\mu\{x \mid \exists n\, I(x(n)) \leq [-\log_2 \mu[x(n)]] - m\}\\ &\leq \sum\{\mu[w] \mid w \in 2^{<\omega}, I(w) \leq [-\log_2 \mu[w]] - m\}.\end{aligned}$$

But if $I(w) \leq [-\log_2 \mu[w]] - m$, then $2^{-I(w)-m} \geq \mu[w]$. The above sum is therefore majorized by

$$\sum\{2^{-I(w)-m} \mid I(w) \leq [-\log_2 \mu[w]] - m\} \leq 2^{-m} \cdot \sum_w 2^{-I(w)} \leq 2^{-m}.$$

"*If*". Let N be the universal recursive sequential test with respect to μ. By abuse of notation, let N also denote the set of pairs $\langle w, m\rangle$ such that $x \in N$ iff $\forall m \exists n$ $\langle x(n), m\rangle \in N$. Put $N_m := \{w \mid \langle w,m\rangle \in N\}$. We may assume that N_m are prefixfree (a simple construction will make them so). In this case,

$$\sum_{w \in N_m} \mu[w] = \sum_{w \in N_m} \mu[w] \leq 2^{-m}.$$

Strings w which occur in some N_m will turn out to have fairly low complexity. In particular, we show that for some constant c, if $\langle w,m\rangle \in N$, then $I(w) \leq [-\log_2 \mu[w]] - \frac{1}{2}m + 2 + c$. This estimate is a consequence of Lemma 5.1(b), since

$$\sum_m \sum_{w \in N_m} \mu[w] \cdot 2^{m/2-2} \leq \frac{1}{4} \sum_m 2^{-m/2} \leq 1.$$

Hence, if x is in N, i.e. if $\forall m \exists n\, x(n) \in N_m$, we have

$$\forall m \exists n\, I(x(n)) \leq [-\log_2 \mu[x(n)]] - \tfrac{1}{2}m + 2 + c$$

which implies that if $\exists m \forall n\, I(x(n)) > [-\log_2 \mu[x(n)]] - m$, then x is in $R(\mu)$. ■

The author recently noticed that a similar proof can be pieced together from two articles by Dies [1]).

5.2. Perhaps somewhat surprisingly, the characterisation of randomness involves constraints on the lower bound of I only. An upper bound can be derived from Lemma 5.1.

LEMMA 5.2. *Let μ be a computable measure on 2^ω. Then for some constant c, $\forall w$*

$$I(w) \leq [-\log_2 \mu[w]] + 2\log_2 |w| + 1 + c.$$

PROOF.

$$\sum_w \mu[w] \cdot |w|^{-2} \cdot 2^{-1} = \frac{1}{2} \sum_n n^{-2} \leq 1.$$ ■

In 0.8 we introduced the expression $I(x(n))/n$ (or rather $K(x(n))/n$), called the *compression coefficient*. Combining the upper and lower bounds on I, we shall derive a theorem on the asymptotic behaviour of $I(x(n))/n$, which is an analogue of the classical noiseless coding theorem.

THEOREM 5.2. *Let μ be an ergodic computable measure on 2^ω with entropy $H(\mu)$. Then for μ-almost all x in 2^ω,*

$$\lim_{n\to\infty} I(x(n))/n = H(\mu).$$

PROOF. If μ is ergodic, the Shannon-McMillan-Breiman theorem (see e.g. Petersen [1, p. 259]) states that for μ-a.a. x,

$$\lim_{n\to\infty} -\log_2 \mu[x(n)]/n = H(\mu).$$

The result then follows from Theorem 5.1 and Lemma 5.2. ■

This improves a result of Brudno [1, p. 139]. For noncomputable measures Theorem 5.2 still holds, but the proof is more elaborate, since we cannot profit from Lemma 5.2.

