3 A New Start: Martin-Löf's Definition

3.1 Introduction At the close of the Geneva conference on probability theory (see 2.6) it became clear that von Mises' axiomatisation of the probability calculus had lost the day. Although *sub specie aeternitatis* almost none of the objections brought against von Mises was cogent, Kolmogorov's measure theoretic formalism, which did not attempt to define probability explicitly, was henceforth universally accepted.

With the acceptance of a measure theoretic foundation of probability theory, the necessity of providing a rigorous definition of randomness disappeared. Consequently, from the publication of Ville' s book [99] in 1939 to 1963, interest in the problem dwindled. In 1963, however, Kolmogorov came to the conclusion that the frequency interpretation stood in need of a precise formulation after all. He published a definition of randomness for finite sequences [47] which contains the germ of *Kolmogorov–complexity* (defined in Chapter 5). Martin-Löf, investigating sequences with high Kolmogorov–complexity, gave a definition of randomness [62] involving a particular type of statistical test, namely, *significance tests*. This definition is nowadays the one most generally accepted. In this chapter we introduce Martin-Löf's definition and several variants and discuss their respective merits.

As a consequence of the criticism voiced by Fréchet and Ville, the problem of defining randomness was now conceived as follows: a random sequence (with respect to some probability measure) should satisfy all probabilistic laws for that measure; in other words, the set of random sequences should be the intersection of all properties of probability one. Of course, in this form, the demand is impossible to satisfy, since the required intersection is empty. Hence we have to choose among the properties of probability one; and Martin-Löf's definition is one such choice.

The main result of the previous chapter is that this way of introducing Kollektivs has not much more than the name in common with von Mises' ideas. For one thing, it completely reverses the attitude von Mises expressed in the slogan "Erst das Kollektiv, dann die Wahrscheinlichkeit". What's more, for von Mises a Kollektiv x in 2^ω induces a probability distribution on {0,1}, not on 2^ω itself; so from his point of view, there is no immediate relation between properties of probability one in 2^ω and Kollektivs x in 2^ω.

Speaking mathematically, a distribution (1–p,p) on {0,1} determines a *measure* $\mu_p = (1-p,p)^\omega$ on 2^ω, but this measure is a *probability* only if it is induced by a Kollektiv $\xi \in (2^\omega)^\omega$. To be sure, such a measure can be extremely helpful in proving existence theorems; for instance, in this way we proved that the set of Church random sequences C(p) has μ_p– measure one

(theorem 2.5.2.3). But this result should not be construed as implying that a "true" random sequence should at least be Church random (*because* Church randomness is a property of probability one).
Another consequence of strict frequentism is that the distribution (1–p,p) on {0,1} in no way determines a unique distribution on 2^ω, to wit, μ_p. Indeed, the distribution on (1–p,p) would lead uniquely to μ_p if it were a property of each coordinate, as in the propensity interpretation. But, according to strict frequentism, a Kollektiv x in 2^ω allows no such conclusion: p is really only a limiting relative frequency. It follows that all measures which, in a sense to be made precise in Chapter 4, determine the same limiting relative frequency p, should be treated on equal footing, and existence theorems should not be sensitive to which measure (from the class of measures which determine the same relative frequencies) we choose. Some notation we introduced in Chapter 2 was intended to reflect this point: e.g. the set of Church random sequences with parameter p was denoted C(p), to emphasize the fact that only the limiting relative frequency p is relevant. In Chapter 4 we shall show that, roughly speaking, C(p) has measure one for measures which determine the same p.
The randomness notions which we shall introduce in this chapter are, on the other hand, very sensitive to the underlying measure. This is emphasized by the notation R(μ), meaning "the set of sequences random with respect to the measure μ". Exactly *how* sensitive to the choice of a measure these notions are, will be investigated in Chapter 4.

Although we may have so far given the impression that the definition of randomness of Martin-Löf and its variants, being conceived in sin, are ipso facto unsatisfactory, this is not our purpose. The preceding chapter should have convinced the reader that randomness defined as the satisfaction of "all" properties of probability one is anathema to the strict frequentist. It is not, however, implied that such a definition does not make sense on any view of probability. In particular, if you subscribe to some variant of the propensity interpretation, which views probability primarily as a physical property of an experimental set-up, it does make sense to have randomness defined with respect to some unique probability distribution on 2^ω.
Indeed, the widespread belief that Kollektivs should satisfy the law of the iterated logarithm, and that probability zero of an outcome should exclude that this outcome occurs infinitely often (at least for a discrete sample space), probably testifies to an instinctive acceptance of the propensity interpretation. Accordingly, the mathematical differences between the two definitions, investigated in detail in Chapter 4, may be seen as a contribution towards the study of the philosophical differences between these two interpretations of probability.

This chapter is organized as follows. In 3.2 we introduce the definitions of randomness of Martin-Löf [62] and Schnorr [88] and we prove some recursion theoretic properties of these definitions (3.2.2-3).

Although most of the results occur already in Schnorr's book, the proofs have been simplified, e.g. by using the so-called Basis Theorem from recursion theory. Apart from added elegance, we thus introduce a technique that will be helpful in Chapter 5.

Having thus prepared the ground, we turn to some problems not usually treated in the literature. For one thing, there is a notable lack of concrete examples of properties which random sequences satisfy. E.g. in Schnorr's book, only the validity of the law of large numbers is verified, not even that of the law of the iterated logarithm. This fact is slightly ironical, since the non-validity of the law of the iterated logarithm for von Mises' Kollektivs was the main impetus behind the new approach.

One of the goals of this thesis is, therefore, to exhibit more examples of properties of random sequences. For a start we prove in 3.3 effective versions of the Borel-Cantelli lemmas, which allow one to show that random sequences satisfy the usual probabilistic laws.

So far, random sequences are considered only from the point of view of probability theory. Martin-Löf's original introduction of random sequences proceeded slightly differently: a sequence was defined to be random with respect to some statistical hypothesis H if it is not rejected by some (effective) statistical test for H at arbitrarily small levels of significance. From this perspective, it is not immediately clear that Martin-Löf's definition is the correct one to use, since there is some controversy surrounding the notion of significance test employed in the definition.

To set the stage for the discussion, we introduce *Martingales* in 3.4. Martingales were first mentioned in 2.6.2, in connection with Ville's construction, as formalisations of gambling strategies. We shall briefly examine this aspect of Martingales, but our main interest lies in their statistical meaning, as *likelihood ratios.* In 3.5 we explain the controversy surrounding significance tests and we discuss some alternatives to Martin-Löf's definition. A conclusion follows in 3.6.

The relation between Martin-Löf's definition and that of von Mises is discussed in Chapter 4, which is considerably more technical than Chapter 3.

3.2 The definitions of Martin-Löf and Schnorr

3.2.1 Randomness via probabilistic laws Ville ended his book [99] on a note of resignation: a random sequence should satisfy all properties of probability one; that's impossible, so which probabilistic laws should we choose? Ville had shown that, in a sense, any probabilistic law can be represented by a Martingale (see lemma 3.4.7 below), so the question could equivalently be posed as: which gambling strategies should one choose? Any choice seemed to be arbitrary, thus causing the definition of random sequences to be arbitrary as well. Of course Ville didn't mind, not being a strict frequentist.

In [62], Martin-Löf proposed a canonical choice for the class of probabilistic laws: the class of those laws which can be proved effectively. To explain this notion of effectiveness, we must look at proofs of probabilistic laws.

A probabilistic law, according to the usual interpretation, is a statement of the form:

$$\mu\left\{ x \in 2^{\omega} |\, A(x) \right\} = 1,$$

where A is some formula. The discussion in 2.4.3 should have made clear that this is not von Mises' concept of a probabilistic law; but we are in a different circle of ideas now.

Typically, a proof of such a statement proceeds in either of the two following ways (examples will be given in 2.3):

(i) One constructs a sequence (O_n) of open sets such that (a) $\{x|A(x)\}^c \subseteq O_n$ for all n, (b) $\mu O_n \leq 2^{-n}$ (or any other recursive function of n which decreases to 0), (c) the O_n are recursively enumerable unions of cylinders, or at least unions recursively enumerable in μ and (d) similarly, the function which associates to each n a Gödelnumber for O_n is recursive in μ.

(ii) One uses the two Borel-Cantelli lemmas (Feller [25,200-2]):

(a) if (A_n) is a sequence of sets such that $\sum_n \mu A_n < \infty$, then

$$\mu \bigcap_n \bigcup_{m \geq n} A_m = 0$$

(b) if (A_n) is a sequence of independent events such that $\sum_n \mu A_n = \infty$, then

$$\mu \bigcap_n \bigcup_{m \geq n} A_m = 1.$$

Usually such a sequence (A_n) satisfies properties analogous to (c) and (d) in (i).

Roughly speaking, a probabilistic law is effective if it can be proved according to (i) or (ii). Not all probabilistic laws are effective in this sense; the ergodic theorem (see 7.4) may be a case in point[1].

Martin-Löf's definition of randomness may be seen as a formalisation of procedure (i). Procedure (ii) will receive separate treatment in 3.3.

Let us first introduce two notions of a measure being computable.

3.2.1.1 Definition The probability measure μ on 2^{ω} is called *computable* if there exists a recursive function g: $2^{<\omega} \times \omega \to \mathbb{Q}$ such that for all w,k: $|\mu[w] - g(w,k)| < 2^{-k}$.

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

$W_> := \{\langle w,a \rangle \in 2^{<\omega} \times \mathbb{Q}^+ |\, \mu[w] > a\}$ and $W_< := \{\langle w,a \rangle \in 2^{<\omega} \times \mathbb{Q}^+ |\, \mu[w] < a\}$.

A slightly stronger concept of computability for measures results if we demand that these sets be Δ_1: a measure μ is *strongly computable* if the associated sets $W_<$, $W_>$ are Δ_1.

Evidently a strongly computable measure is computable, but not conversely: strong computability excludes measures μ such that it cannot be decided whether $\mu[w]$ is rational, a case not very likely to occur in practice. In section 3.4 we have to introduce still another notion of computability for measures, this time weaker than those above.

