2 Roots of Randomness: Von Mises' Definition of Random Sequences

2.1 Introduction In 1919 Richard von Mises (1883–1957) 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 set of axioms given in Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung* of 1933. The defeat of von Mises' theory was sealed at a conference on probability theory in Geneva (1937), where Fréchet gave a detailed account of all the objections that had been brought to bear against von Mises' approach.

We believe that this debate, for all its vigor, has failed to produce a careful analysis of von Mises' views and the new concepts he introduced. In fact, when one reads the various contributions, one is immediately struck by its monotony: the same objections and refutations are repeated over and over again, with scarcely any new elements being brought in. (There is one major exception: the objections based on a construction due to Ville (1937).) When one takes into account the considerable scientific acumen of the participants in the debate, this monotony may be a cause for surprise.

In the following pages, we shall attempt both to analyse von Mises' theory in detail and to examine the reasons why the debate which ensued after its publication failed to lead to satisfaction. Our guiding principles in this analysis will be twofold.

First, we believe that von Mises' characterisation of random sequences has great intuitive appeal, for all its imprecision. We do not regard the lack of precision itself as objectionable. Instead, we subscribe to Kreisel's doctrine of *informal rigour*:

> The 'old fashioned' idea is that one obtains rules and definitions by analysing intuitive notions and putting down their properties. This is certainly what mathematicians thought they were doing when defining length or area, or, for that matter logicians when finding rules of inference or axioms (properties) of mathematical structures such as the continuum. [...] What the 'old fashioned' idea assumes is quite simply that the intuitive notions are *significant* , be it in the external world or in thought (and a *precise* formulation of what is significant in a subject is the result, not the starting point of research into that subject).
> Informal rigour wants (i) to make this analysis as precise as possible (with the means available), in particular to eliminate doubtful properties of the intuitive notions when drawing conclusions about them; in particular not to leave undecided questions which can be decided by full use of evident properties of these intuitive notions [52,138].

It will be seen, for instance, that the notion of Kollektiv is at least clear enough to refute the often repeated allegiation of inconsistency. We do not, however, claim to have reached the limits of analysis (but perhaps the idea of the ultimate analysis does not even make sense).

Second, we try to explain the sterility of the debate by assuming that the participants had widely diverging, but in part unarticulated, opinions on the foundations of mathematics and probability. We shall meet instances of this phenomenon when we discuss the alleged inconsistency of Kollektivs (in 2.3.3) and the force of Ville's objection (in 2.6.2).
The conclusion of our analysis will be that the criticisms directed against von Mises' theory are either misguided (such as the charge that von Mises was working with a wrong concept of what axiomatisation should be) or based on foundational views which are not his (the alleged inconsistency, or the objection that Kollektivs do not always satisfy the law of the iterated logarithm). One may then pursue the debate at the level of foundational issues, but here, it is much more difficult to decide who is right and who is wrong. And for our purpose, the conclusion that different views on the foundations of probability may lead to different requirements on definitions of random sequences, is sufficient to motivate the technical work of subsequent chapters.

The plan of this chapter is as follows. In 2.2 we examine von Mises' version of the frequency interpretation, its surprising consequences and its possible rival, the propensity interpretation. In 2.3 we introduce Kollektivs and discuss their metamathematical status. 2.4 is centered around the demonstration that any form of the frequency interpretation assumes that the phenomena to which it is applicable are Kollektivs. In 2.5 we study some of the attempts to achieve precision in the definition of Kollektivs. 2.6 is devoted to a discussion of the objections brought forth by Fréchet. Our conclusions will be summed up in 2.7.
It will be clear from this outline that we shall mostly be concerned with two problems and their relation: the interpretation of probability and the definition of random sequences.