Theorem 5.2 sheds further light on the notion of a random sequence. The noiseless coding theorem of information theory states that $H(\mu)$ is the greatest lower bound of the average compression of all possible coding schemes (if the statistics of the source are described by μ). We thus see that a random sequence is one which cannot be coded more efficiently than on an average. Since $H(\mu)$ can be approximated arbitrarily well using coding procedures which require only the statistical structure of the given sequence, a loose rendering of Theorems 5.1 and 5.2 might read as follows: *A sequence x is random with respect to μ iff it contains only statistical regularities with respect to μ.* More regularities would cause the compression coefficient to drop considerably. This is indeed one of the basic intuitions behind randomness; we can find it for instance in von Mises [1, p. 59], where he claims to have proven:

"SATZ 5. *Ein Kollektiv wird durch seine Verteilung volkommen bestimmt; die vollständige Angabe der Zuordnung im einzelnen ist nicht möglich.*"

From this perspective, we can also understand the often repeated query: "How can a random sequence exhibit statistical regularities, since randomness entails the absence of regularities?" In a sense, the implied objection is right; we might even concede that it is illustrated by the failure of our putative definition of irregularity, namely $\exists m \forall n\, I(x(n)) > n + I(n) - m$. This definition turned out to be unsatisfactory (i.e. not matching our previous definition of randomness) precisely because a statistical regularity brought about a decrease of I (see the remarks following Definition 0.8.6).

We see, however, that some regularities are more regular than others; in particular statistical regularities are not simple, that is, they do not lead to a significant decrease in complexity.

5.3. We now use complexity to shed light on the theorems in §§2 and 3 and the philosophical considerations, expounded in §0, which gave rise to them. In our comments on Theorems 3.1–3.3 we granted that the difference between admissible and nonadmissible place selections does not show up in the statement of these theorems. Can we do better than that?

The main motivation for the technical work in §§2 and 3 was an attempt to elucidate von Mises' phrase "Auswahl ohne Benützung der Merkmalunterschiede". The difficulties encountered in making this phrase mathematically acceptable ultimately led Church to propose as a canonical set of place selections (admissible by definition) those determined by recursive functions $\phi: 2^{<\omega} \to \{0, 1\}$. In this way, the originally intended character of admissibility, namely as a relation of independence between Kollektiv and selection procedure, disappeared and a new problem arose, to wit: "What's so special about the recursively defined place selections?"

We shall try to reinstate the relational character of admissible choice and to show that what matters about place selections is not so much recursiveness as having "zero entropy".

As to the first task, we have to answer the question: when is a strictly increasing sequence of integers, i.e. a y in 2^ω, an admissible place selection with respect to a Kollektiv x? Combining von Mises' axiom, complexity theory and wishful thinking, we are led to consider the possibility that y is admissible with respect to x if y does not contain any information about x. Several alternatives to formalise this statement suggest themselves: via conditional complexity $I(x(n) \mid y(k))$ or using I^y, that is, the information function defined with perfixalgorithms recursive in y. We choose the latter possibility and define

DEFINITION 5.3.1 If $x \in R(\mu_p)$ then y is an *admissible place selection with respect to* x if

$$\exists m \forall n\, I^y(x(n)) > [-\log_2 \mu_p[x(n)]] - m.$$

A definition of the form: y is an admissible place slection with respect to x if $\exists m \forall n\, I^y(x(n)) > I(x(n)) - m$ would be rather more elegant, but I have as yet not been able to show that, e.g., for all x in $R(\lambda)$,

$$\lambda\{y \mid \exists m \forall n\, I^y(x(n)) > I(x(n)) - m\} = 1.$$

It now easily follows that

(a) *If* y is admissible with respect to x and $x \in R(\mu_p)$, then $x/y \in R(\mu_p)$.

(b) The set of place selections y not admissible with respect to x is constructively small (cf. §3).

For (a), note that we may relativise Theorem 5.1 as follows:

$\exists m \forall n\, I^y(x(n)) > [-\log_2 \mu_p[x(n)]] - m$ *iff* $x \in R^y(\mu_p)$ (the latter set with the obvious definition).

For (b), we have to show that if $x \in R(\mu_p)$, then for every computable ν

$$\nu\{y \mid \exists m \forall n\, I^y(x(n)) > [-\log_2 \mu_p[x(n)]]^p - m\} = 1.$$

By Theorem 2.2.1, it suffices to show that

$$\{\langle x, y\rangle \mid \forall m \exists n\, I^y(x(n)) \leq [-\log_2 \mu_p[x(n)]] - m\}$$

is a recursive sequential test with respect to $\mu_p \times \nu$.

The Π_2 form is obvious, and

$$\begin{aligned}&\mu_p \times \nu\{\langle x, y\rangle \mid \exists n\, I^y(x(n)) \leq [-\log_2 \mu_p[x(n)]] - m\}\\ &= \int \mu_p\{x \mid \exists n\, I^y(x(n)) \leq [-\log_2 \mu_p[x(n)]] - m\}\, d\nu(y) \leq \int 2^{-m}\, d\nu(y) = 2^{-m},\end{aligned}$$

the inequality following from the relativised version of Theorem 5.1. A trivial combination of Theorems 2.2.1 and 5.1 thus allows us to capture at least some of the content of the randomness axiom. (Note that Theorem 3.3 also follows from (a) and (b).)

