

## RANDOMNESS AND FOUNDATIONS OF PROBABILITY: VON MISES' AXIOMATISATION OF RANDOM SEQUENCES

MICHIEL VAN LAMBALGEN\*  
*University of Amsterdam*

### Abstract

We discuss von Mises' notion of a random sequence in the context of his approach to probability theory. We claim that the acceptance of Kolmogorov's rival axiomatisation was due to a different intuition about probability getting the upper hand, as illustrated by the notion of a martingale. We also discuss the connection between randomness and the axiom of choice.

*to David Blackwell*

**1. Introduction.** In 1937, the Université de Genève organized a conference on the theory of probability, part of which was devoted to foundational problems (the proceedings of this part have been published as [9]). The focal point of the discussion was von Mises' axiomatisation of probability theory [21], and especially its relation to the newly published axiomatisation by Kolmogorov. In 1919 Richard von Mises (1883-1957) had published an (in fact the first) axiomatisation of probability theory, which was based on a particular type of disorderly sequences, so called *Kollektivs*. The two features characterizing *Kollektivs* are, on the one hand, existence of limiting relative frequencies within the sequence (global regularity) and, on the other hand, invariance of these limiting relative frequencies under the operation of "admissible place selection" (local irregularity). An admissible place selection is a procedure for selecting a subsequence of a given sequence  $x$  in such a way that the decision to select a term  $x_n$  does not depend on the value of  $x_n$ .

After several years of vigorous debate, which concerned not only von Mises' attempted characterisation of a class of random phenomena, but also his views on the interpretation of probability, it became clear that most probabilists were critical of von Mises' axiomatisation and preferred the simple

---

\* AMS 1991 *subject classifications*. Primary 60A05; secondary 01A60, 03E25.

*Key words and phrases*: Probability, randomness, axiom of choice.

The author gratefully acknowledges support from the Netherlands Organisation for Scientific Research (NWO), grant PGS 22 - 262.

set of axioms given in Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung* [14] of 1933. The defeat of von Mises' theory was sealed at the conference in Geneva, where Fréchet gave a detailed account of all the objections that had been brought against von Mises' approach. While this history may now seem old hat, we contend that the discussion itself is still of interest, for the following reasons: a) the demise of von Mises' theory seems due to a different *intuition about probability* getting the upper hand, and b) the notion of a Kollektiv remains mathematically fruitful.

Of course, the usual picture is that von Mises' theory is inconsistent, too weak and in general misguided, so that the transition to Kolmogorov's axiomatisation was in fact the most rational course of events. We believe that none of the usual objections stand up to scrutiny, and that there is actually still something to be learned from von Mises' rigorous discussion of foundations. This will be illustrated below by a comparison of von Mises' views with those of Sklar's "Physics and Chance" [31], a thoughtful and scholarly treatise on the foundations of statistical mechanics.

More importantly, the years 1900-1940 represent a very interesting period in the history of probability, starting with Hilbert's injunction (in his 6<sup>th</sup> problem) to axiomatise probability and ending with the acceptance of Kolmogorov's axiomatisation. To modern eyes, Kolmogorov's axioms look very simple, and one may well wonder why it took such a long time for probability theory to mature. One reason appears to be that probability was considered to be a branch of mathematical physics (this is how Hilbert presented it), so it was not immediately apparent which part of the real world should be incorporated in the axioms. Here, von Mises and Kolmogorov chose different options. Another reason is that attempts to articulate axioms were very much guided by widely divergent intuitions about probability and the foundations of mathematics in general. This will be illustrated below using von Mises and Fréchet as protagonists. It will be seen that games played a prominent role here.

As a preliminary example of probability's movement towards articulation, let us consider Borel's paper "Les probabilités dénombrables et leurs applications arithmétiques" of 1909 (reprinted as [1]). When he introduces the considerations which lead up to the strong law of normal numbers, he states [1,194-5]

Nous nous proposons d'étudier la probabilité pour qu'une fraction décimale appartienne à un ensemble donné en supposant que

1. Les chiffres décimaux sont indépendants;
2. Chacun d'eux a une probabilité  $\frac{1}{q}$  (dans le cas de la base  $q$ ) de prendre chacune de ces valeurs possibles:  $0, 1, 2, 3, \dots, q - 1$ .

Il n'est pas besoin d'insister sur le caractère partiellement arbitraire de

ces deux hypothèses; la première, en particulier, est nécessairement inexacte, si l'on considère *comme on est toujours forcé de le faire dans la pratique*, un nombre décimal défini par *une loi*, quelle que soit d'ailleurs la nature de cette loi. Il peut néanmoins être intéressant d'étudier les conséquences de cette hypothèse, afin que précisément de se rendre compte de la mesure dans laquelle les choses se passent *comme si* cette hypothèse est vérifiée.

Borel felt that he was developing a theory *different* from measure theory to deal with probability. It is clear from the above passage that Borel considers the 'practical' continuum to consist of lawlike reals only; hence the practical continuum is countable and has measure zero with respect to any absolutely continuous measure. Borel wanted to circumvent this problem. In the introduction to [1] he explicitly states that 'dénombrable' refers to the cardinality of the sample space, and *not* to the  $\sigma$ -additivity of the measure. This was universally misunderstood by authors (such as Fréchet [7]) who view Borel as a predecessor of the measure theoretic approach to probability.

Now the consequence of Borel's two hypotheses, the strong law, was by no means considered to be self-evident; in fact one expected the opposite result. Here is Hausdorff's comment [10,420]

Dieser Satz ist merkwürdig. Auf der einen Seite erscheint er als plausible Übertragung des "Gesetzes der großen Zahlen" ins Unendliche; andererseits ist doch die Existenz eines Limes für eine Zahlenfolge, noch dazu eine vorgeschriebene Limes, ein sehr spezieller Fall, den man a priori für sehr unwahrscheinlich halten sollte.

And in 1923 Steinhaus still called the strong law of normal numbers *le paradoxe de Borel* [32,286]. Evidently, the strong law was considered to be paradoxical because a regularity such as the existence of limiting relative frequencies was felt to be incompatible with chance.

These brief indications should suffice to convince the reader that the maturation of probabilistic notions did not come overnight and that probability was so much intertwined with other concepts that a 'pure' axiomatisation was long in coming.

## 2. The frequency interpretation and the laws of large numbers.

Von Mises was an ardent advocate of the frequency interpretation of probability (cf. [24]) and took this as the basis of his axiomatisation. Since this interpretation, and especially its relation to the laws of large numbers, is often misunderstood (cf. for example Sklar's book on the foundations of statistical mechanics [31,97]; but also Feller's famous treatise [6,204]), it is worthwhile to explain it briefly here.