2.2 The frequency interpretation of probability

2.2.1 Methodological considerations. In the early thirties, two books were published on the foundations of probability theory, which express widely divergent attitudes: the *Wahrscheinlichkeitsrechnung*, von Mises' definitive treatise (1931) and the *Grundbegriffe der Wahrscheinlichkeitsrechnung* by Kolmogorov (1933). A convenient starting point for a discussion of von Mises' views is given by the following juxtaposition of quotations:

> Die Wahrscheinlichkeitstheorie als mathematische Disziplin soll und kann genau in demselben Sinne axiomatisiert werden wie die Geometrie oder die Algebra. Das bedeutet, daß, nachdem die Namen der zu untersuchenden Gegenstände und ihrer Grundbeziehungen sowie die Axiome, denen diese Grundbeziehungen zu gehorchen haben, angegeben sind, die ganze weitere Darstellung sich ausschließlich auf diese Axiome gründen soll und keine Rücksicht auf die jeweilige konkrete Bedeutung dieser Gegenstände und Beziehungen nehmen darf.
>
> Dementsprechend wird im §1 der Begriff eines *Wahrscheinlichkeitsfeldes* als eines gewissen Bedingungen genügenden Mengensystems definiert. Was die Elemente dieser Mengen sind, ist dabei für die mathematische Entwicklung der Wahrscheinlichkeitsrechnung völlig gleichgültig (man vergleiche die Einführung der geometrische Grundbegriffe in HILBERTs "Grundlagen der Geometrie" oder die Definitionen von Gruppen, Ringen und Körpern in der abstrakten Algebra).
>
> Jede axiomatische (abstrakte) Theorie läßt bekanntlich unbegrenzt viele konkrete Interpretationen zu. In dieser Weise hat auch die mathematische Wahrscheinlichkeitstheorie neben derjenigen ihrer Interpretationen, aus der sie aufgewachsen ist, auch zahlreiche andere. Wir kommen so zu Anwendungen der mathematische Wahrscheinlichkeitstheorie auf Untersuchungsgebiete, die mit den Begriffen des Zufalls und der Wahrscheinlichkeit im konkreten Sinne dieser Begriffe nichts zu tun haben (Kolmogorov [44,1]).

> Die Wahrscheinlichkeitstheorie wird in dieser Vorlesungen aufgefaßt als eine *mathematische Naturwissenschaft* von der Art etwa wie die Geometrie oder die Mechanik. Ihr Ziel ist es, für eine bestimmte Gruppe beobachtbarer Erscheinungen, die Massenerscheinungen und Wiederholungsvorgänge, eine übersichtliche Beschreibung zu geben, wie sie die Geometrie für die räumlichen, die Mechanik für die Bewegungserscheinungen liefert. An der Spitze einer derartigen Theorie stehen Aussagen, durch die die Grundbegriffe definiert werden und die man oft Axiome nennt; in ihnen kommen allgemeine Erfahrungseinhalte zur Verwertung, ohne daß sie unmittelbar als Erfahrungssätze angesprochen werden dürften. Aus den Axiomen werden dann auf deduktivem Wege, oder, wie man jetzt besser sagt, durch "tautologische Umformungen" mannigfache Sätze gewonnen, die vermöge des Zusammenhanges, der zwischen den Grundbegriffen und der Erfahrungswelt besteht, bestimmten, durch Beobachtung nachprüfbaren Tatbeständen entsprechen. So weist die Theorie am Anfang und am Ende jeder Gedankenreihe Berührung mit der Welt der Beobachtungen auf; ihren eigentlichen Inhalt aber, der uns vorzugsweise beschäftigen wird, bilden die rein *mathematischen Überlegungen, die zwischen dem Anfang und dem Ende stehen* (von Mises [68,1]).

These quotations emphasize two different aspects of the mathematical method. The quotation from Kolmogorov is concerned mainly with faultless derivations from axioms, which should proceed regardless of the actual meanings of the primitive concepts involved.