For computable measures, the clauses "recursive in" in (c) and (d) of (i) can be replaced by "recursive" pure and simple. We shall now formally introduce procedure (i) under the name of "recursive sequential test". This name, coined by Martin-Löf, reflects the statistical origin of these sets, statistical rather than probabilistic. The statistical view will be explained in 3.5.

3.2.1.2 Definition Let μ be a computable measure. $N \subseteq 2^\omega$ is a *recursive sequential test* with respect to μ if N can be written as a Π_2 set $\bigcap_n O_n$, where $O_n \in \Sigma_1$, the function $n \to O_n$ is recursive, $O_{n+1} \subseteq O_n$ and $\mu O_n \leq 2^{-n}$.

We shall see below that probabilistic laws such as the law of the iterated logarithm or the law of large numbers can indeed be proven by constructing recursive sequential tests covering the sets of sequences not satisfying these laws. In fact, these proofs usually show something more: with the notation as in the preceding definition, one usually has that the μO_n are computable *uniformly in* n, i.e. that for some recursive function $f: \omega \times \omega \to \mathbb{Q}$,

$$\forall n,k \; |\mu O_n - f(n,k)| < 2^{-k}$$

This added feature is present in Schnorr's definition of *total recursive sequential test* [88,63].

3.2.1.3 Definition With the notation of 3.2.1.2: N is a *total recursive sequential test* with respect to μ if μO_n is computable uniformly in n.

Schnorr's reasons for preferring this definition will be examined in 3.2.3 and 3.4. In 3.2.3 we shall see that indeed some recursive sequential tests are not total.
Abstractly, we may now introduce definitions of randomness as follows:

3.2.1.4 Definition Let μ be a computable measure. $x \in 2^\omega$ is *random* with respect to μ (denoted $x \in R(\mu)$) if for all recursive sequential tests N with respect to μ, $x \notin N$.

3.2.1.5 Definition Let μ be a computable measure. $x \in 2^\omega$ is *weakly random* with respect to μ (denoted $x \in R_w(\mu)$) if for all total recursive sequential tests N with respect to μ, $x \notin N$. (Schnorr calls *hyperzufällig* what we call random, and *zufällig* what we call weakly random.)
3.2.1.6 Lemma $R(\mu) \subseteq R_w(\mu)$ and $\mu R(\mu) = \mu R_w(\mu) = 1$.

Proof Each (total) recursive sequential test has measure zero and there are only countably many of them. □

These definitions are very abstract, much more so than that of von Mises. For example, while a probabilistic law gives rise to a (total) recursive sequential test, via procedure (i) on p. 58, the converse does not seem to be obvious: does every recursive sequential test correspond to a bona fide probabilistic law? In order to answer such questions, one must have some kind of representation or classification of recursive sequential tests. Sections 3-5 of this chapter, and also Chapter 4, contain some efforts in this direction. The rest of 3.2 develops some recursion theoretic properties of the above definitions and settles a question left open by lemma 3.2.1.6, namely: is every weakly random sequence also random?

3.2.2 Recursive sequential tests A surprising property of recursive sequential tests is:

3.2.2.1 Lemma (Martin-Löf [62]) Let μ be a computable measure. (a) The collection of recursive sequential tests with respect μ to is recursively enumerable. (b) There exists a universal recursive sequential test with respect to μ, i.e. a test U such that for all recursive sequential tests N with respect to μ, $N \subseteq U$.

A curious consequence of the preceding lemma is that $R(\mu)$ and, a fortiori $R_w(\mu)$, have elements which are rather simple. Although neither set contains recursive sequences if μ is non-atomic (for if x is recursive, $\bigcap_n[x(n)]$ is a total recursive sequential test with respect to any non-atomic computable μ; cf. remark 3.2.3.11), $R(\mu)$ does contain Δ_2-definable sequences. This is a consequence of the following

3.2.2.2 Basis Theorem (Soare [92,109]) Any non-empty Π_1 subset of 2^ω has a Δ_2-definable element.

Proofsketch A Π_1 subset of 2^ω can be viewed as the set of infinite paths through a recursive binary tree T. Call $w \in T$ *admissible* if $\forall n>|w|\, \exists v \in 2^n (w \subseteq v \;\&\; v \in T)$. (By König's Lemma, w is admissible iff there is an infinite branch of T through w.) The set of admissible words is Π_1. Since the subset is non-empty, T has an infinite branch. The *leftmost* infinite branch can be constructed recursively in the set of admissible words, which is Π_1; hence this branch must itself be Δ_2. □

3.2.2.3 Lemma Let μ be a non-atomic computable measure. Then $R(\mu)$ contains Δ_2-, but no Δ_1-, definable sequences.

Proof (See also Schnorr [88, 56].) By 3.2.2.1, $R(\mu)$ is a Σ_2 set of measure 1. Pick a Π_1 set A $\subseteq R(\mu)$ such $\mu A > 0$ and apply the Basis Theorem. If x is recursive and μ computable and non–atomic, then $\bigcap_n[x(n)]$ is a total recursive sequential test with respect to any non-atomic computable μ; cf. remark 3.2.3.11. □

Although Δ_2 sequences may thus possess all statistical properties associated with randomness, in another sense they can be completely deterministic.

$$\lim_{k\to\infty} (\xi_k)_n.$$

In words: Δ_2 sequences x can be produced by Turing machines if the machine is allowed to correct itself a finite number of times per x_n. This is a far cry from the usual mechanisms that produce random sequences: indeterministic systems such as those of quantum mechanics, or deterministic systems that have been subject to coarse graining (see Chapter 5). The finer tools of Kolmogorov complexity will allow us to distinguish between Δ_2 definable random sequences and those which are not so simply definable.

3.2.3 Total recursive sequential tests The requirement of uniform computability of the μO_n is strong; to prove that a recursive sequential test is in fact total sometimes demands considerable effort. Fortunately, nullsets bearing a strong resemblance to total recursive sequential tests were already known in constructive mathematics, so we can draw upon the large reservoir of proof techniques developed there (see, e.g., the books by Bishop [5], Bridges [9] and Bishop-Bridges [6]) Although not every total recursive function is acceptable in constructive mathematics (since the proof that the function is in fact total must itself be constructively valid), arguments involving constructive functions usually carry over directly to recursive functions; when the result is simple we shall not bother to write down proofs. For instance, we shall often have occasion to use the following comparison principle:

3.2.3.1 Lemma (See [5,30].) Let (a_n), (b_n) be recursive sequences of computable reals such that $0 \leq a_n \leq b_n$ and $\sum_n b_n < \infty$ is computable. Then $\sum_n a_n$ is also computable.

To compute the measure of a Σ_1 set, it is often helpful to have such sets presented in normal form, namely as a disjoint union of sets of the form [w]. For if A in Σ_1 is brought in such a form, i.e. $A = \bigcup_i [w^i]$, then $\mu A = \sum_i \mu[w^i]$.

3.2.3.2 Definition A subset S of $2^{<\omega}$ is called *prefixfree* if for distinct $w,v \in S$: neither $w \subseteq v$ nor $v \subseteq w$.

If S is prefixfree, the open set determined by S, namely $[S] = \{x \mid \exists n(x(n)) \in S)\}$ can be written as

$$[S] = \dot{\bigcup}_{w \in S} [w] \quad \text{(where } \dot{\bigcup} \text{ denotes disjoint union).}$$

3.2.3.3 Lemma For every Σ_1 set $A \subseteq 2^\omega$, one can effectively determine a recursively enumerable prefixfree set $S \subseteq 2^{<\omega}$ such that $A = [S]$.

Proof A is of the form [T], $T \subseteq 2^{<\omega}$ r.e. Generate T. S is obtained as a union $\bigcup_n S_n$, $S_n \subseteq S_{n+1}$. Suppose S_n has been constructed. Consider the n+1[th] word w in T. (a) If w is a prolongation of some v in S_n, put $S_{n+1} = S_n$. (b) If w is an initial segment of some v in S_n, replace w by all its prolongations of length $|v|$ and apply (a) and (b) to each of these prolongations. This process comes to a halt; let S_n be the union of S_n and the finite list thus obtained and proceed. (c) In all other cases, put $S_{n+1} = S_n \cup \{w\}$.[2] □

Using this lemma one can easily show

3.2.3.4 Lemma Let μ be a computable measure on 2^ω; A, B Σ_1 subsets of 2^ω with μA, μB computable. Then $\mu(A \cup B)$, $\mu(A \cap B)$ are computable.

Proof We do the first case only. We may suppose that A is written as a *disjoint* union $\bigcup_n [w^n]$; let $B = [v]$. Then $\mu(A \cup B) = \sum_n \mu([w^n] \cup [v])$ and we may apply lemma 3.2.3.1 with $a_n = \mu([w^n] \cup [v])$ and $b_n = \mu[w^n] + \mu[v]$. For the general case, write B as a disjoint union $\bigcup_m [v^m]$; then $\mu(A \cup B) = \sum_m \mu(A \cup [v^m])$. Apply 3.2.3.1 with $a_m = \mu(A \cup [v^m])$ (which is compuable by the first part of the proof) and $b_m = \mu A + \mu[v^m]$. □

We now come to an essential feature of Σ_1 sets O such that μO is computable. If O is just Σ_1, it may be the case that all recursive sequences are contained in O; this is for instance true of the levels U_n of a universal recursive sequential test U. Not so for Σ_1 sets O with μO computable:

3.2.3.5 Lemma Let μ be a computable measure, O in Σ_1 and μO computable. Then for any word w such that $\mu([w] \cap O) < \mu[w]$, there exists a *recursive* x in $[w] \cap O^c$.

Proof This is just a formalisation of an old intuitionistic result; see e.g. Schnorr [88,64-5]. Alternatively, one could show that, if μO is computable, it can be written as a recursive union of cylinders [w] and then apply the lemma proved in footnote 2. □

3.2.3.6 Corollary Let μ be a computable measure which is positive on open sets, A a $\prod_1$ set without recursive elements. Then either μA = 0 or μA is not computable (both cases occur).