The fundamental primitive in von Mises' axiomatisation is the *Kollektiv*, a mathematical abstraction representing an infinite series of independent trials. Probability itself is a defined notion: the probability of an attribute in a *Kollektiv* is the limiting relative frequency of that attribute in the *Kollektiv*; in von Mises' words: "Erst das Kollektiv, dann die Wahrscheinlichkeit". This may seem rather trivial, but it is not.

1. Formally, it means that probability is not a primitive, as it is in Kolmogorov's *Grundbegriffe* [14]. This has led critics (such as Feller in his talk at the Geneva conference [4]; see also Fréchet [7]) to argue that von Mises' conception of a mathematical theory confuses empirical and mathematical considerations. Clearly however, the choice of primitive terms is free as long as the result is a rigorous system. Kolmogorov was well aware of this [14,2].

2. Von Mises' definition is not the only one which establishes some connection between probability and relative frequency. We shall use the term *strict frequentism* for any interpretation of probability which *explicitly* defines probability in terms of relative frequency. Von Mises also thinks that there is more to probability than the definition :

Die Wahrscheinlichkeit, Sechs zu zeigen, ist *eine physikalische Eigenschaft* eines Würfels, von derselben Art, wie sein Gewicht, seine Wärmedurchlässigkeit, seine elektrische Leitfähigkeit usw. [24,16]

but this aspect of probability does not figure in the definition. In particular it does not make sense to speak about probabilities of singular events, such as 'the outcome of *this* toss is heads'. This has consequences for the role of the laws of large numbers. An influential interpretation of probability superficially related to strict frequentism, the *propensity interpretation*, holds that probability should primarily be thought of as a physical characteristic. Now von Mises could concede this much but, contra von Mises, the propensity interpretation claims to be able to *derive* the frequency interpretation from the strong law of large numbers together with an auxiliary hypothesis. (Some use the weak law for this purpose; see the quotation from Fréchet [7] given in section 5.) In other words, propensity theorists claim that it is possible to derive statements on relative frequencies from premisses which are (almost) probability-free. Typically one argues as follows.

Suppose we have a coin; after a thorough examination of its physical characteristics (weight, center of mass etc.) we conclude that the probability, *as a physical characteristic or propensity*, of coming up heads will be  $p$ . The strong law of large numbers is then invoked to conclude that the set of outcome sequences which show limiting relative frequency of heads equal to  $p$  has measure one (w.r.t. the product measure determined by  $(1 - p, p)$ ). Now the auxiliary hypothesis comes in. All we need to assume is that zero probability (or zero measure) means, in the case of random events, *a probability*

*which may be neglected as if it were an impossibility* (quoted from Popper [30,380]).

Von Mises (cf. [24]) declines any use of the laws of large numbers in the way indicated above. He rightly remarks that this use amounts to an adoption of the frequency interpretation for certain special values of the probabilities, namely those near to 0 and 1 (or equal to 0 or 1 if you use the strong law), and asks: Why not adopt the frequency interpretation from the start, for *all* values of the probability distribution? The obvious answer is that the above procedure explains (or at least pretends to) the frequency interpretation. As Popper puts it:

Thus, there is no question of the frequency interpretation being *inadequate*. It has merely become *unnecessary*: we can now derive consequences concerning frequency limits even if we do not assume that probability means a frequency limit; and we thus make it possible to attach to "probability" a wider and vaguer meaning, without threatening the bridge on which we can move from probability statements on the one side to frequency statements which can be subjected to statistical tests on the other (Popper [30, 381]).

So what *is* the role of the laws of large numbers in strict frequentism? Here, von Mises [24] adopts Kolmogorov's point of view (cf. the latter's [13]) that the laws of large numbers are actually statements about fluctuations of averages in finite sequences. These laws are derivable in probability theory *because* Kollektivs are invariant under admissible place selections, not the other way around. (For a detailed discussion of these derivations, cf. van Lambalgen [16]).

3. The proposed reduction of the frequency interpretation of probability to the laws of large numbers has its exact analogue in the use of ergodic theory to explain the equality of phase averages (with respect to standard volume measure) and time averages in equilibrium statistical mechanics. Motivated by his strict frequentism, von Mises declined any such use of ergodic theory (see the last chapter of [22]), so in this context it may be worthwhile to compare the two approaches. The traditional argument runs as follows. Prove that the system at hand is ergodic (with respect to volume measure). It stands to reason that phase averages have to be computed with respect to some equilibrium measure. Assume that any reasonable equilibrium distribution is absolutely continuous with respect to volume measure. By the ergodic theorem, it then follows that the equilibrium distribution *is* volume measure. Also, the ergodic theorem gives us the equality of phase averages and time averages for a set of initial states with volume measure one. If we now assume that the anomalous trajectories do not occur, we are done. A salient feature of the argument is that the premisses do not refer to empir-

ically determined probabilities, whereas the conclusion does (in the form of time averages). Clearly there are problems with the proposed reduction; to mention but two, extensively discussed in Sklar [31]: what is the relation of the actual initial microstate and the distribution over microstates? why can we safely assume that (some) sets of volume measure zero have probability zero, or perhaps cannot even occur?

Von Mises, on the other hand, adopts a fully probabilistic approach, without reference to the underlying dynamics. He dispenses with microstates but uses coarse graining only. The data consist of two kinds of probabilities: the probability for the system to be in a certain cell, and the probability for the transition from one cell to another. He then assumes, as a plausible generalization from experience, that this set-up determines an irreducible Markov shift and proceeds to prove (a weak form of) the ergodic theorem for this situation. Note that here there is no pretense at all to reduce probabilistic behaviour to the dynamics of the system: the probabilistic data are taken from experience, not justified a priori by an appeal to ergodicity. (Von Mises also had a physical reason for this: quantum mechanics tells us that the assumption of an underlying deterministic dynamics is false.)

4. To appreciate the strictness with which von Mises himself applied his doctrine, it is instructive to consider the case of attributes of probability zero ([24,38]). If the sample space is uncountable, then probability zero cannot mean impossibility; but in the case of a finite sample space probability zero is generally equated to impossibility. Not so for von Mises. When our Kollektiv is infinite, as the precise version of the explicit definition of probability requires, then probability zero of an attribute is compatible with the attribute occurring infinitely often.

**3. Axiomatising Kollektivs.** Von Mises' formal set-up is as follows ([21,57]). Let  $M$  (for "Merkmalraum") be a sample space, i.e. the set of possible outcomes of some experiment. The doctrine of strict frequentism says that probabilities  $P(A)$  for  $A \subseteq M$  must be interpreted as the relative frequency of  $A$  in some Kollektiv. In our mathematical description the probability  $P(A)$  will be identified with the limiting relative frequency of the occurrence of  $A$  in some infinite Kollektiv  $x \in M^\omega$ .