Von Mises, of course, does not deny the importance of this procedure, but he stresses the role of mathematics in describing real structures, in as much detail as is necessary, a feature less prominent in Kolmogorov's book.
The following quotation from von Mises' *Wahrscheinlichkeit, Statistik und Wahrheit* [70] further clarifies the sense in which probability theory is *mathematische Naturwissenschaft* :

> Die Wahrscheinlichkeitsrechnung (oder die Theorie der zahlenmäßig erfaßbaren Wahrscheinlichkeiten) ist die Theorie bestimmter, der Beobachtung zugänglicher Erscheinungen, der Wiederholungs– und Massenvorgänge etwa vom Typus der Glücksspiele, der Bevölkerungsbewegung, der Bewegung Brownscher Partikel usf. Das Wort "Theorie" ist hier in demselben Sinn gemeint wie die Hydromechanik Theorie der Flüssigkeitsströmungen, die Thermodynamik Theorie der Warmevorgänge, die Geometrie Theorie der räumlichen Erscheinungen heißt [70,128].

Statements such as these have led critics (e.g. Feller in his talk at the Geneva conference on probability theory [23,9]; see also Fréchet [28]) to object that von Mises' conception of a scientific theory was not true to the example set by Hilbert's *Grundlagen* and confused mathematical and empirical considerations; and since Kolmogorov's theory did not fall prey to this alleged confusion, it had to be preferred.

This objection is untenable. The axiomatisations of Kolmogorov and von Mises both attempt to provide a *rigorous* mathematical foundation for probability theory, but they choose, as we shall see, different sets of primitive terms. In particular, perhaps somewhat surprisingly, the term "probability" does not occur in von Mises' axioms, but is a defined notion, whereas it *is* a primitive term in the Kolmogorov axioms. These different languages reflect different motives, as Kolmogorov was well aware. Von Mises believed that only the frequency interpretation of probability makes sense and attempts to say in mathematical terms what this interpretation amounts to. Kolmogorov's preferences are expressed in the continuation of the passage cited above:

> Die Axiomatisierung der Wahrscheinlichkeitsrechnung kann auf verschiedene Weisen geschehen, und zwar beziehen sich diese verschiedenen Möglichkeiten sowohl auf die Wahl der Axiome als auch auf die der Grundbegriffen und Grundrelationen. Wenn man allerdings das Ziel der möglichen Einfachheit des Axiomensystems und des weiteren Aufbaus der darauf folgenden Theorie im Auge hat, so scheint es am zweckmäßigsten, die Begriffe eines zufälligen Ereignisses und seiner Wahrscheinlichkeit zu axiomatisieren. Es gibt auch andere Begründungssysteme der Wahrscheinlichkeitsrechnung, nämlich solche, bei denen der Wahrscheinlichkeitsbegriff nicht zu den Grundbegriffe zählt, sondern durch andere Begriffe ausgedrückt wird [a footnote refers to von Mises]. Dabei wird jedoch ein anderes Ziel angestrebt, nämlich der größtmögliche Anschluß der mathematischen Theorie an die empirische Entstehung des Wahrscheinlichkeitsbegriffes [44,2].

Although Kolmogorov is clearly aware of the possibility of different axiomatisations of probability theory, this passage has generally been overlooked by von Mises' critics (a phenomenon which will recur again). Accordingly, Kolmogorov was used unwillingly as support for a cause that was not his.

Our attitude toward the problem of axiomatising probability theory is as follows. There is no need to deviate from the Kolmogorov axioms in purely mathematical investigations. But von Mises' theory is a useful (indeed necessary) counterpart to that of Kolmogorov, since it attempts to provide a frequentistic interpretation for the theorems of probability theory which are, strictly speaking, statements about *measure* only. Interestingly, this attempt does not always succeed, as with the law of the iterated logarithm when formulated as a theorem about infinite sequences. Such cases lead one to question the empirical content of some of the results of measure theoretic probability theory. (Ideally, a derivation from the Kolmogorov axioms should be followed by a derivation from von Mises' axioms, to see what the result really means.)