For our purpose the most important consequence is

3.2.3.7 Corollary (a) Let μ be a computable measure. If N is a total recursive sequential test with respect to μ, there exists a recursive x ∉ N. (b) If μ is non-atomic, there exists no universal total recursive sequential test with respect to μ.

Proof (a) Write $N = \bigcap_n O_n$ as in definition 3.2.1.3. Observe that $\mu O_1 < 1$ and apply lemma 3.2.3.5. (b) Otherwise, by (a), there would exist a recursive sequence outside this universal test. □

Schnorr sees in the preceding lemma a mark of the superiority of total recursive sequential tests over recursive sequential tests. The construction of a recursive x outside N implies that we can construct a model of the probabilistic law corresponding to N, so that we can visualize the property stated by the law (von Mises considered this use of recursive "Kollektivs" in [69]). This is indeed not an unreasonable requirement for probabilistic laws which purport to be effective. But the requirement is satisfied by other types of tests as well (see footnote 2 and section 3.4). Furthermore, the existence of recursive sequences satisfying a probabilistic law does not imply visualizability of that law in any real, practical, sense: there must exist recursive absolutely normal numbers (i.e. numbers which are normal to every base), but there are no examples of absolutely normal numbers which are as easily described as the example of a normal number in lemma 2.5.1.5. It therefore seems more correct to say that, whenever a probabilistic law can be associated with a *total* recursive sequential test, the possibility of a visualizable model for that law is at least not excluded.
We now state a technical lemma which, besides being useful later, will imply that the collection of total recursive sequential tests (with respect to a given measure) is not r.e.

3.2.3.8 Lemma (Schnorr [88,65]) Let μ be a computable measure and $(N_k)_k$ a recursively enumerable collection of total recursive sequential tests with respect to μ. Then $\bigcup_k N_k$ is a contained in a total recursive sequential test M with respect to μ.

Proof Let $N_k = \bigcap_n O_{k,n}$. Put $M = \bigcap_n \bigcup_k O_{k,(n+k)}$. M is a recursive sequential test with respect to μ.

To compute $\mu \bigcup_k O_{k,(n+k)}$, note that for $n+1 < i < j$:

$$\mu\bigcup_{k=1}^{j} O_{k,(n+k)} - \mu\bigcup_{k=1}^{i} O_{k,(n+k)} \leq \sum_{k=i}^{j}\mu O_{k,(n+k)} \leq \sum_{k=i}^{j} 2^{-k-n}$$

hence lemma 3.2.3.4 implies that

$$\left(\mu\bigcup_{k=1}^{j} O_{k,(n+k)}\right)_{j\in\mathbb{N}}$$

is a recursive sequence of computable reals which is recursively Cauchy, so converges to a computable real (see [5,27]). ∎

3.2.3.9 Corollary Let μ be a non-atomic computable measure. The collection of total recursive sequential tests with respect to μ is not r.e.

Proof Otherwise the M constructed in lemma 3.2.3.8 would be universal. ∎

We now come to the main result of this section: that $R(\lambda) \subset R_w(\lambda)$. This observation is due to Schnorr [88,77], whose proof uses Martingales and a detour via a different randomness concept.

3.2.3.10 Theorem Let μ be a computable measure. Then there exists a sequence which is weakly random, but not random, with respect to μ.

Proof Let $(N_k)_{k\in\mathbb{N}}$ be an enumeration of the collection of total recursive sequential tests with respect to μ. By lemma 3.2.3.8, we may assume that each N_k is of the form $\bigcap_n O_{k,n}$, where

$$O_{k,n} = \bigcup_{i=1}^{k-1} O_{i,(n+i)}.$$

We construct a weakly random, but non–random x as a pointwise limit of a sequence $(\xi_k)_{k\in\mathbb{N}}$, where $\xi_k \in 2^\omega$. Let $U = \bigcap_n U_n$ be the universal recursive sequential test with respect to μ.
By lemma 3.2.3.5, we can construct a recursive ξ_1 not contained in $O_{1,1}$. Since μ is non-atomic, U contains all recursive sequences. Determine k_1 such that $[\xi_1(k_1)] \subseteq U$. Since $[\xi_1(k_1)]\cap(O_{1,1})^c \neq \emptyset$, there exists a recursive ξ_2 such that $\xi_2(k_1) = \xi_1(k_1)$ and ξ_2 not contained in $O_{2,1}$. Determine $k_2 > k_1$ such that $[\xi_2(k_2)] \subseteq U$. Proceeding inductively we construct recursive ξ_k not contained in $O_{k,1}$. Put $x_n = \lim_{k\to\infty} (\xi_k)_n$. We show that for all k, $x \notin O_{k,k+1}$. For if $x \in O_{k,k+1}$, say $[x(m)] \subseteq O_{k,k+1}$, we can determine k'>k such that $\xi_{k'}(m) = x(m)$. Since $x_{k'}$ is not contained in $O_{k',1}$ and

$$O_{k',1} = \bigcup_{i=1}^{k'-1} O_{i,i+1} ,$$

$\xi_{k'}$ is not contained in $O_{k,k+1}$, a contradiction. □

3.2.3.11 Remark If M is a total recursive sequential test with respect to μ, $M = \bigcap_n O_n$, then the conventional upper bound on μO_n is 2^{-n}. This requirement may be relaxed. For if $M = \bigcap_n O_n$ is a Π_2 μ-nullset and each μO_n is computable, then M is contained in a total recursive sequential test N: since for each k, $\mu \bigcap_{n \leq k} O_n$ is computable (uniformly in k) by lemma 3.2.3.2, there exists a total recursive g: $\omega \to \omega$ such that for all m,

$$\mu \bigcap_{n \leq g(m)} O_n \leq 2^{-m};$$

if we then put

$$O'_m = \bigcap_{n \leq g(m)} O_n ,$$

$N := \bigcap_m O'_m$ is the required recursive sequential test.

3.2.4 An appraisal and some generalisations Do the definitions of Martin-Löf and Schnorr really amount to a canonical choice of a class of probabilistic laws, thus providing an *absolute* concept of randomness? Martin-Löf must have had his doubts, since he later proposed to define the set of random sequences as the intersection of all hyperarithmetical sets of measure one [64], the reason being that "the specific Borel sets considered [in probability theory] are always obtained by applying the Borelian operations to recursive sequences of previously defined sets, which means precisely that they are hyperarithmetical" [64,74]. Nor is it clear that this is really the end: why not consider all Borel sets of measure one with codes in some admissible set, the theory of admissible sets being the natural generalisation of recursion theory?
Even if we assume that a random sequence should satisfy all "effective" laws of probability theory, still "effectiveness" is an open-ended notion, so we can't expect to arrive at some definitive notion of randomness in this way. The question is, whether we would be much happier with such a definition.

We believe that the alleged "problem of the relativity of randomness" is a pseudo-problem, born from an excessive concern with abstract things. The fundamental concept of mathematics, set, is relative (with respect to axioms and models for set theory), but that doesn't imply that the notion is useless; only that we should stick to those properties which are uncontroversial, whenever possible. Very few mathematicians are willing to forego sets, just because the contours of the universe of sets are hazy. Some, notably Kreisel, even believe that philosophical analysis of the notion of set may help to enlarge the charted domain.

The situation with respect to random sequences is different in so far as it is quite possible to do mathematics without them; and one is of course much less willing to bear with a problematic concept if one can forego it. We have seen in the previous chapter, however, that random sequences are necessary for a frequentist foundation of probability and in particular that random sequences should minimally be invariant under admissible place selections. Invariance under place selections also suffices to explain the applicability of probability theory, so that Martin-Löf's definition is threatened by relativity only because it disregards the function of random sequences in von Mises' probability theory. The propensity interpretation does nothing to remove this relativity.

We therefore propose to investigate the modern definitions of randomness, not with a view to single out one as *the* definition, but rather to establish reasonable (or just interesting) properties of random sequences. This attitude entails that we do not introduce sets which are more complex (in the sense of the arithmetical hierarchy) than those occurring in definitions 3.2.1.2-3, unless we are forced to do so (see below). We wish to remain agnostic about the exact boundary of the set of properties a random sequence has to satisfy (when these sequences are not considered in their role as foundation for probability theory). The fact that we shall almost never consider sets which are more complex than those in definitions 3.2.1.2-3 does *not* imply that we believe that all (total) recursive sequential tests are reasonable probabilistic laws, since it depends on one's views on, e.g., statistics (does significance testing make sense in the absence of an alternative hypothesis? what exactly *is* an alternative hypothesis?) which properties of random sequences to accept. All in all, then, we regard the definitions of Martin-Löf and Schnorr as convenient way-stations, as technically elegant, concise descriptions of probabilistic laws. But we think that, in their present form, these definitions are too abstract and that questions such as "Is Martin-Löf's definition the right one to use?" do not make sense. Moreover, worrying about the recursive aspects of the definition might easily lead to a neglect of its more urbane questionable aspects.

We shall now examine possible reasons for enlarging the framework. Up till now, we have considered only computable measures. What happens if, for some reason or other, we wish to consider measures which are not computable? A moment's reflection on how a measure μ occurs in a probabilistic law (or a glance at section 3.3) will show that the most useful concept in this context is 3.2.1.2. with "Π_2" replaced by "Π_2 in μ". Most theorems hold for the new concept if we put in "recursive in μ" in the appropriate places; section 3.3 will provide illustrations of this point. Consequently, allowing non-computable measures does not really amount to a generalization.

We *do* get a generalization if we drop the requirement in 3.2.1.2 that μO_n be bounded by 2^{-n};

that leaves us with just a bare Π_2 μ-nullset. Once we're on this slippery slope, we could replace the Π_2 set by a Π_n set, for arbitrary n. This is indeed what happens in Gaifman and Snir [34]. They introduce

3.2.4.1 Definition Let μ be a computable measure. x is n–*random* with respect to μ (Notation: $x \in R_n(\mu)$) if for all Π_n μ-nullsets N, $x \notin N$.