**Definition 3.1.** A sequence  $x \in M^\omega$  is called a Kollektiv if

- (i) For all  $A \subseteq M$ ,  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_A(x_k)$  exists; call this limit  $P(A)$ .
- (ii) Let  $A, B \subseteq M$  be non-empty and disjoint; and suppose that  $A \cup B$  occurs infinitely often in  $x$ . Derive from  $x$  a new sequence  $x'$ , also in  $M^\omega$ , by deleting all terms  $x_n$  which do not belong to either  $A$  or  $B$ . Now let  $\Phi$  be an admissible place selection, i.e. a selection of a subsequence  $\Phi x'$  from  $x'$  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.”*

Then

$$P'(A) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_A((\Phi x')_k)$$

and

$$P'(B) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_B((\Phi x')_k)$$

exist and  $P'(A)P(B) = P'(B)P(A)$ .

A few remarks on the above definition are in order.

1. The quantifier “for all  $A \subseteq M$ ” should not be taken too seriously. In the *Wahrscheinlichkeitsrechnung* [22,17] von Mises remarks that all one needs to assume is that (i) and (ii) hold for “simply definable” sets. For definiteness, we may substitute “Peano-Jordan measurable” for “simply definable”. Note that we cannot take  $A$  to range over Borel sets; for one thing because  $\{x_n \mid n \in \omega\}$  is countable, hence (in reasonable spaces) Borel. We shall come back to this point in section 5.
2. Of course the enigmatic condition (ii) will take pride of place among our considerations. In the relevant literature the first part (replacing  $x$  by  $x'$ , obtained from  $x$  by deleting terms not in  $A \cup B$ ) is usually omitted. For the paradigmatic case of coin tossing, the sample space  $M$  equals  $2 = \{0, 1\}$  and condition (ii) reduces to: if  $\Phi$  is an admissible place selection,  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} (\Phi x')_k = P(\{1\})$ .

The more elaborate condition is necessary in order to ensure the validity of the rule for conditional probabilities:  $P(A \mid B)P(B) = P(A \cup B)$ . It is interesting that the validity of this rule has to be built in blatantly into the axioms, thus emphasizing its empirical origin. (Wald [35, 41-2] claims that, also in the general case, condition (ii) can be reduced as for Kollektivs in  $2^\omega$ ; but his proof uses evidently non-admissible place selections.)

**3.1. Games and place selections.** Admissible place selections may be viewed as gambling strategies with *fixed* stakes: if  $n$  is chosen, that means that a bet is placed on the outcome of the  $n^{\text{th}}$  trial; otherwise, the  $n^{\text{th}}$  trial is skipped. Von Mises called condition (ii) the *Regellosigkeitsaxiom* or the *Prinzip vom ausgeschlossenen Spielsystem*. Apparently, von Mises thinks that a gambling strategy gaining unlimited amounts of money can operate only by selecting a subsequence of trials in which relative frequencies are

different. It was shown by Ville [34] that this idea is mistaken: there exist gambling strategies, namely martingales, which cannot always be represented as place selections. We shall come back to this point in section 5.

In the examples below we consider the simplest kind of game, cointossing; in other words, Kollektivs  $x$  in  $2^\omega$ .

- (a) Choose  $n$  if  $n$  is prime. (This strategy caused Doob to remark that its only advantage consists in having increasing leisure to think about probability theory in between bets.)
- (b) Choose  $n$  if the  $n - 9^{\text{th}}, \dots, n - 1^{\text{th}}$  terms of  $x$  are all equal to 1. (The strategy of a gambler who believes in “maturity of chances”.)
- (c) Now take a second coin, supposed to be independent of the first is so far as that is possible (no strings connecting the two coins, no magnetisation etc.). Choose  $n$  if the outcome of the  $n^{\text{th}}$  toss with the second coin is 1.

If one thinks of successive tosses of the coin as being independent, then condition (ii) is intuitively satisfied in all three cases, although in (c) a heavy burden is put upon the double meaning of the word “independent”. We shall call selections of type (a) and (b) *lawlike* (since they are given by some prescription) and those of type (c) *random*. Note that the treatment of independence in von Mises’ theory differs from the standard one: he tries to capture the independence of successive tosses directly, without invoking the product rule.

At this point a natural question may arise: why do we need Kollektivs at all? Why isn’t it sufficient to use the distribution (as in effect happens in Kolmogorov’s theory) instead of the unwieldy formalism of Kollektivs? The answer is that *Kollektivs are a necessary consequence of the frequency interpretation*, in the sense that if one interprets probability as limiting relative frequency, then infinite series of outcomes will exhibit Kollektiv-like properties. Therefore, if one wants to axiomatise the frequency interpretation, these properties have to be built in. That infinite series of outcomes satisfy Kollektiv-like properties is *not* because of the laws of large numbers, but for the following reason.

Consider first the case where we toss two coins, supposedly independent, simultaneously. The two coins are represented by Kollektivs  $x$  and  $y$ . We expect that the distribution in  $(\langle x_n, y_n \rangle)_n$  is given by the product rule. Now the limiting relative frequency of observing  $\langle 1, 1 \rangle$  equals the limiting relative frequency of 1 in  $y$  times the limiting relative frequency of 1 in the subsequence of  $x$  determined by  $y_n = 1$ . Hence we obtain the expected answer if the place selection in example (c) is admissible.

A slightly more complicated example is the following. We want to know the probability that successive tosses of a coin yield 1. This is calculated as follows: take a Kollektiv  $x$  corresponding to the coin; by admissible place selection (of type (a)) form Kollektivs  $(x_{2n-1})_n$  and  $(x_{2n})_n$ ; we may now apply the preceding argument if we can prove that the Kollektivs  $(x_{2n-1})_n$  and  $(x_{2n})_n$  are independent, i.e. that  $(\langle x_{2n-1}, x_{2n} \rangle)_n$  is a Kollektiv with respect to the product distribution. In order to show this, we have to use the fact that the place selection determined by the prescription 'choose  $x_k$  if  $k$  is even and  $x_{k-1} = 1$ ' is admissible. This is a place selection of type (b).

In this way a Kollektiv  $x \in 2^\omega$  determines a distribution on the set of finite binary words. Kollektivs invariant under place selections of type (a) and (b) so that they determine a distribution on the set of finite binary words have been studied under the name of 'Nachwirkungsfreie Folgen' (Popper [30]) or 'Bernoulli sequences'. Of course, if the probability equals  $1/2$ , these are just normal numbers.