Suppose we now ask for a characterization of those place selections y which are absolutely admissible, i.e. y such that for all x in $R(\lambda)$ (say), x/y is also in $R(\lambda)$. Recursive sequences are absolutely admissible and they should be, since it seems plausible that a recursive sequence cannot contain any information about a Kollektiv. In fact, a recursive sequence is an example of a sequence which does not contain any information at all (asymptotically), as $\lim_{n\to\infty} I(y(n))/n = 0$. Sequences y satisfying $\lim_{n\to\infty} I(y(n))/h = 0$ exist in profusion (they arise for instance from irrational rotations of the circle) and one might conjecture that it is this property, rather than being r.e., that qualifies a sequence to be an absolutely admissible place selection. There exist results in the literature which lend credibility to such a conjecture. Using a concept of entropy *for sequences* related to complexity, Kamae [1] has shown that (modulo the technical restriction that lim inf $(1/n)\sum_{k=1}^{a} y_k > 0$) sequences y having zero entropy define absolutely admissible place selections for Bernoulli sequences: if $B(p)$ is the set of Bernoulli sequences with parameter p (cf. Definition 0.4.2), y has zero entropy and $x \in B(p)$, then $x/y \in B(p)$; conversely, if $B(p) \subseteq (/y)^{-1} B(p)$, then y has zero entropy.

We can draw two lessons from this theorem. First, we see that the relativity which hovered threateningly above the Kollektivs ever since Kamke's "proof of impossibility" and Wald's theorems might be only apparent: the Bernoulli sequences admit a canonical set of absolutely admissible place selections (in so far as they depend on indices only). The only remaining relativity resides in the choice of Bernoulli sequences as models for Kollektivs.

Second, we may discern from the theorem that what makes a place selection y admissible is the existence of an "entropy barrier" between itself and the process which generates the Kollektivs (and hence, via Theorem 5.2, between itself and the Kollektivs). Indeed, it is this property, not recursiveness, which seems to matter.

When we learn that a sequence y having zero entropy in the sense of Kamae (the definition is too complicated to be given here) implies $\lim_{n\to\infty} I(y(n))/n = 0$ (Brudno [1, p. 140]), and recall the connection between I and statistical tests, we can model our conjecture on Kamae's theorem as follows:

"Suppose $\liminf (1/n)\sum_{k=1}^{n} y_k > 0$. Let $p \in (0,1)$ be computable. Then the following are equivalent:

(i) $\lim_{n\to\infty} I(y(n))/n = 0$.

(ii) For all x in $R(\mu_p)$, x/y is in $R(\mu_p)$."

I have at present no idea how to prove this, but let me note that at least the implication from (ii) to (i) is most likely in view of Kamae's theorem and the relation between Kamae-entropy and complexity.

§6. Appendix: notation and definitions.

6.1. Notations for sequences. 2^{ω} is the set of infinite binary sequences. If x is in 2^{ω}, then $x(n)$ is the initial segment of x of length n, and x_n is the nth coordinate of x.

The mapping $T\colon 2^{\omega} \to 2^{\omega}$ (called the left shift) is defined by $(Tx)_n = x_{n+1}$.

$2^{<\omega}$ is the set of all finite binary sequences. An element w of $2^{<\omega}$ is also called a word. The length of a word w is denoted $\mathrm{lh}(w)$. 2^n is the set of all w such that $\mathrm{lh}(w) = n$. If $m \le \mathrm{lh}(w)$, then $w(m)$ is the initial segment of w of length m, and w_m is the mth coordinate.

6.2. Topology on 2^{ω}. If B is a set, we let 1_B denote the characteristic function of B.

Let $2 = \{0,1\}$ have the discrete topology, and form the product topology on 2^{ω}. The open sets are the unions of the *cylinders* $[w] := \{x \mid x(\mathrm{lh}(w)) = w\}$, for a word w.