It will turn out (in Chapter 4) that the concept is actually most useful for strongly computable measures, which were defined in 3.2.1.1. Again, if we wish to consider arbitrary measures μ, it is best to replace "Π_n" by "Π_n in μ".

It is doubtful whether we really do need this generality. I know of one probabilistic law which may not be effective in the sense introduced in 3.2.1: the ergodic theorem (which is stated in the appendix, 6.4). In this case, e.g. the set

$$\left\{x \in 2^\omega \mid \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k > \mu[1] > \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k\right\}$$

is Σ_3 in μ, i.e. a countable union of Π_2 μ-nullsets; so here at least is some use for 2–randomness.

Let us therefore in conclusion of this part compare 2–randomness (definition 3.2.4.1) with randomness (definition 3.2.1.4).

3.2.4.2 Lemma Let μ be a non-atomic computable measure. (a) There is no universal Π_2 μ-nullset. (b) There exist sequences which are random, but not 2–random, with respect to μ.

Proof (a) Suppose U were a universal Π_2 μ-nullset. Then $\mu U^c = 1$ and U^c is Σ_2. It then follows from the Basis Theorem (3.2.2.2) that U contains a Δ_2 definable sequence x. But then $\{x\}$ is a Π_2 set and $\mu\{x\} = 0$ by non-atomicity of μ. (b) If not, then $R(\mu)^c$ would be a universal Π_2 μ-nullset. □

In fact, as an application of the techniques developed in Chapter 4 we shall show in 4.7 that for some continuous measure μ: $\mu(R(\lambda) \cap R_2(\lambda)^c) = 1$.

3.3 Probabilistic laws After these abstract considerations, let us now exhibit some concrete examples of probabilistic laws which are satisfied by (weakly) random sequences. The main technical tools here are effective versions of the two Borel-Cantelli lemmas (Feller [25,200-2]).

3.3.1 Lemma Let μ be a computable measure, $(A_n)_{n\in\mathbb{N}}$ a recursive sequence of Σ_1 sets in 2^ω such that each μA_n is computable (uniformly in n) and $\sum_n \mu A_n$ converges recursively[3]. Then $N := \bigcap_n \bigcup_{k\geq n} A_k$ is a total recursive sequential test with respect to μ.

Proof Obviously N is Π_2. $\mu\bigcup_{k\geq n} A_k$ is computable since for $m_2 > m_1$,

$$\mu\bigcup_{k=n}^{m_2} A_k - \mu\bigcup_{k=n}^{m_1} A_k \leq \sum_{k=m_1}^{m_2} \mu A_k$$

and decreasing to 0 since $\sum_n \mu A_n$ converges. Now apply remark 3.2.3.11. ∎

Seeing that one automatically obtains a *total* recursive sequential test, starting from the natural condition that $\sum_n \mu A_n$ converges constructively, one might wonder whether there exists some condition which yields only recursive sequential tests. There is, namely:

$$\text{for some total recursive } f: \omega \to \omega, \text{ for all n: } \sum_{k\geq f(n)} \mu A_k \leq 2^{-n};$$

but, in practice, whenever in an application of the first Borel-Cantelli lemma the latter condition is satisfied, so is the more exacting condition of lemma 3.3.1. This illustrates a general phenomenon: it is hard to come up with *natural* examples of recursive sequential tests which are not total (they may come from the theory of Martingales, to which the next section is devoted). Nevertheless, it will become clear in the sequel and especially in Chapter 5, that Martin-Löf's concept has immense technical advantages.

Likewise we have the following effective analogue of the second Borel-Cantelli lemma:

3.3.2 Lemma Let μ be a computable measure, $(A_n)_{n\in\mathbb{N}}$ a recursive sequence of independent Σ_1 sets in 2^ω such that $\sum_n \mu A_n$ diverges and μA_n is computable (uniformly in n). Then $\bigcup_n \bigcap_{k\geq n} A_k^c$ is contained in a total recursive sequential test with respect to μ.

Proof By the second Borel-cantelli lemma (Feller [25,201]), $\mu\bigcap_{k\geq n} A_k^c = 0$, for each n. $\bigcap_{k\geq n} A_k^c$ is a Π_1 set, which by remark 3.2.3.11, can be taken to be a total recursive sequential test. Now apply lemma 3.2.3.8. ∎

As an application of the preceding material, we shall now prove the strong law of large numbers for (weakly) random sequences. The probabilistic argument is copied from Feller [25,259], but we have to complicate the construction to ensure computability.

3.3.3 Theorem Let $\mu = \prod_n(1-p_n,p_n)$ be a computable product measure. For a recursive and dense (in (0,1)) set of computable reals ε, the sets

$$\{x \in 2^\omega | \forall m \exists n \geq m \, |\frac{1}{n}\sum_{k=1}^{n} x_k - \frac{1}{n}\sum_{k=1}^{n} p_k| > \varepsilon\}$$

are contained in a total recursive sequential test with respect to μ.

Proof Choose > 0 and rational. Let

$$A_k := \{ x \in 2^\omega | \exists n(2^{k-1} < n \leq 2^k \,\&\, |\frac{1}{n}\sum_{k=1}^{n} x_k - \frac{1}{n}\sum_{k=1}^{n} p_k| > \varepsilon)\}$$

The obvious candidate for a total recursive sequential test is $\bigcap_n \bigcup_{k\geq n} A_k$, but there is a slight problem here: μA_k need not be computable, even if ε is rational; for we might not be able to decide whether

$$\frac{1}{n}\sum_{k=1}^{n} x_k - \frac{1}{n}\sum_{k=1}^{n} p_k = \varepsilon$$

for pathological μ. One may circumvent this problem by restricting μ to be strongly computable (definition 3.2.1.1) or by choosing ε such that we know *in advance* that this situation cannot occur. Now every number

$$|\frac{1}{n}\sum_{k=1}^{n} x_k - \frac{1}{n}\sum_{k=1}^{n} p_k|$$

is of the form

$$|\frac{m}{n} - \frac{1}{n}\sum_{k=1}^{n} p_k| =: a_{mn}, \text{ where } m \leq n.$$

Obviously each a_{mn} is computable and the sequence $(a_{mn})_{m,n\in\mathbb{N}}$ is recursive. By repeated diagonalisation one may then construct a recursive sequence of computable reals $(\varepsilon_j)_{j\in\mathbb{N}}$ such that $\lim_{k\to\infty} \varepsilon_j = 0$ and for all j, n and m: $\varepsilon_j \neq a_{mn}$.

Now if we set, in the definition of A_k, ε equal to ε_j, we do have that A_k is Σ_0, (A_k) is recursive and μA_k is computable (uniformly in k). (A similar argument occurs in 4.4, where we need an effective version of the Baire Category Theorem to effect the iterated diagonalisation.) The argument then follows familiar probabilistic lines: if s_n is the variance of μ at the n^{th} coordinate, then $s_n = p_n\cdot(1-p_n)$ and since for all n, $s_n\cdot n^{-2} \leq n^{-2}$, $\sum_n s_n\cdot n^{-2} \leq \sum_n n^{-2} = \pi^2/6$ converges constructively by lemma 3.2.3.1. By Kolmogorov's inequality (Feller [25,234]),

$$\mu A_k \le 4\cdot\varepsilon_j^{-2}\cdot s_{2^k}\cdot 2^{-2k}$$

hence

$$\sum_k \mu A_k \le 4\cdot\varepsilon_j^{-2}\cdot\sum_k 2^{-2k}\cdot\sum_{n=1}^{2^k} s_n = 4\cdot\varepsilon_j^{-2}\sum_n s_n\cdot\sum_{2^k\ge n} 2^{-2k} \le 8\cdot\varepsilon_j^{-2}\sum_n s_n\cdot n^{-2}.$$

Now apply lemma 3.3.1. □

The law of the iterated logarithm can be proved similarly, this time using both effective Borel-Cantelli lemmas and the proof of the law of the iterated logarithm in Feller [25,205]. In Chapter 4 we shall construct examples of probabilistic laws not hitherto considered in the literature.

In conclusion of this section, let us investigate what happens if we drop the requirement in lemma 3.3.3, that the product measure μ be computable. Since there is now no sense in requiring the μA_k to be computable, we may choose rational $\varepsilon > 0$. We then have that the sequence (A_k) is recursive in μ and that the upper bounds on μA_k are given by a recursive function of k, by the inequality $\sum_n s_n\cdot n^{-2} \le \sum_n n^{-2}$. This illustrates our claim in 3.2.4, that the most useful concept of effective probabilistic law for *arbitrary* μ is obtained if we replace in definition 3.2.1.2, "Π_2" by "Π_2 in μ".

3.4 Martingales As a technical prelude to 3.5, where we examine Martin-Löf's original way of introducing random sequences, we present a different a characterisation of random sequences, using Martingales, Ville's formalisation of the concept of a gambling strategy.

Von Mises' axioms for Kollektivs were stated in terms of admissible place selections and did not mention gambling strategies. The second axiom, however, was explained informally as the "principle of the excluded gambling strategy"; so it is natural to ask whether all gambling strategies can be represented as place selections. As we have seen in 2.6.2, Ville [99] showed that such is not the case. He argued that place selections left one essential element of gambling strategies out of consideration: the possibility to vary one's stakes from one bet to the next. We now give a rapid introduction to the definition and main properties of gambling strategies with variable stakes, so-called *Martingales*, and afterwards discuss their interpretation.

The stakes are given by functions B_0, B_1: $2^{<\omega} \to \mathbb{R}^+$ as follows: we bet $B_0(w)$ on the event that w is followed by 0 and $B_1(w)$ on the event that w is followed by 1. If V(w) denotes our capital after the sequence w has occurred, we must have (we exclude loans): $B_0(w) + B_1(w) \le$

V(w). We say that the game played with strategy V is *fair* if, for each n, the expected capital after the n+1th trial is equal to the capital after the nth trial. To formalize this condition of fairness we need a probability measure μ on 2^ω. Having a probability measure, we may then define Martingales.