**4. Inconsistency?** Von Mises himself was aware that Kollektivs cannot be explicitly constructed, so that the consistency of the theory can be established only indirectly. Indeed (arguing informally), suppose that  $x$  is a Kollektiv given by an explicit function  $n \rightarrow x_n$ . Then this function can be used to define an admissible place selection selecting a subsequence of  $x$  in which the distribution is different from that in  $x$ . Von Mises comments

daß man die "Existenz" von Kollektivs nicht durch eine analytische Konstruktion nachweisen kann, so wie man etwa die Existenz stetiger, nirgends differentierbarer Funktionen nachweist. Wir müssen uns mit der abstrakten logischen Existenz begnügen, die allein darin liegt, daß sich mit den definierten Begriffe widerspruchsfrei operieren läßt [21,60].

In other words, Kollektivs are *new* mathematical objects, not constructible from previously defined objects. Hence in one place [22,15] (see also [24,112]) von Mises compares Kollektivs to Brouwer's free choice sequences [2], one extreme example of which is the sequence of outcomes produced by successive casts of a die. In another place he contrasts his approach with that of Borel, in a way which makes clear that Kollektivs are not to be thought of as numbers, i.e. *known* objects:

... die von Borel u.a. untersuchten Fragen (z.B. über das Auftreten einzelner Ziffern in den unendlichen Dezimalbrüchen der irrationalen Zahlen), wo das Erfülltsein oder Nicht-Erfülltsein der Forderung II [i.e. 3.1.(ii)] ohne Bedeutung ist [21,65].

Von Mises' argument that Kollektivs cannot be explicitly constructed, was often turned against him, as an argument showing that Kollektivs do

not exist. Here is Kamke's version, in a report to the Deutsche Mathematiker Verein [12,23]: suppose that  $x \in 2^\omega$  is a Kollektiv which induces a distribution  $P$  with  $0 < P(\{1\}) < 1$ . Consider the set of strictly increasing sequences of (positive) integers. This set can be formed independently of  $x$ ; but among its elements we find the strictly increasing infinite sequence  $\{n \mid x_n = 1\}$ , and this sequence defines an admissible place selection which selects the subsequence 11111... from  $x$ . Hence  $x$  is not a Kollektiv after all. (Fréchet reiterated this objection in his [7], while referring to von Mises' stubbornness in refusing to accept it.)

Clearly, this argument is very insensitive to von Mises' intentions, and he had no trouble dismissing it: the set  $\{n \mid x_n = 1\}$  does not define an admissible place selection since it uses *Merkmalunterschiede* in a most extreme manner. The real problem is, rather, to understand why the argument was considered to be convincing at all. The first obvious reason is that von Mises could not come up with an unassailable consistency proof. A second, less obvious, reason may be that von Mises and his adversaries had very different views on the foundations of mathematics; as his reference to Brouwer shows, he was willing to admit objects into mathematics about which we have incomplete information only, whereas for instance Kamke (the author of a textbook on set theory!) stood firmly in the classical, Platonist, tradition. Here we shall not deal with attempts to make von Mises' definition precise in terms of classical concepts, except to note the following points.

Efforts were first directed toward defining a class  $\mathcal{H}$  of place selections for which it could be shown that the set of Kollektivs invariant under that class is non-empty. To this end (and considering the simplest case) place selections were conceived of as functions  $\Phi : 2^\omega \rightarrow 2^\omega$ , generated by some  $\tau : 2^{<\omega} \rightarrow \{0, 1\}$  (here,  $2^{<\omega}$  is the set of finite binary sequences) in the obvious way: interpret ' $\tau(x(n)) = 1$ ' as 'select  $x_{n+1}$ '. It can be shown that (the inverses of)  $\Phi$ 's thus defined preserve null sets (with respect to any product measure), hence as long as  $\mathcal{H}$  is countable there will always be Kollektivs invariant with respect to  $\mathcal{H}$ . Two choices of  $\mathcal{H}$  imposed themselves: take  $\mathcal{H}$  to be the set of place selections determined by recursive  $\phi$  (Church [3]), or use the so-called Bernoulli selections, which are determined by  $\phi_w$  satisfying  $\phi_w(u) = 1$  iff  $w$  is a final segment of  $u$ . The Bernoulli sequences are precisely the sequences invariant under Bernoulli selections.

Second, once one has a set of Kollektivs invariant under some  $\mathcal{H}$ , one can ask whether it is perhaps also invariant under place selections not in  $\mathcal{H}$ . A very interesting example in this area is Kamae's work on place selection by means of entropy zero sequences (Kamae [11]). Entropy zero sequences are deterministic in the following weak sense: as  $n$  grows, given  $x(n)$ , we can predict ever longer segments  $x(n+m)$ . Such deterministic sequences can for example be obtained from irrational rotations of the circle (so-called

Sturmian trajectories). Kamae showed that if  $x$  is a Bernoulli sequence and  $y$  is deterministic, then the subsequence of  $x$  determined by: choose  $x_n$  if  $y_n = 1$ , is again a Bernoulli sequence.

Unfortunately we have to forego any discussion of Kolmogorov complexity, where random sequences are defined as those sequences which do not have 'easily describable' regularities. We refer the reader to Li and Vitanyi [20] for historical and technical details.

In the remaining few paragraphs of this section, we sketch an approach to randomness which tries to stay closer to von Mises' in that it is axiomatic, introduces objects about which we have incomplete information, and takes the notion of admissibility as a starting point.

If one carefully looks at von Mises' *Regellosigkeitsaxiom*, what seems to be fundamental is the notion of *independence*: whether a trial should be included in the subsequence should be independent of the outcome of that trial. In section 3 it was shown that there exists a close connection between independence of successive outcomes on the one hand, and independence of two sequences of outcomes on the other. It therefore seems promising to take independence between sequences of outcomes as a primitive of the axiomatisation. This choice can also be motivated in a different way. We have seen that recursive sequences or entropy zero sequences can define admissible place selections. What seems essential here is that these sequences have low information content, so that *a fortiori* they have little information about the Kollektiv. This suggests taking as primitive the relation " $y$  has no information about  $x$ ". The formal properties of this relation are much the same as those of independence (cf. van Lambalgen [17]); in fact they satisfy most axioms of matroid theory, i.e. the theory of linear or algebraic independence. For example, for all proposed definitions it can be shown (with some effort) that they satisfy the Steinitz exchange property: if  $x$  is independent of  $\{y, z_1, \dots, z_n\}$  and  $y$  is independent of  $\{z_1, \dots, z_n\}$ , then  $y$  is independent of  $\{x, z_1, \dots, z_n\}$ . Hence we axiomatise randomness by means of matroid theory and some additional properties. For reasons of space, here we shall give an informal description only; for details we refer the reader to van Lambalgen [18], [19].