6.3. Measures on 2^{ω}. A measure on the Borel σ-algebra on 2^{ω} is completely determined by its values on the cylinders. Let $\{p_n\}$ in $[0,1]^{\omega}$ be a sequence of reals. This sequence determines a *product measure* on 2^{ω}, denoted $\prod_n(1-p_n, p_n)$, by putting

$$\prod_n (1-p_n, p_n)[w] = \prod_{k=1}^{\mathrm{lh}(w)} \hat{p}_k,$$

where

$$\hat{p}_k = \begin{cases} p_k, & w_k = 1. \\ 1 - p_k, & w_k = 0. \end{cases}$$

One product measure on 2^{ω} occurs so often that it is given a special name: $\lambda = \prod_n(\frac{1}{2}, \frac{1}{2})$. λ is the image of the Lebesgue measure on the unit interval under the natural map, and will also be called Lebesgue measure.

The following relationships between measures μ and ν are of special importance.

μ is *singular with respect to* ν (denoted $\mu \perp \nu$) if there exists a Borel set A such that $\mu A = 1$ and $\nu A = 0$.

μ is *absolutely continuous with respect to* ν (denoted $\mu \ll \nu$) if for all Borel A such that $\nu A = 0$, also $\mu A = 0$.

μ and ν are *equivalent* (denoted: $\mu \sim \nu$) if $\mu \ll \nu$ and $\nu \ll \mu$.

Let $\{\mu_n\}$ be a sequence of measures. We say that μ_n *converges weakly* to ν if for every Borel set A such that $\nu(\mathrm{Cl}(A) - \mathrm{Int}(A)) = 0$, we have

$$\lim_{n\to\infty} \mu_n A = \nu A.$$

The portmanteau theorem states that weak convergence is equivalent to convergence on the cylinders. We say that μ_n *converges strongly* to ν if for all Borel sets A, $\lim_{n\to\infty} \mu_n A = \nu A$.

6.4. Computability. A function $f\colon \omega \to \mathbf{R}$ is called *computable* if there exists a recursive function $g\colon\omega \times \omega \to \mathbf{Q}$ such that, for all n and k, $|f(n) - g(n,k)| < 2^{-k}$. Using codes for the cylinders, it makes sense to take of *computable measures* on 2^ω.

Note that, if μ is a computable measure, the following sets are Σ_1:

$$W_> = \{\langle w, a\rangle \mid w \in 2^{<\omega}, a \in \mathbf{Q}^+, \mu(w) > a\},$$
$$W_< = \{\langle w, a\rangle \mid w \in 2^{<\omega}, a \in \mathbf{Q}^+, \mu(w) < a\}.$$

A slightly stronger concept results if we demand that these sets be Δ_1. That is, a measure μ on 2^ω is *strongly computable* if the sets $W_>$ and $W_<$ are Δ_1.

6.5. Ergodic theory. A measure μ on 2^ω is called *stationary* if for all Borel sets A, $\mu T^{-1} A = \mu A$.

μ is *ergodic* if $T^{-1}A = A$ implies $\mu A = 0$ or 1.

Associated to a stationary measure μ is its (metric) entropy $H(\mu)$, defined by

$$H(\mu) = \lim_{k\to\infty} -\frac{1}{k} \sum_{w \in 2^k} \mu[w] \cdot \log \mu[w].$$

REFERENCES

V. M. Alekseev and M. V. Yakobson [1], *Symbolic dynamics and hyperbolic dynamical systems*, ***Physics Reports***, vol. 75 (1981), pp. 287–325.

E. Bishop and H. Cheng [1], *Constructive measure theory*, ***Memoirs of the American Mathematical Society***, no. 116 (1972).

A. Brudno [1], *Entropy and the complexity of the trajectories of a dynamical system*, ***Transactions of the Moscow Mathematical Society***, 1983, no. 2 (44), pp. 127–151.

G. J. Chaitin [1], *A theory of program size formally identical to information theory*, ***Journal of the Association for Computing Machinery***, vol. 22 (1975), pp. 329–340.

——— [2], *Algorithmic information theory*, ***IBM Journal of Research and Development***, vol. 21 (1977), pp. 350–3509.

A. Church [1], *On the concept of a random sequence*, ***Bulletin of the American Mathematical Society***, vol. 46 (1940), pp. 130–135.

J.-E. Dies [1], *Information et complexité*, ***Annales de l'Institut Henri Poincaré, Section B: Calcul des Probabilitiés et Statistique, Nouvelle Série***, vol. 12 (1976), pp. 365–390; vol. 14 (1978), pp. 113–118.

W. Feller [1], ***Introduction to probability theory and its applications.*** Vol. I, Wiley, New York, 1968.

——— [2], Idem, Vol. II, Wiley, New York, 1971.