3.4.1 Definition Let μ be a measure on 2^ω. $V: 2^{<\omega} \to \mathbb{R}^+$ is a (positive) Martingale with respect to μ if $V(\langle\rangle) < \infty$ and for all w:

$$V(w) = \frac{\mu[w0]}{\mu[w]} \cdot V(w0) + \frac{\mu[w1]}{\mu[w]} \cdot V(w1).$$

The relation to the usual probabilistic concept (see e.g. Feller [26] and Neveu [77]) should be clear: let B_n denote the algebra generated by the cylinders of length n, $V_n: 2^{<\omega} \to \mathbb{R}^+$ the function defined by $V_n(x) = V(x(n))$, then the sequence (V_n) is a Martingale (in the usual sense) with respect to μ and the filtration (B_n).

We say that a Martingale V is *successful* on a sequence x if $\limsup_{n\to\infty} V(x(n)) = \infty$. The following lemma, called Kolmogorov's inequality for Martingales by Feller [26,242], but which occurs already in Ville [99,100], shows that Martingales (with respect to μ) are almost never (again with respect to μ) successful.

3.4.2 Lemma Let V be a Martingale with respect to μ, then for $a \in \mathbb{R}^+$

$$\mu\left\{x \in 2^\omega \mid \exists n\,(V(x(n)) > a)\right\} \le \min\left(\frac{V(\langle\rangle)}{a}, 1\right).$$

As a consequence,

$$\mu\left\{x \in 2^\omega \mid \limsup_{n\to\infty} V(x(n)) = \infty\right\} = 0.$$

3.4.3 Examples

1. Let Φ be a place selection (see definition 2.5.1.1). Choose $p,q \in (0,1)$. Define a Martingale V_q with respect to the measure μ_p by
 (i) $V_q(\langle\rangle) = 1$
 (ii) if $\phi(w) = 0$, let $V_q(w) = V_q(w0) = V_q(w1)$
 (iii) if $\phi(w) = 1$, put $V_q(w0) = V_q(w)\cdot(1-q)/(1-p)$ and $V_q(w1) = V_q(w)\cdot q/p$.

Then V_q is a Martingale with respect to μ_p, and one can show that $\Phi(x) \notin$ LLN(p) iff for some q, $\limsup_{n\to\infty} V_q(x(n)) = \infty$ (see Schnorr [88,78-82]). (For the definition of LLN(p), see 2.3.2.3.) So Martingales are indeed generalisations of place selections.

2. Likelihood ratios. Let μ_0, μ_1 be probability measures on 2^ω. Put $V(w) = \mu_0[w]/\mu_1[w]$, then V is called the *likelihood ratio* of μ_0 and μ_1 and V is a Martingale with respect to μ_1:

$$\frac{\mu_1[w0]}{\mu_1[w]} \cdot \frac{\mu_0[w0]}{\mu_1[w0]} + \frac{\mu_1[w1]}{\mu_1[w]} \cdot \frac{\mu_0[w1]}{\mu_1[w1]} = \frac{\mu_0[w]}{\mu_1[w]}.$$

Note that some of the Martingales V defined in 1. are also of this form: if the place selection Φ is the identity, $V_q(w) = \mu_q[w]/\mu_p[w]$. In fact, any Martingale in the sense of definition 3.4.1 can be written in the form of a likelihood ratio: if V is a Martingale with respect to μ with $V(\langle\rangle) = 1$, and if we define $\mu'[w] := V(w)\cdot\mu[w]$, them μ' determines a probability measure and V is the likelihood ratio of μ' and μ.

In order to obtain a rich supply of recursive sequential tests, we now introduce some computability considerations, in particular a weak notion of computability for measures.

3.4.4. Definition A measure μ on is called *subcomputable* if the set

$$\{<w,a> \in 2^{<\omega} \times \mathbb{Q} \mid \mu[w] > a\}$$

is Σ_1. A Martingale V is called *subcomputable* if the set

$$\{<w,a> \in 2^{<\omega} \times \mathbb{Q} \mid V(w) > a\}$$

is Σ_1.

These concepts are not very natural from the point of view of probability theory, but the representation of recursive sequential tests in terms of Martingales will make clear why they are useful. The following two lemmas can be found in Schnorr [88, 38-44], but, stripped of their recursive content, they go back to Ville [99,87-93].

3.4.5 Lemma Let V be a subcomputable Martingale with respect to some measure μ. Then $\{x \mid \forall k \exists n V(x(n)) > 2^k\}$ is a recursive sequential test with respect to μ.

Proof By subcomputability, the set $\{x \mid \forall k \exists n V(x(n)) > 2^k\}$ is Π_2. Without loss of generality we may assume $V(\langle\rangle) \leq 1$; then by lemma 3.4.2, $\mu\{x \mid \exists n V(x(n)) > 2^k\} \leq 2^{-k}$. □

3.4.6 Example Likelihood ratios. Let μ_0, μ_1 be computable measures on such that μ_1 is not absolutely continuous with respect to μ_0. Then there exists a recursive sequential test N with respect to μ_0 such that $\mu_1 N > 0$. Indeed, put $N = \{x \mid \forall k \exists n V(x(n)) > 2^k\}$, where $V(w) = \mu_1[w]/\mu_0[w]$. By the preceding lemma, N is a recursive sequential test with respect to μ_0. The Lebesgue decomposition of μ_1 with respect to μ_0 can be written as

$$\mu_1 = \int \lim_{n\to\infty} V(x(n))d\mu_0(x) \;+\; 1_N d\mu_1,$$

so that if μ_1 is not absolutely continuous with respect to μ_0, then $\mu_1 N > 0$.

We now prove a converse to lemma 3.4.5.

3.4.7 Lemma Let N be a recursive sequential test with respect to some computable measure μ. Then there exists a subcomputable Martingale V with respect to μ such that $N \subseteq \{x| \forall k\exists n V(x(n)) > 2^k\}$.

Proof Write $N = \bigcap_n O_n$ as in definition 3.2.1.2. Put $V(w) := \sum_n n\cdot\mu([w]\cap O_n)\cdot\mu[w]^{-1}$. Then V is a Martingale with respect to μ:

$$\frac{\mu[w0]}{\mu[w]}\cdot V(w0) + \frac{\mu[w1]}{\mu[w]}\cdot V(w1) =$$

$$\frac{\mu[w0]}{\mu[w]}\sum_n n\cdot\mu([w0]\cap O_n)\cdot\mu[w0]^{-1} + \frac{\mu[w1]}{\mu[w]}\sum_n n\cdot\mu([w1]\cap O_n)\cdot\mu[w1]^{-1} = V(w).$$

Furthermore, $V(\langle\rangle) = \sum_n n\cdot\mu O_n \leq \sum_n n\cdot 2^{-n} < \infty$. V is subcomputable since for any set O in Σ_1, $\{<w,a> \in 2^{<\omega} \times \mathbb{Q} \mid \mu([w]\cap O) > a\}$ is itself Σ_1.

Lastly, $N \subseteq \{x| \forall k\exists n V(x(n)) > 2^k\}$:

if $x \in \bigcap_n O_n$, then $\forall n\exists m\geq n\forall m'\geq m\ (\mu([x(m')]\cap O) = \mu[x(m')])$, which implies $\forall n\exists m\geq n\forall m'\geq m\ (V(x(m')) \geq n)$ and this in turn implies $\lim_{n\to\infty} V(x(n)) = \infty$. □

The preceding lemmas may be combined to obtain a characterisation of random sequences along the lines suggested by Ville, namely as sequences which do not admit a successful gambling strategy (where the latter are taken to be Martingales):

3.4.8 Lemma Let μ be a computable measure. Then $x \in R(\mu)$ iff for all subcomputable Martingales V with respect to μ: $\limsup_{n\to\infty} V(x(n)) < \infty$. (Note that, as a consequence of the proof, the latter condition is in turn equivalent to: for all subcomputable Martingales V with respect to μ: $\lim_{n\to\infty} V(x(n)) < \infty$.)

We may now give a more precise discussion of Ville's objection, that not all gambling strategies can be represented as place selections. Recall that Ville could construct $x \in 2^\omega$ which satisfy (where $C(\frac{1}{2})$ is the set of Church-random sequences defined in 2.5.1.7.):

$$x \in C(\tfrac{1}{2}) \text{ and for all n, } \frac{1}{n}\sum_{k=1}^{n} x_k \geq \tfrac{1}{2}.$$

The second property is in contradiction with the law of the iterated logarithm. By the results in section 3.3, the set of sequences not satsfying the law of the iterated logarithm (for the measure λ) is a (total) recursive sequential test with respect to λ. The last lemma then implies that for some Martingale V with respect to λ: $\lim(\sup)_{n\to\infty} V(x(n)) = \infty$. This Martingale V cannot be obtained from a place selection (in contradistinction to the Martingales V_q defined in example 3.4.3). Hence, to give a precise formulation of the "principle of the excluded gambling strategy", one should define Kollektivs using Martingales, not just place selections.
We do not think that this result is a problem for von Mises, who after all does not require that there is no successful gambling strategy, *of whatever kind,* on a Kollektiv. Furthermore, Ville's argument assumes without further ado that Martingales constitute a good formalisation of fair games and indeed that the notion of fairness is itself clear and unproblematic. But that may not be so.

We formulated fairness as follows: a game is fair if, for each n, the expected capital after the $n+1^{th}$ trial is equal to the capital after the n^{th} trial. But taking expectations requires some probability measure; and which probability measure should one consider? Adopting the standpoint of strict frequentism, one might be inclined to say that expectations have to be computed with respect to the measures P_n on 2^n, induced by Ville's Kollektiv x via combination as explained in 2.4 (so that in this case the measures P_n are uniform distributions on 2^n). In other words, one might think that the pay-offs for a game *on x* should be determined by the limiting relative frequencies *in x*. Ville's example shows that, when two people agree to play a game according to *this* concept of fairness, one of them may have a successful gambling strategy on Kollektivs of the type constructed by Ville. What's more, in Chapter 4 we shall show that there exist product measures $\mu = \prod_n(1-p_n,p_n)$ with $\mu C(\tfrac{1}{2}) = 1$, but $\mu\{x | \limsup_{n\to\infty} V(x(n)) = \infty\} = 1$, for some computable Martingale V (for instance, one may take $p_n = \tfrac{1}{2}(1 + (n+1)^{-\frac{1}{2}})$). Thus, the first tentative "operational" definition of fairness apparently has to be rejected: although it applies for games with fixed stakes (i.e. place selections), it is not applicable to games with variable stakes. However, it does not seem to follow from the strict frequency interpretation that this is the *only* way in which fairness can be defined.