First some notation. If  $z$  is an infinite binary sequence, let  $[z]$  denote the set  $\{x \mid x \text{ differs from } z \text{ on at most a finite initial segment}\}$ . Let  $z^0$  denote the sequence of even coordinates of  $z$ , and let  $z^1$  denote the sequence of odd coordinates of  $z$ . Using this notation, we may now extrapolate some of von Mises' ideas on independence as follows:

- (a) if  $z$  is random sequence, so are  $z^0$  and  $z^1$
- (b) since we think of the successive choices (or coin tosses) generating  $z$  as independent, it follows that  $z^0$  and  $z^1$  are independent of each other

- (c) conversely, if  $x$  and  $y$  are independent randomly generated sequences, and if  $z$  is such that  $z^0 = x$  and  $z^1 = y$ , then  $z$  is itself random.

‘Independent’ is used as an intuitive concept here, but (a-c) have some simple consequences which can be stated without using the concept of independence. For example, suppose  $z$  is randomly generated; if  $x$  is such that  $x^0 = z^1$ ,  $x^1 = z^0$ , then by (c),  $x$  is also random.

The second ingredient of the axiomatisation refers back to Brouwer’s theory of free choice sequences [2]. Fundamental to this theory is the following observation (in this form due to Troelstra [33]): if we randomly or freely generate an infinite sequence  $\alpha$ , then at any stage we know only a finite initial segment of that sequence. This principle is called the Axiom of Open Data. Formulated in this manner, Open Data contradicts classical logic, but it can be given a form in which it does not, if we restrict the class of properties to which it applies. Suppose  $A$  is a property which is insensitive to initial segments in the following sense: if  $A(\beta)$  is true and  $\gamma$  differs from  $\beta$  only in a finite initial segment, then  $A(\gamma)$  is also true. Call such a property  $A$  *asymptotic* (clearly  $A$  defines a tailset). Then we may paraphrase Open Data as follows: if  $A$  is an asymptotic property and  $A$  holds for *some* randomly generated  $\alpha$ , then  $A$  holds for *all* randomly generated  $\alpha$ . (If  $A$  were false for some randomly generated  $\beta$ , then  $A$  would make distinctions among randomly generated sequences; but the only way to make distinctions is on the basis of finite initial segments, which  $A$  cannot do because it is asymptotic.) Clearly, this reformulation of Open Data is an abstract version of the 0-1 law.

It is shown in van Lambalgen [19] (using forcing) that the axioms for independence together with the 0-1 law are consistent with Zermelo-Fraenkel set theory (plus the axiom of dependent choices), thus showing that von Mises’ intuition did not deceive him. However, somewhat surprisingly they do contradict the axiom of choice.

Intuitively, one may argue as follows. Let  $\mathcal{C}$  be the collection of pairs  $\{\{[z^0], [z^1]\} \mid z \text{ randomly chosen}\}$ . We show that it is impossible to choose an element from each pair of  $\mathcal{C}$ . Working toward a contradiction, suppose that  $g$  is a function which picks one element from each pair in  $\mathcal{C}$ . For definiteness, let us suppose that there is a random  $z$  such that  $g\{[z^0], [z^1]\} = [z^0]$ . Now the property “ $g\{[z^0], [z^1]\} = [z^0]$ ” is asymptotic in  $z$ , so by the 0-1 law it follows that for all randomly chosen  $x$ ,  $g\{[x^0], [x^1]\} = [x^0]$ . Consider a very special  $x$ , namely a sequence which is defined by  $x_{2n} = z_{2n+1}$ ,  $x_{2n+1} = z_{2n}$ . The sequence  $x$  is just as random as  $z$  (by (c) above), and we have  $[z^0] = [x^1]$  and  $[z^1] = [x^0]$ , hence  $g\{[z^0], [z^1]\} = g\{[x^0], [x^1]\}$  also equals  $[z^1]$ , a contradiction. (The formal version of the argument takes account of the fact that  $z^0$  and  $z^1$  should be *independent given g*.)

Philosophically this seems interesting, because it shows that a funda-

mental probabilistic notion, at least when taken to the limit, fits somewhat uneasily in the standard mathematical framework.

**5. The Geneva conference: Fréchet's objections.** As mentioned in section 1, during the Geneva conference on probability the prevailing attitude towards von Mises' ideas was critical. A fairly complete list of objections was drawn up in Fréchet's survey lecture on the foundations of probability [9,23-55]. Von Mises himself was absent, but his rebuttals of the objections were published in the proceedings [25]. To no avail: the same objections were reiterated in Fréchet's [8]; and, for that matter, ever since. Fréchet's criticism has more or less become the standard wisdom on the subject and for this reason we shall present it in some detail. Our conclusion will be that most of the objections, those based on Ville's famous construction included, are unfounded.

**5.1. Fréchet's philosophical position.** As stated in section 1, we shall adopt as working hypothesis that the lack of mutual comprehension between von Mises and his critics is due to widely differing views on the foundations of mathematics as well as on the foundations of probability. In particular, we shall assume that Fréchet is an adherent of the propensity interpretation. This hypothesis will explain at least in part why Fréchet thought that Ville's theorem dealt such a devastating blow to von Mises' program. We compile some passages from Fréchet [7,45-7] to show that he indeed subscribes to the propensity interpretation.

[...] "la probabilité d'un phénomène est une propriété de ce phénomène qui se manifeste à travers sa fréquence et que nous mesurons au moyen de cette fréquence".

Voici donc comment nous voyons répartis les différents rôles dans la théorie des probabilités. Après avoir constaté comme un fait pratique, que la fréquence d'un événement fortuit dans un grand nombre d'épreuves se comporte comme la mesure d'une constante physique attachée à cet événement dans une certaine catégorie d'épreuves, constante qu'on peut appeler probabilité on en déduit, par des raisonnements dont la rigueur n'est pas absolue, les lois des probabilités totales et composées et on vérifie pratiquement ces lois. La possibilité de cette vérification enlève toute importance au peu de rigueur des raisonnements qui ont permis d'induire ces lois. Ici s'arrête la synthèse inductive.

On fait correspondre maintenant à ces réalités (toutes entachées d'erreurs expérimentales), un modèle abstrait, celui qui est décrit dans l'ensemble des axiomes, lesquels ne donnent pas - contrairement à ceux de M. de Mises - une définition constructive de la probabilité, mais une définition descriptive. [...]

Sur l'ensemble d'axiomes est bâtie la théorie déductive ou mathématique des probabilités. Enfin la valeur du choix de cet ensemble est soumise au contrôle des faits, non par la vérification directe, mais par celle des conséquences qui en ont été déduites dans la théorie déductive. La vérification la plus immédiate se présentera en général de la façon suivante: on adopte comme mesures expérimentales de certaines probabilités  $p, p', \dots$  les fréquences  $f, f', \dots$ , correspondantes dans les groupes d'épreuves nombreuses. Certains théorèmes de la théorie déductive établissent les expressions de certaines autres probabilités,  $P, P', \dots$ , en fonction de  $p, p', \dots$ . Ayant calculé  $P, P', \dots$  au moyen de ces expressions où l'on a remplacé approximativement  $p, p', \dots$  par  $f, f', \dots$  la vérification consistera à s'assurer que les valeurs approchées ainsi obtenus pour  $P, P', \dots$  sont aussi approchées des fréquences  $F, F', \dots$  qui sont les mesures expérimentales directes de  $P, P', \dots$ .