J. FORD [1], *How random is a coin toss?* ***Physics Today***, vol. 36 (1983), no. 4 (April), pp. 40–47.

P. GACS [1], *On the relation between descriptional complexity and algorithmic probability*, ***Theoretical Computer Science***, vol. 22 (1983), pp. 71–93.

H. GAIFMAN and M. SNIR [1], *Probabilities over rich languages, randomness and testing*, this JOURNAL, vol. 47 (1982), pp. 495–548.

S. KAKUTANI [1], *On the equivalence of infinite product measures*, ***Annals of Mathematics***, ser. 2, vol. 49 (1948), pp. 214–224.

T. KAMAE [1], *Subsequences of normal sequences*, ***Israel Journal of Mathematics***, vol. 16 (1973), pp. 121–149.

E. KAMKE [1], *Über neuere Begründungen der Wahrscheinlichkeitsrechnung*, ***Jahresbericht der Deutschen Mathematiker-Vereinigung***, vol. 42 (1932), pp. 14–27.

A. KECHRIS [1], *Measure and category in effective descriptive set theory*, ***Annals of Mathematical Logic***, vol. 5 (1973), pp. 337–384.

A. N. KOLMOGOROV [1], *On tables of random numbers*, ***Sankhyā, Series A***, vol. 25 (1963), pp. 369–376.

——— [2], *Combinatorial foundations of information theory and the calculus of probabilities*, ***Russian Mathematical Surveys***, vol. 38 (1983), no. 4, pp. 29–40.

L. A. LEVIN [1], *On the notion of a random sequence*, ***Soviet Mathematics Doklady***, vol. 14 (1973), pp. 1414–1416.

——— [2], *Various measures of complexity for finite objects*, ***Soviet Mathematics Doklady***, vol. 17 (1976), pp. 522–526.

A. J. LICHTENBERG and M. A. LIEBERMAN [1], ***Regular and stochastic motion***, Springer-Verlag, Berlin, 1983.

P. MARTIN-LÖF [1], *The definition of random sequences*, ***Information and Control***, vol. 9 (1966), pp. 606–619.

——— [2], *Complexity oscillations in infinite binary sequences*, ***Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete***, vol. 19 (1971), pp. 225–230.

R. VON MISES [1], *Grundlagen der Wahrscheinlichkeitsrechnung*, ***Mathematische Zeitschrift***, vol. 5 (1919), pp. 52–99.

——— [2], ***Wahrscheinlichkeit, Statistik und Wahrheit***, Springer-Verlag, Vienna, 1928. (The English edition, Dover, New York, 1981, is cited as [2a].)

——— [3] (with H. GEIRINGER), ***Mathematical theory of probability and statistics***, Academic Press, New York, 1964.

J. C. OXTOBY [1], ***Measure and category***, 2nd ed., Springer-Verlag, Berlin, 1980.

K. PETERSEN [1], ***Ergodic theory***, Cambridge University Press, Cambridge, 1983.

G. SACKS [1], *Measure theoretic uniformity in recursion theory and set theory*, ***Transactions of the American Mathematical Society***, vol. 142 (1969), pp. 381–424.

C. P. SCHNORR [1], ***Zufälligkeit und Wahrscheinlichkeit***, Lecture Notes in Mathematics, vol. 218, Springer-Verlag, Berlin, 1971.

——— [2], *Process complexity and effective random tests*, ***Journal of Computer and System Sciences***, vol. 7 (1973), pp. 376–388.

E. TORNIER [1], *Comment*, ***Mathematische Annalen***, vol. 108 (1933), p. 320.

J. VILLE [1], *Étude critique de la notion de collectif*, Gauthier-Villars, Paris, 1939.

A. WALD [1], *Die Widerspruchsfreiheit des Kollektivbegriffs in der Wahrscheinlichkeitsrechnung*, ***Ergebnisse eines Mathematischen Kolloquiums***, vol. 8 (1938), pp. 38–72.

Note. The proceedings of the 1936 Geneva conference on probability theory were published in the series Actualités Scientifiques et Industrielles, Hermann, Paris, 1938 (see especially nos. 734 and 735).

Note added in proof. One can show that for the measure μ defined in Lemma 4.2.3, also $\mu(\mathrm{LIL}(\lambda)^c \cap C(1/2)) = 1$.

DEPARTMENT OF PHILOSOPHY
TECHNICAL UNIVERSITY OF DELFT
DELFT, THE NETHERLANDS

Current address: Department of Mathematics, University of Amsterdam, Amsterdam, The Netherlands.