The intuitive idea behind fairness seems to be that it makes sense to speak of "probability of heads at the n toss". This notion of fairness is clear on the propensity interpretation (or perhaps

one should say: not less clear than the propensity interpretation), so it is not surprising that Ville has no qualms about fairness. But, as we have seen in the previous chapter, from the point of view of strict frequentism one may speak of probabilities at specific coordinates only with reference to Kollektivs $\xi \in (2^{\omega})^{\omega}$. In particular, one must consider infinitely many (infinite) runs of the mechanism that produces the Kollektivs (with which the game has to be played) and then count the limiting relative frequencies *in each coordinate*; and *these* probabilities must determine the pay-offs. Now with this definition, a Martingale with respect to the uniform distribution would no longer be considered fair for a game played with Kollektivs of Ville's type: if each ξ_k is of this type, then the probability of 1 at the n^{th} coordinate will be larger than $\frac{1}{2}$.
In conclusion, we may say that Ville's argument is not relevant for the question how to define Kollektivs, but rather for the examination of the probabilistic assumptions that go into the intuitive notion of a fair game. For games with variable stakes, fairness seems to involve a reference to probabilities at some specified coordinate. An adherent of the propensity interpretation will have no difficulty recognizing such probabilities, but the strict frequentist can only introduce them using a Kollektiv of Kollektivs. If for some reason or other his data consist in only one Kollektiv $x \in 2^{\omega}$, in other words, if his data consist only in a distribution over $\{0,1\}$, he cannot decide whether some proposed game is in fact fair. To some, the strict frequentist conception of fairness may seem artificial; but this seeming unnaturalness serves to confirm the impression that the *instinctively* adopted interpretation of probability is the propensity interpretation. Interestingly, the only reference to Martingales that I could find in von Mises' published works expresses his incomprehension:

> Jusqu'ici je n'ai pu encore saisir l'idée essentielle qui serait à la base de la notion de "martingale" et de toute la théorie de M. Ville. Mais je ne doute point que, une fois son livre paru, on s'apercevra à quel point il aurait réussi à concilier les fondements classiques du calcul des probabilités avec la notion moderne du collectif [72,67].

Needless to say, there are no technical obstacles to a treatment of Martingales in von Mises' theory; as for the interpretation of the results obtained, we need not repeat here the observations made in 2.4.3 à propos of the strong limit laws.

We now continue our discussion of the technical aspects of the relationship between randomness and Martingales. In section 3.5 we need more detailed information on the Martingale constructed in lemma 3.4.7. This construction has the following analytical meaning:

3.4.9 Corollary Let N be a recursive sequential test with respect to μ and let V be the Martingale constructed in the proof of lemma 3.4.7. If we put $\mu'[w] := V(w)\cdot\mu[w]$, then μ' is absolutely continuous with respect to μ.

Proof Put

$$f(x) := \sum_n n \cdot 1_{O_n}(x)$$

then f is in $L^1(\mu)$ and f is the density of μ' with respect to μ:

$$\mu'[w] = \sum_n n \cdot \mu([w] \cap O_n) = \int_{[w]} f d\mu.$$

If μ is Lebesgue measure, one can show that the distribution function of μ' has derivative equal to $+\infty$ at all points of N. □

The parallel theory for *total* recursive sequential tests is considerably less smooth.

3.4.10 Definition The Martingale V is *computable* if for some recursive function g: $2^{<\omega} \times \omega \to \mathbb{Q}$: $\forall n \forall w\ |V(w) - g(w,n)| < 2^{-n}$.

Inspecting the proof of lemma 3.4.7 we see that

3.4.11 Lemma Let μ be a computable measure and let N be a total recursive sequential test with respect to μ. Then there exists a computable Martingale V such that N is contained in $\{x | \forall k \exists n\ V(x(n)) > 2^k\}$.

Proof Write $N = \bigcap_n O_n$ as in definition 3.2.1.3 and define V as in lemma 3.4.7. It suffices to show that the expression $\sum_n n \cdot \mu([w] \cap O_n)$ is computable uniformly in w. Since $n \cdot \mu([w] \cap O_n) \leq n \cdot \mu O_n$ and $\sum_n n \cdot \mu O_n$ is computable, this follows from lemma 3.2.3.1. □

In this case the converse, namely

> If V is a computable Martingale with respect to a computable measure μ, then $N = \{x | \forall k \exists n\ V(x(n)) > 2^k\}$ is a total recursive sequential test with respect to μ,

causes some trouble. Obviously N, so defined, is a recursive sequential test; but we also need to show that $\mu\{x | \exists n\ V(x(n)) > 2^k\}$ is computable (uniformly in k). The obvious way to do this, is to use lemma 3.2.3.1 and first passage times: $\mu\{x | \exists n\ V(x(n)) > 2^k\} = \sum_m \mu\{x | V(x(m-1)) \leq 2^k < V(x(m))\}$; and one could hope that there is some recursive sequence of computable reals (a_m) such that $\mu\{x | V(x(m-1)) \leq 2^k < V(x(m))\} \leq a_m$ and $\sum_m a_m$ converges recursively.

However, it is impossible to choose such a sequence (a_m) independent of the Martingale under consideration, since for each m, one may construct a Martingale V' such that $\mu\{x|V'(x(m-1)) \leq 2^k < V'(x(m))\} = 2^{-k}$. Hence, knowledge of the specific structure of the Martingale is necessary. This is the reason why the Martingale convergence theorem in Bishop [5,225] has to be proven under additional assumptions on the Martingales.

In order to circumvent this problem, Schnorr [88,70-7] proposed a different definition of the total recursive sequential tests associated with computable Martingales.

3.4.12 Definition Let $f: \mathbb{N} \to \mathbb{R}^+$ be a computable function, V a computable Martingale. The set $N = \{x| \limsup_{n\to\infty} V(x(n))\cdot f(n)^{-1} > 0\}$ is called the *nullset of order* f *associated to* V.

In other words, only those sequences are put into the nullset on which V can grow sufficiently fast. With the help of the following lemma one may then show that N is indeed contained in a total recursive sequential test.

3.4.13 Lemma (Schnorr [88,72]) Let V be a computable Martingale. For any rational $\varepsilon > 0$, one can construct a *recursive* Martingale $V': 2^{<\omega} \to \mathbb{Q}^+$ such that for all w, $V'(w) \geq V(w)$ and $V'(w) - V(w) \leq \varepsilon$.

3.4.14 Lemma Let V be a computable Martingale with respect to μ and let be N as in definition 3.4.12. Then N is contained in a total recursive sequential test with respect to μ.

Sketch of proof The total recursive sequential test can be defined by

$$M = \{x|\forall k\exists n\, (V'(x(n)) > 2^k\cdot V'(\langle\,\rangle)\ \&\ V'(x(n)) > f(n))\},$$

where V' is the Martingale constructed in the previous lemma. For a verification that M is indeed a total recursive sequential test, see Schnorr [88,73][4]. □

Although Schnorr claims that the concept of randomness itself suggests consideration of Martingales together with order functions (a sequence should be non-random only if we can detect the non-randomness sufficiently fast [88,70]), we think that definition 3.4.12 is interesting only in those cases in which it follows *from the definition* of a Martingale V that it must grow with speed f on some given nullset. Schnorr has established some results of this kind (see chapter 10 of [88]). In other cases, Schnorr's way out seems to be adhoc.

The considerations of this section therefore suggest a concept of randomness which might be different from that of Schnorr.

3.4.15 Definition Let μ be a computable measure. x is called *Martingale-random* with respect to μ (notation: $x \in R_M(\mu)$) if for all computable Martingales V with respect to μ:
$\limsup_{n\to\infty} V(x(n)) < \infty$.

By lemma 3.4.11, $R_M(\mu) \subseteq R_w(\mu)$; it is difficult to say whether we in fact have equality.

In conclusion of this section we point out that tests of the form $\{x | \forall k \exists n\ V(x(n)) > 2^k\}$, for computable Martingales V, share one of Schnorr's desiderata with total recursive sequential tests: the existence of recursive sequences outside these sets (cf. corollary 3.2.3.7 and the discussion which follows it).

3.4.16 Lemma Let V be a computable Martingale. Then for some recursive x:
$\limsup_{n\to\infty} V(x(n)) < \infty$.

Proof Let V' be the Martingale constructed in lemma 3.4.13. Choose rational $\delta > 0$ and define a recursive binary tree T by T := $[w | V'(w) < V'(\langle\rangle) + \delta\}$. For every $w \in T$, $w0 \in T$ or $w1 \in T$ by the Martingale property, and we can decide which by the computability of V. Consequently the leftmost infinite branch of T is recursive. □

3.5 Randomness via statistical tests

Originally, Martin-Löf [62] introduced the set of random sequences $R(\lambda)$ as follows: a sequence is random wirh respect to if λ it is not rejected at arbitrarily small levels of significance by any (effective) statistical test for λ. Since this way of introducing randomness raises some interesting problems of its own, we shall now give it a separate treatment. To do so, we must first recall some elementary notions concerning statistical tests. As always, we consider an experiment (or measurement) with two outcomes, 0 and 1.

3.5.1 Types of statistical tests We want to test the hypothesis H_0, that the probability of the outcome 1 of an experiment equals p. We may divide tests of H_0 into two classes:
(a) We may distinguish between tests of H_0 which refer to some alternative hypothesis, the so-called *hypothesis tests*, and *significance tests*, which reject H_0 when an outcome sequence is observed which has sufficiently low probability under the hypothesis H_0, without consideration of alternative hypotheses;

(b) We may also distinguish between tests which use a *fixed sample size*, i.e. tests where the number of repetitions of the experiment is fixed before the execution of the experiment, and tests which are *sequential*, where the data themselves decide how large the sample is to be.