On peut d'ailleurs réduire beaucoup les difficultés pratiques de ces vérifications. Si l'on appelle  $P_n$  la probabilité pour que la fréquence dans  $n$  épreuves d'un événement de probabilité  $p$ , diffère de  $p$  de plus de  $\varepsilon$ , alors d'après le théorème de Bernoulli,  $P_n$  converge vers zéro avec  $\frac{1}{n}$ . Si donc on se contente de vérifier expérimentalement qu'un événement de probabilité assez petite est pratiquement très rare et même qu'un événement de probabilité extrêmement petite est pratiquement impossible, le théorème de Bernoulli se traduit pratiquement ainsi: quel que soit le nombre  $\varepsilon > 0$ , la fréquence dans  $n$  épreuves pourra pratiquement être considérée comme différant de la probabilité correspondante, de moins de  $\varepsilon$ , si le nombre des expériences est assez grand. Autrement dit, il est inutile d'opérer, pour toutes les valeurs de la probabilité  $p$ , la vérification qu'on se proposait. On peut se contenter de la faire quand  $p$  est petit. Or cela est beaucoup plus facile; il n'est pas nécessaire de faire de longs relevés.

Except for the use of the weak law of large numbers where Popper uses the strong law, Fréchet's version of the propensity interpretation follows the lines laid out in 2 (although Fréchet seems to be much less aware of his assumptions than e.g. Popper!). It is evident from [7] and [8] that Fréchet considers the propensity interpretation to be much simpler than the strict frequency interpretation. Superficially, this is indeed so: much of what von Mises struggled to formulate precisely is relegated here to the "synthèse inductive", where "c'est l'intuition qui domine et cherche à dégager comme elle peut, l'essentiel de la complexité des choses" [7,45]. In particular, as we have seen, the rules of probability do not have to be rigorously derived from the interpretation, in contrast with von Mises' approach. Similarly, Fréchet can do without *limiting* relative frequencies and Kollektivs. We need not reiterate our views here.

**5.2. Weakness of Kollektivs: Ville's construction.** Fréchet's interpretation of probability lies at the root of what he considers to be the most forceful objection against von Mises. To understand this objection, we have to state the law of the iterated logarithm.

**Law of the iterated logarithm (LIL)** Let  $p \in (0, 1)$ . 'Almost all' refers to the product measure  $(1 - p, p)^\omega$ .

- (a) For  $\alpha > 1$ , for a.a.  $x \in 2^\omega \exists k \forall n \geq k | \sum_{j \leq n} x_j - np | < \alpha [2p(1-p)n \log \log n]^{\frac{1}{2}}$
- (b) For  $\alpha < 1$ , for a.a.  $x \in 2^\omega \forall k \exists n \geq k [ \sum_{j \leq n} x_j - np > \alpha [2p(1-p)n \log \log n]^{\frac{1}{2}}$   
and  
for a.a.  $x \in 2^\omega \forall k \exists n \geq k [ np - \sum_{j \leq n} x_j > \alpha [2p(1-p)n \log \log n]^{\frac{1}{2}} ]$ .

Part (b) in particular shows that the quantities  $\sum_{j \leq n} x_j - np$  and  $np - \sum_{j \leq n} x_j$  exhibit fairly large oscillations. This observation provides the starting point for Ville's construction [34,55-69], which proceeds in two stages (actually, our presentation is slightly anachronistic, since Ville uses Lévy's Law, a precursor of the law of the iterated logarithm, instead of the latter).

1. Given any countable set  $\mathcal{H}$  of place selections  $\Phi : 2^\omega \rightarrow 2^\omega$ , Ville is able to construct a sequence  $x \in 2^\omega$  with the following properties:

- (i)  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} x_k = \frac{1}{2}$  and  $x$  is invariant under place selections from  $\mathcal{H}$
- (ii) for all except finitely many  $n$ ,  $\sum_{k \leq n} x_k \geq \frac{n}{2}$ .

Part (ii) means that the relative frequency of 1 approaches its limit from above, a property which is atypical in view of the law of the iterated logarithm. Ville's construction is algebraic, but it can be given a measure theoretic form as follows (cf. van Lambalgen [15]). Let  $\mu$  be a product measure of the form  $\prod_n (1 - p_n, p_n)$  such that  $p_n$  converges to  $\frac{1}{2}$ . Then for any such  $\mu$ ,  $\mu\{x \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} x_k = \frac{1}{2} \text{ and } x \text{ is invariant under place selections from } H\} = 1$ , but for  $(p_n)$  converging sufficiently slowly,  $\mu\{x \mid x \text{ satisfies LIL}\} = 0$ .

2. In the second stage of the construction, Ville temporarily adopts von Mises' viewpoint and interprets probability measures on  $2^\omega$  as in effect being induced by Kollektivs  $\xi \in (2^\omega)^\omega$ . Hence if  $\lambda$  is Lebesgue measure (i.e. the product measure generated by the uniform distribution on  $\{0, 1\}$ ) on  $2^\omega$ ,  $\lambda A = 1$  must mean

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_A(\xi_k) = 1.$$

So far we have considered only Kollektivs in  $2^\omega$ ; in particular, we have not defined what place selections  $\Psi : (2^\omega)^\omega \rightarrow (2^\omega)^\omega$  are. Fortunately, we need

not do so here, since we may, for the sake of argument, assume that Ville has done so in a satisfactory manner (for those interested in the details, see [34,63-67]). Then Ville shows the following, using 1. :

For any countable set  $\mathcal{H}$  of place selections  $\Psi : (2^\omega)^\omega \rightarrow (2^\omega)^\omega$ , there exists  $\xi \in (2^\omega)^\omega$  such that

(iii)  $\xi$  induces  $\lambda$  and is invariant under place selections from  $\mathcal{H}$

(iv)  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_{LIL}(\xi_k) = 0$ .