We now proceed to a detailed description. Let us first assume that we have a fixed sample size, say n; hence the set of possible outcomes, the sample space, is 2^n. Under the hypothesis H_0 an outcome sequence w in 2^n is assigned probability $\mu_p[w]$[5]. In essence, a significance test for the hypothesis H_0 consists in a partition of the sample space 2^n in disjoint pieces S_0 and S_1. Observation of an outcome sequence w in S_0 leads to rejection of H_0. Observation of w in S_1 does *not* lead to rejection of H (in practice, this will mean that H_1 is given the benefit of the doubt). S_0 is often called a *critical region*. The probability of S_0 under H_0, namely

$$\sum_{w \in S_0} \mu_p[w]$$

is called the *size* of the test and can be interpreted as the relative frequency of unwarranted rejections of H_0 were this test to be executed very often. Obviously we want the size to be small; how small depends on the importance we attach to the hypothesis.
Usually S_0 and S_1 are determined via a *test statistic*, a function $t: 2^{<\omega} \to \mathbb{R}^+$ which can be seen as a measure of the discrepancy between hypothesis and data. Accordingly, the critical region S_0 is of the form:

$$S_0 = \{ w \in 2^n \mid t(w) > a \}$$

where a is adjusted so as to have, for some preassigned *significance level* α,

$$\sum_{t(w)>a} \mu_p[w] \leq \alpha.$$

How should we choose such test statistics? Obviously not every S_0 of small probability can reasonably be interpreted as a critical region for H_0; e.g. for n = 1000 and p = ½, the set of words in 2^n with 500 ones has very small probability, but to take this set for our critical region would be a silly choice indeed.
This line of reasoning shows that the choice of a test statistic is a delicate matter, and it is still a subject of lively debate whether this choice can be effected at all without the consideration of hypotheses alternative to H_0. In the survey by Cox [18], the issue is stated as follows:

> The central philosophical point concerns whether it is sensible to find evidence against a hypothesis solely because an outcome of relatively low probability has occurred, and without regard to possible alternative explanations. If the labelling of the sample points in the sample space is totally arbitrary and no other information is available, there seems to be no option but to use the absolute test [i.e. significance test in the sense defined above]; such situations do, however, seem quite exceptional in applications [18,53].

Cox' first question is answered with an emphatic *no* by the founding fathers of modern statistical theory, Neyman and Pearson:

> It is indeed obvious, upon a little consideration, that the mere fact that a particular sample may be expected to occur very rarely in sampling from [a certain population] would not in itself justify the rejection of the hypothesis that it had been so drawn [from that population], if there were no other more probable hypothesis conceivable [78,4].

It is clear from Martin-Löf's statistical work [65;66] that he rejects this view (or perhaps one should say that his concept of "alternative hypothesis" is much wider than that of Neyman and Pearson); but let us first expound the view of Neyman and Pearson.

To eliminate the possibility of disastrous choices of the test statistic, Neyman and Pearson propose to introduce the consideration of alternative hypotheses. In the simplest case, we have only one alternative H_1 to H_0, where H_1 states that the outcome 1 has a different probability q ≠ p. A test for H_1 against H_0 is again specified by a partition (S_0,S_1) of 2^n: S_0 corresponds to rejection of H_0 (and acceptance of H_1) and S_1 corresponds to acceptance of H_0 (and rejection of H_1).
In this case, there are two possibilities for wrong decisions: rejecting H_0 when it is true (*type I error* ; the probability of type I error is called the *size* of the test) and accepting H_1 when it is in fact false (*type II error*; 1 – the probability of type II error is called the *power* of the test) . As in the case of a significance test, the probability of type I error is equal to

$$\sum_{w\in S_0} \mu_p[w].$$

But whereas it makes no sense to speak of type II error for a significance test, for lack of an alternative hypothesis, we may compute the probability of type II error here as

$$\sum_{w\in S_1} \mu_q[w].$$

The interpretation of power is the same is that of size: it measures the performance of the test were it used a large number of times.

The distinction between type I and type II errors allows us to discredit the test defined on p. 81, which rejects H_0: p = ½, upon observation of an outcome sequence of length 1000 with 500 ones. Clearly, in this case, for any q ≠ ½,

$$\sum_{w\in S_1} \mu_q[w]$$

is large; and it will be required of a good test that both types of errors are simultaneously small (they are of course not independent).

Call a test of H_0 against H_1 *most powerful* of level α if

$$\sum_{w \in S_1} \mu_q[w]$$

is as small as is compatible with

$$\sum_{w \in S_0} \mu_p[w] \leq \alpha.$$

In this particular situation, most powerful tests exist and can even be given explicitly; this is the content of the

Neyman-Pearson Lemma[6] For suitable constant c (depending on α, the sample size n as well as on the hypotheses involved): if a partition (S_0,S_1) of 2^n is defined by

$$S_0 = \left\{ w \in 2^n \mid \frac{\mu_q[w]}{\mu_p[w]} > c \right\},\ S_1 = 2^n - S_0,$$

then (S_0,S_1) is the most powerful level α test of H_0 against H_1.

The preceding exposition of significance tests and hypothesis tests proceeded on the assumption of a fixed sample size. We now relax this assumption and generalize the description to situations in which the sample size is not fixed beforehand. The following description of sequential tests is borrowed from Wald [102,22].
An essential feature of the sequential test, as distinguished from the [fixed sample size test] is that the number of observations required by the sequential test depends on the outcome of the observations and is therefore not predetermined but a random variable. The sequential method of testing a hypothesis H may be descibed as follows. A rule is given for making one of the following decisions at any stage of the experiment (at the m^{th} trial for each integral value of m):
(1) to accept the hypothesis H
(2) to reject the hypothesis H
(3) to continue the experiment by making an additional observation.
Thus, such a test procedure is carried out sequentially. On the basis of the first observation, one of the aforementioned three decisions is made. If the first or the second decision is made, the process is terminated. If the third decision is made, a second trial is performed [...]. The process is continued until either the first or the second decision is made. The number n of observations required by such a test procedure is a random variable, since the value of n depends on the outcome of the observations.

Formally, a sequential test for *hypothesis* testing may be described as follows. We have a test statistic t: $2^{<\omega} \to \mathbb{R}^+$ and constants A,B such that

(1) if t(w) > A and for all $v \subset w$, $B \leq t(v) \leq A$, reject H_0 (accept H_1);

(2) if t(w) < B and for all $v \subset w$, $B \leq t(v) \leq A$, reject H_1 (accept H_0);

(3) if for all $v \subseteq w$, $B \leq t(v) \leq A$, go on testing.

If the measure μ_p corresponds to H_0, and μ_q to H_1, the probabilities of type I and type II errors can be computed as follows:

$$\text{size } = \mu_p\{x | \exists n\, (t(x(n)) > A \;\&\; \forall m<n\, (B \leq t(x(m)) \leq A))\}$$

$$1 - \text{power} = \mu_q\{x | \exists n\, (t(x(n)) < B \;\&\; \forall m<n\, (B \leq t(x(m)) \leq A))\}.$$

Obviously, we want both types of errors to be as small as possible. Again, for the simple situation of testing one hypothesis against another, there is an optimum result: for given significance level α, one can determine constants A and B such that the *likelihood ratio test* defined by putting $t(w) := \mu_q[w]/\mu_p[w]$ in the decision rules above, is the most powerful test of significance level α.

In a sequential *significance* test we are concerned with one hypothesis H_0 only. In this case the set-up is as follows: we have a test statistic t and a constant A such that H_0 is rejected on the basis of data w if t(w) > A; otherwise we go on testing. Of course A is adjusted so as to achieve a prescribed significance level α.

The difficulties we pointed out for fixed sample size tests seem to be even more severe in the sequential case. Not only does the choice of a test statistic present a problem in the absence of an alternative hypothesis; but there seems to be no rational basis for a decision to give H_0 the benefit of the doubt, since there does not appear to be a non-arbitrary way to determine a constant B such t(w) < B entails the decision to stop testing.

So it seems that sequential significance tests are useful for rejecting hypotheses, rather than for accepting them. This point should be borne in mind when we discuss Martin-Löf's definition of randomness via statistical tests.

3.5.2 Effective statistical tests It is now easy to view definition 3.2.1.2 as a formalisation of sequential (significance and hypothesis) tests. Let μ be a computable measure on 2^{ω}. μ need not be of the form μ_p, since we also wish to study tests applicable in situations not involving independent repetitions of the same experiment. We interpret μ as the nullhypothesis to be tested. Typically, μ contains information about the underlying model (Markov chain, independent repetitions) as well as about the parameters of the model. We are interested in arbitrarily small levels of significance; we may take these levels to be of the form 2^{-k}, $k \in \mathbb{N}$.

Now, in practice, a test statistic t: $2^{<\omega} \to \mathbb{R}^+$ will be a computable function. This implies that the set $\{x | \exists n\, (t(x(n)) > A \;\&\; \forall m<n\, (t(x(n)) \leq A))\}$ is Σ_1 for suitable choices of A, namely for

those computable A which do not occur in the range of t. As in section 3.3.3, we may construct a recursive and dense (in $\mathbb{R}^+$) set of such A's by iterated diagonalisation.
Clearly, if $(A_k)_{k\in\mathbb{N}}$ is a recursive set of computable reals which do not occur in the range of t, the set $\{x|\forall k\exists n\ (t(x(n)) > A_k\ \&\ \forall m<n\ (t(x(m)) \leq A_k))\}$ is $\prod_2$.

If the sequence (A_k) is such that $\mu\{x|\exists n\ (t(x(n)) > A_k\ \&\ \forall m<n\ (t(x(m)) \leq A_k))\} \leq 2^{-k}$, we have arrived at a recursive sequential test with respect to μ. This, in a nutshell, is the statistical motivation of definition 3.2.1.2. Note that the test statistics are subject only to restrictions of a recursion theoretic nature.