**Remark.** The reader may well wonder what “induces” in (iii) means in view of (iv), since we defined “ $\xi$  induces  $P$ ” to mean:

$$\text{for 'all' } B \subseteq 2^\omega, P(B) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_B(\xi_k)$$

but since  $P(A) = 0$  (by (iv)), the induced measure  $P$  cannot be equal to  $\lambda$  as claimed by (iii). Therefore (iii) should be understood as follows. A  $\sigma$ -additive measure on  $2^\omega$  is determined completely by its values on the cylinders  $[w]$ , for finite binary words  $w$ ; and we do have for the  $\xi$  constructed by Ville:

$$\lambda[w] = 2^{-|w|} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_{[w]}(\xi_k).$$

It can then be shown that also  $\lambda A = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} 1_A(\xi_k)$  for Peano-Jordan measurable  $A$ , i.e. for  $A$  such that the boundary of  $A$  is a nullset. Clearly, LIL, being first category, is not Peano-Jordan measurable. Ville’s construction is thus a very interesting case of the phenomenon that limiting relative frequency is not a  $\sigma$ -additive measure; since if the induced  $P$  were  $\sigma$ -additive, it would coincide with  $\lambda$ .

Again, this construction can be given a measure theoretic form: choose a product measure  $\mu$  whose marginals converge sufficiently slowly to  $\frac{1}{2}$ , put  $\mu_n = \mu T^{-n}$  (where  $T$  is the leftshift) and let  $\nu$  be the product measure  $\mu_1 \times \mu_2 \times \mu_3 \times \dots$ . Then for  $\nu$ -a.a.  $\xi \in (2^\omega)^\omega$  (iii) and (iv) hold.

From 1. and 2., Fréchet and Ville derived the following three objections to von Mises’ theory.

- (a) (From 2.) The theory of von Mises is weaker than that of Kolmogorov, since it does not allow the derivation of the law of the iterated logarithm.
- (b) (From 1.) Kollektivs do not necessarily satisfy all asymptotic properties proved by measure theoretic methods and since the type of behaviour exemplified by (ii) will not occur in practice (when tossing a fair coin), Kollektivs are not satisfactory models of random phenomena.

- (c) (From 1.) Von Mises' formalisation of gambling strategies as place selections is defective, since one may devise a strategy (a martingale) which makes unlimited amounts of money of a sequence of the type constructed in 1., whereas ipso facto (by (i)), there is no place selection which does this.

Von Mises reacted to these objections with a cavalier dismissal: "J'accepte ce théorème mais je n'y vois pas une objection" [25,66]. In fact, von Mises to some extent anticipated Ville's construction in his discussion of the meaning of probability zero [24,38]. As we have seen, von Mises thought that an event having zero probability might occur infinitely often in a Kollektiv. But in this case, the limiting relative frequency is necessarily approached unilaterally, as for the sequence constructed by Ville. We now discuss objections (a), (b) and (c).

Objection (a) is easiest to dispose of; in fact we have done so already in section 2, when we discussed the meaning of the strong limit laws in von Mises' theory. Stage 2 of Ville's construction shows that, although the version of the law of the iterated logarithm for finite sequences is derivable in von Mises' theory (which implies that it can be interpreted via relative frequency), the version for infinite sequences is not so derivable. This means that the theorem does not have a frequency interpretation (in the space of *infinite* binary sequences). To be more precise: von Mises distinguishes between measure theoretic and probabilistic derivations. LIL, as a statement about infinite sequences of trials, is not (probabilistically) derivable using operations such as place selections, although it is of course measure theoretically derivable using properties of the infinite product measure (essentially the Borel-Cantelli lemmas). Von Mises' rules are set up so that they preserve the frequency interpretation; this no longer holds for the limiting operations of measure theory.

Far from being a drawback of the theory, this seems to be a very interesting subtlety, which illuminates the status of the law of the iterated logarithm and which nicely illustrates Kolmogorov's note of caution when introducing  $\sigma$ -additivity:

Wenn man die Mengen (Ereignisse)  $A$  aus  $\mathcal{E}$  [which in this case is the algebra generated by the cylinders  $[w]$ ] als reelle und (vielleicht nur annäherungsweise) beobachtbare Ereignisse deuten kann, so folgt daraus natürlich nicht, daß die Mengen des erweiterten Körpers  $B(\mathcal{E})$  [the  $\sigma$ -algebra generated by  $\mathcal{E}$ ] eine solche Deutung als reelle beobachtbare Erscheinungen vernünftiger Weise gestatten. Es kann also vorkommen, daß das Wahrscheinlichkeitsfeld  $(\mathcal{E}, P)$  als ein (vielleicht idealisiertes) Bild reeller zufälliger Erscheinungen betrachtet werden kann, während das erweiterte Wahrscheinlichkeitsfeld  $(B(\mathcal{E}), P)$  eine reine mathematische Konstruktion ist [14,16].

Objection (b) consists of two parts:

- (b<sub>1</sub>) Kollektivs are not satisfactory models of random phenomena, since a unilateral approach of the limit will not occur in practice;
- (b<sub>2</sub>) Kollektivs apparently do not necessarily satisfy all asymptotic laws derived by measure theoretic methods; it is an arbitrary decision to demand the satisfaction of one asymptotic law, viz. the strong law of large numbers at the expense of another, the law of the iterated logarithm.

These objections make sense only from the point of view of the propensity interpretation. "In practice" we see only finite sequences. Kollektivs were so designed as to be able to account for all statistical properties of finite sequences and they do so perfectly. To that end, a certain amount of idealisation, in particular the consideration of infinite sequences turned out to be convenient. But the consideration of infinite sequences was not an end in itself and von Mises certainly had no intention to model infinite random "phenomena".

The only criterion for accepting or rejecting properties of infinite Kollektivs was their use in solving the finitary problems of probability theory and for that purpose, assuming invariance under place selections suffices. Objection (b<sub>2</sub>) claims that in fact there *does* exist another criterion: satisfaction of asymptotic laws derived by measure theoretic methods. But we have seen that, according to von Mises, limiting relative frequencies in Kollektivs do not owe their existence to the law of large numbers. Neither are they invariant under admissible place selections *because* place selections are measure preserving transformations. Similarly, the fact that the law of the iterated logarithm has been derived (for infinite sequences) does not in itself entail that Kollektivs should satisfy it.

On the propensity interpretation, objection (b<sub>2</sub>) of course makes sense. Formally, here one may view the relation between probabilistic statements and experience as follows: if a strong law of large numbers has been derived, a typical outcome sequence should satisfy it. Clearly, however, to turn this into a definition of typicality requires a non-arbitrary choice of a set of strong laws, a difficult task.

Lastly, we come to objection (c): von Mises' formalisation of gambling strategies (as place selections) is not the most general possible, since one can construct a strategy (a martingale) which may win unlimited amounts of money on the type of sequence constructed in 1.. For the present discussion, a martingale is given by a function  $V : 2^{<w} \rightarrow R^+$ , where  $V(w)$  denotes the capital which the gambler, having played according to the strategy, possesses after  $w$  has occurred, and such that  $V(w)$  equals the expected capital after move  $|w| + 1$ . Ville exhibits a martingale  $V$  such that for the sequence  $x$

constructed in 1.,  $\limsup_{n \rightarrow \infty} V(x(n)) = \infty$ ; but, since  $x$  is a Kollektiv, no gambling strategy in the sense of von Mises can win unlimited amounts of money on  $x$ . The interesting point about this objection is not that it undermines von Mises' approach; after all, it was not his purpose to formalise the concept of an infinite sequence for which no successful gambling strategy exists. What is really interesting is the following. 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 [25,67].

We believe that von Mises was actually forced to not understand martingales by his interpretation of probability. Martingales were introduced to capture the notion of fairness of a game: a game is fair if, for each  $n$ , the expected capital after the  $n + 1^{\text{th}}$  trial is equal to the capital after the  $n^{\text{th}}$  trial. But taking expectations requires some probability measure; and which probability measure should one consider? The intuitive idea behind fairness seems to be that it makes sense to speak of "probability of heads at the  $n^{\text{th}}$  toss". This notion of fairness is clear on the propensity interpretation. Adopting the standpoint of strict frequentism, one might be inclined to say that the pay-offs for a game *on*  $x$  should be determined by the limiting relative frequencies *in*  $x$ . Ville's example shows that if gambling house adopts this pay-off policy, it may lose money. Although this policy is alright for games with fixed stakes (i.e. place selections), it is not applicable to games with variable stakes. Hence the gambling house must have knowledge of the probabilities of individual coordinates.

But, as we have seen, 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  $\xi_k$  is of this type, then the probability of 1 at the  $n^{\text{th}}$  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.

## REFERENCES

- [1] E. BOREL, Les probabilités dénombrables et leurs applications arithmétiques, reprinted as Note V in: E. Borel, *Leçons sur la théorie des fonctions*, Gauthiers-Villars (1914), 182-216.
- [2] L.E.J. BROUWER, Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten I: Allgemeine Mengenlehre, *Ned. Acad. Wetensch. Verh. Tweede Afd. Nat.* (1918) 12/5.
- [3] A. CHURCH, On the concept of a random sequence, *Bull. AMS* **46** (1940), 130-135.
- [4] W. FELLER, Sur les axiomatiques du calcul des probabilités et leurs relations avec les expériences, in [9], 7-21.
- [5] W. FELLER, Über die Existenz sogenannter Kollektive, *Fund. Math.* **32** (1939), 87-96.
- [6] W. FELLER, *Introduction to probability theory and its applications*, Volume 1, 3<sup>rd</sup> ed. Wiley (1968).
- [7] M. FRÉCHET, Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités, in [9], 22-55.
- [8] M. FRÉCHET, The diverse definitions of probability, lecture at the fourth International Congress for the Unity of Science, *Erkenntnis* (1938).
- [9] *Colloque consacré au calcul des probabilités*, Proceedings of a conference held at the Université de Genève in 1937. The papers concerning the foundations of probability were published in the series *Actualités Scientifiques et Industrielles*, **735**, Hermann (1938).
- [10] F. HAUSDORFF, *Grundzüge der Mengenlehre*, von Veit (1914).
- [11] T. KAMAE, Subsequences of normal sequences, *Isr. J. Math.* **16** (1973), 121-149. See also: T. Kamae, B. Weiss, Normal numbers and selection rules, *Isr. J. Math.* **21** (1975), 101-110.
- [12] E. KAMKE, Über neuere Begründungen der Wahrscheinlichkeitsrechnung, *Jahresber. DMV* **42** (1932), 14-27.
- [13] A.N. KOLMOGOROV, Das Gesetz des iterierten Logarithmus, *Math. Ann.* **101** (1929), 126-135.
- [14] A.N. KOLMOGOROV, *Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse der Mathematik und ihrer Grenzgebiete*, J. Springer (1933).
- [15] M. VAN LAMBALGEN, Von Mises' definition of random sequences reconsidered, *J. Symb. Logic* **52** (1987), 725 - 755.

- [16] M. VAN LAMBALGEN, *Random Sequences*, Dissertation, Dept. of Mathematics, University of Amsterdam (1987).
- [17] M. VAN LAMBALGEN, The axiomatisation of randomness, *J. Symb. Logic* **55** (1990), 1143 - 1167.
- [18] M. VAN LAMBALGEN, Independence, randomness and the axiom of choice, *J. Symb. Logic* **57** (1992), 1274 - 1304.
- [19] M. VAN LAMBALGEN, Independence structures in set theory, in: W. Hodges et al. (eds.), *Logic Colloquium '93*, Oxford University Press (1996).
- [20] M. LI, P.B.M VITANYI, *An introduction to Kolmogorov complexity and its applications*, Springer (1993).
- [21] R. VON MISES, Grundlagen der Wahrscheinlichkeitsrechnung, *Math. Z.* **5** (1919), 52-99.
- [22] R. VON MISES, *Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und theoretische Physik*, Deuticke (1931).
- [23] R. VON MISES, Über Zahlenfolgen die ein Kollektiv-ähnliches Verhalten zeigen, *Math. Ann.* **108** (1933), 757-772.
- [24] R. VON MISES, *Wahrscheinlichkeit, Statistik und Wahrheit*, 2<sup>nd</sup> ed., J. Springer (1936).
- [25] R. VON MISES, Quelques remarques sur les fondements du calcul des probabilités, in [9], 57-66.
- [26] R. VON MISES, On the foundations of probability and statistics, *Ann. Math. Stat.* **12** (1941), 191-205; 215-216.
- [27] R. VON MISES, H. Geiringer, *Mathematical theory of probability and statistics*, Academic Press (1964).
- [28] R. VON MISES, *Probability, Statistics and Truth* (English translation of 3<sup>rd</sup> ed. of *Wahrscheinlichkeit, Statistik und Wahrheit*; cf. [24]), Dover (1981).
- [29] K.R. POPPER, *Logik der Forschung*, J. Springer (1935). English translation: *Logic of scientific discovery*, Hutchinson & Co. (1975).
- [30] K.R. POPPER, *Realism and the aim of science*, Rowman and Littlefield (1983).
- [31] L. SKLAR, *Physics and chance*, Cambridge University Press (1993).
- [32] H. STEINHAUS, Les probabilités dénombrables et leur rapport à la théorie de la mesure, *Fund. Math.* **4** (1922), 286-310.
- [33] A.S. TROELSTRA, *Choice sequences*, Oxford University Press (1977).
- [34] J. VILLE, *Étude critique de la notion de collectif*, Gauthiers-Villars (1939).
- [35] A. WALD, Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung, *Ergebnisse eines math. Koll.* **8** (1936), 38-72.
- [36] A. WALD, Die Widerspruchsfreiheit des Kollektivbegriffes, in [9], 79-99.

DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF AMSTERDAM  
michiell@fwi.uva.nl