3.5.3 Discussion Seeing that every effective statistical test corresponds to a recursive sequential test, we may now ask for a converse: does every recursive sequential test determine an acceptable statistical test? To settle this question,we have to investigate the influence of the reservations concerning significance tests, expressed above, on the proposed definition of randomness. In essence, these reservations come down to this: it is impossible to construct good test statistics without consideration of alternative hypotheses. "Good" here means: the test based on the statistic should not reject the hypothesis when it is intuitively true.
This danger can be avoided if we require that the critical region is in a sense minimal: only reject the nullhypothesis on the basis of data w if w is more plausible on some other hypothesis. In Lévy's words

> Si donc en présence d'une suite remarquable nous excluons la première hypothèse [of the random origin of the data] ce n'est pas que le hasard ait *a priori* moins de chance de la produire qu'une autre; c'est qu'une cause autre que le hasard a plus de chance de la produire [57,92].

Does the alternative hypothesis necessarily have to be of probabilistic origin, stating a different value of a parameter, or perhaps a different model? In other words, should the condition "if w is more *plausible* on some other hypothesis" be interpreted as "if w is more *probable* on some other hypothesis"? The talk of *chance* in the above quotation strongly suggests so and, as we have seen, this was certainly the view of Neyman and Pearson.

If this is indeed the case, we may be led to a notion of randomness which is likely to be different from that of Martin-Löf (or Schnorr), depending upon the definition of "alternative hypothesis" in this abstract setting. The function of the alternative hypothesis ν is to assign a high probability to events to which μ assigns a low probability. If we take "high" and "low" in an absolute sense, so that "high" means "close to 1" and "low" "close to 0", we may regard ν as an alternative to μ if $\nu\perp\mu$.

3.5.3.1 Definition Let μ be a computable measure. Put

$$R_H(\mu) := \left\{ x \mid \text{for all subcomputable measures } \nu\text{: } \limsup_{n\to\infty} \frac{\nu[x(n)]}{\mu[x(n)]} < \infty \right\}.$$

3.5.3.2 Remark $R_{MH}(\mu)$ may be defined as $R_M(\mu)$, except that we require the measures to be computable. This is obviously the more natural concept, but in this case we have *trivially* $R(\mu) \subset R_{MH}(\mu)$, since, by lemma 3.4.16, the diagonalisation argument of Theorem 3.2.3.10 goes through as well in this case. To guard oneself against a trivial solution of the problem, whether a restriction to hypothesis tests enlarges the class of random sequences, one must therefore allow the alternatives to μ to be subcomputable only. The subscript "H" refers to "hypothesis testing"; $R_{MH}(\mu)$ should be interpreted as "the analogue of $R_M(\mu)$ (definition 3.4.15) when we consider hypothesis tests only". For reasons expounded at length in section 3.4, $R_w(\mu)$ probably has no analogue in this sense.

The following lemma shows that sequences in $R_H(\mu_p)$ and $R_{MH}(\mu_p)$ have some reasonable randomness properties:

3.5.3.3 Lemma If $p \in (0,1)$ is a computable real, then $R_{MH}(\mu_p) \subseteq LLN(p)$; moreover, $R_{MH}(\mu_p)$ is invariant under recursive place selections whose domain has full measure.

Proof Consider for $q \in (-1,1)\cap\mathbb{Q}$ the Martingale V defined by

$$V_q(w) := \frac{\mu_q[w]}{\mu_p[w]}.$$

In chapter 10 of [88] Schnorr shows that $N_q := \{x \mid \forall k \exists n V_q(x(n)) > 2^k\}$ is a total recursive sequential test with respect to μ_p and that $x \in LLN(p)^c$ iff for some q, $x \in N_q$. Obviously $\mu_q \perp \mu_p$. If Φ is a recursive place selection whose domain has full measure, then $\mu_q\Phi^{-1} = \mu_q$, so $\mu_q\Phi^{-1}$ is also singular with respect to μ_p. Now apply Schnorr's result with Φx instead of x.
□

By lemma 3.4.5, we have $R(\mu) \subseteq R_H(\mu)$ and it is likely that in fact $R(\mu) \subset R_H(\mu)$. To prove equality, for each recursive sequential test N with respect to μ, one must be able to construct a computable measure $\nu \perp \mu$, such that N is contained in

$$\left\{ x \mid \limsup_{n\to\infty} \frac{\nu[x(n)]}{\mu[x(n)]} = \infty \right\}.$$

This is probably impossible; but in section 3.4 we showed that recursive sequential tests, which were introduced by Martin-Löf as significance tests, can always be represented via a likelihood ratio of measures ν and μ, *if we allow that* ν *be absolutely continuous with respect to* μ (lemma 3.4.7 and corollary 3.4.9). The meaning of the condition

$$\limsup_{n\to\infty} \frac{\nu[x(n)]}{\mu[x(n)]} = \infty$$

for absolutely continuous ν with respect to μ, is that neighbourhoods of x have probabilities under ν which are *relatively* much larger than their probabilities under μ; in an absolute sense, however, both probabilities may be small. If *this* concept of alternative hypothesis is reasonable, then so is Martin-Löf's definition of randomness (*modulo* the propensity interpretation). We leave this question open.

3.6 Conclusion Using recursion theory, Martin-Löf has provided a definition of (effective) statistical test and of randomness of great generality. How good a definition of randomness this is, depends, among else, on

– the interpretation of probability
– the interpretation of statistical tests.

We need not here repeat at length the remarks on the foundations of probability made in Chapter 2 and in the introduction to this chapter. For the sake of argument, we shall assume the propensity interpretation and the idea that randomness should be defined as satisfaction of certain statistical laws; let us see how far Martin-Löf succeeds in formalizing this idea.

As regards the interpretation of statistical tests, the very generality of Martin-Löf's definition presents a problem. There is a glaring contrast between the careful, piecemeal discussion of statistical tests in the literature (see for instance Cox [18] and Barnett [3]) and Martin-Löf's sweeping generalisation. It seems to me that there is no use in trying to establish once and for all *all* properties of random sequences if we cannot survey this totality and if there are no *general* arguments for the choice of a particular class of properties. In this case, these arguments would have to be supplied by recursion theory. Now the prospects for such general arguments look bleak: without too much effort we could devise several alternatives to the definitions proposed by Martin-Löf and Schnorr.

If these general arguments do not exist, the use of recursion theory may be rather inessential here. After the discovery of a statistical law which should be true of random sequences, we may determine its recursion theoretic structure; but this structure seems to be rather accidental. It is open to doubt whether there really exists such an intimate connection between randomness and recursion theory. Martin-Löf and Schnorr never seem to question this assumption. We saw in Chapter 2 that the only argument given in favour of such an intimate connection, the identification of admissible and lawlike place selections, is defective and that other concepts, such as entropy, seem to be more relevant. In general, hierarchies which have proved to be useful and natural in recursion theory or mathematical logic, might be unnatural or even misleading elsewhere. But if that holds true in this case, a definition of randomness should be founded on principles which are less formal and are more concerned with the content of probabilistic laws than those of Martin-Löf.

Also, if more and more concrete examples pile up, there is no guarantee that they will always fit in the straitjacket of definitions 3.2.1.2 and 3.2.1.3. Our remarks on the ergodic theorem (in 3.2.4) and on Martingales (in 3.4) provide cases in point. We don't have much sympathy either for attempts, reviewed in 3.2.4, to fix an upper bound on the arithmetical complexity of statistical tests which is so large that it is inconceivable that it will ever be attained; and even if it were attained, we might have included *too many* properties, witness the discussion on statistical tests.

We conclude that Martin-Löf's definition provides nothing in the way of a *canonical* choice of properties of randomness. We shall therefore take definitions 3.2.1.2 and 3.2.1.3 with a grain of salt and certainly not as the ultimate truth concerning randomness. If, in the sequel, we shall nonetheless use these definitions, it is because they provide a convenient formalisation of a view which is diametrically opposed to that of von Mises; and as such they will be investigated in Chapter 4.

Notes to Chapter 3

1. For an argument to the effect that the ergodic theorem is not constructively valid, see Bishop [5,233].

2. Schnorr's claim [88,37] that S can be chosen to be recursive is false; the universal recursive sequential test provides a counterexample. This is a consequence of the following **Lemma** Suppose the Σ_1 set $O \neq 2^\omega$ can be written as the union of a *recursive* set of cylinders [w]. Then there exists a recursive sequence in O^c.

Proof Let $O = \bigcup_n [w^n]$. We may assume that the recursive set $\{w^n \mid n \in \mathbb{N}\}$ is *sequential*, i.e. that every prolongation of some w^n occurs among the w^n. O^c is a non-empty Π_1 set, which is given by a recursive binary tree T. Determine a recursive subtree T' of T by throwing out all the w^n. No infinite branch of T is lost in this process, since no infinite branch of T passes through a w^n. Now every word of T' is admissible in the sense of (the proof of) theorem 3.2.2.2: for if no infinite branch of T' passes through a word v, this means that every infinite branch starting with v must belong to O; but then v must be one of the w^n. Since the set of admissible words of T' is recursive, the leftmost infinite branch of T' is recursive. □

Note that a Σ_1 set may be the union of a recursive set of cylinders without having, say, computable Lebesgue measure.

3. A sequence $(a_n)_n$ of computable reals *converges recursively* to a computable real a if there exists a total recursive function g: $\omega \rightarrow \omega$ such that for all k,n: $n \geq g(k)$ implies $|a - a_n| < 2^{-k}$. This is the usual constructive definition of convergence couched in recursion theoretic terminology.

4. Since Schnorr wants to consider only Martingales together with some function indicating growth, he must show that every total recursive sequential test is contained in a set of the form defined in 3.4.12. His Satz (9.5) [88,74] purports to establish this, but the proof contains a mistake.

5. This is so by definition if we assume von Mises' concept of probability. Otherwise, we have to add that the repetitions of the experiment are assumed to be independent.

6. We disregard subtleties having to do with randomization at the boundary to achieve the exact significance level α.