Furthermore, the frequency interpretation is not so crystal clear as to render superfluous attempts at a precise formulation; even a rough formalisation shows that there exist essentially different versions (see 2.2.3). Not least among the merits of von Mises' theory is that it pursues one such interpretation, called *strict frequentism* in 2.2.3, to the bitter end.

2.2.2 Kollektivs (informal exposition). When we look at the list of examples of phenomena to which probability theory should be applicable (see the quotation from von Mises' [70] on p.9): coin tossing, demographic events, Brownian motion, it is clear that these examples exhibit a common trait: either an unlimited repetition of an experiment or a great number of events is involved. But the examples also differ in some probabilistic properties; in modern parlance, we would say that coin tossing is a Bernoulli process, whereas Brownian motion is a Markov process. The essence of von Mises' theory is, that it uses properties of games of chance such as coin tossing as a tool to deduce properties of other processes *and* as an instrument to define probability. In order to have at our disposal a technical term for this privileged case, we introduce the word *Kollektiv*.

Informally, a Kollektiv is a sequence of elements of a sample space (which are also called *attributes*), which is akin to a typical sequence of events produced by coin tossing. To say precisely what "akin to" means, we have to list some of the properties of coin tossing which we regard as essential. Two of these properties, amply verified by experience, are:

(i) Approximate stability of the relative frequency of an attribute if the number of observations (or experiments) is increased;

(ii) The impossibility of a successful gambling strategy, that is, the impossibility of making unlimited amounts of money in a game of chance, using some kind of system. A gambling strategy may roughly be thought of as a rule for betting on some trials and skipping others.

The informal statement of these properties of Kollektivs is sufficient to explain von Mises' version of the frequency interpretation. A more formal statement will be given in 2.3; but, in a sense, all the subsequent chapters are devoted to a formalisation of properties 2.2.2(i) and (ii).

2.2.3 Strict Frequentism: "Erst das Kollektiv, dann die Wahrscheinlichkeit"

We may now give an explicit, albeit informal, definition of probability.

2.2.3.1 Definition The *probability* of an attribute in a Kollektiv equals the relative frequency of that attribute within the Kollektiv.

In 2.3 a more formal definition (involving infinite Kollektivs and limiting relative frequencies) will be given, but the salient points can be illustrated as well using the finite version. Von Mises summarizes his attitude in the slogan: "Erst das Kollektiv, dann die Wahrscheinlichkeit", an innocuous–sounding formula with far reaching implications.

1. There is no probability of an individual event, e.g. that of Rachel dying at age 40, as such. One may, however, *metaphorically* assign various probabilities to this event, corresponding to each Kollektiv to which Rachel belongs: that of female heavy smokers, that of sports car drivers and so forth. So far, only the first property of Kollektivs, 2.2.(i), is used.
2. The second property of Kollektivs gives a special twist to the definition of probability and severely restricts its applicability. Basically, defining the probability of an attribute with respect to a Kollektiv only, means that probability enjoys a multiplicative property. Details will be given in 2.4, but an example will make clear what we mean. The paradigmatic example of a Kollektiv is a sequence of tosses with a fair coin. The probability of heads in such a Kollektiv will (approximately) be $\frac{1}{2}$; and the probability of heads on two consecutive tosses will be $\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$. Now the relative frequency $\frac{1}{2}$ may be called a probability *only* if this multiplicative property holds. In this respect, von Mises' nomenclature differs from that of Kolmogorov, who requires of a probability only that it be a positive measure with norm one. The multiplicative property creeps in only afterwards, when he defines the notion of independence, two events being independent if the probability of their joint occurrence equals the product of the probabilities of the events themselves. He then duly remarks that it is this notion of independence which distinguishes probability theory from measure theory [44,8]. Of course, mass phenomena which do not satisfy the second property of Kollektivs can be

handled as well in the theory, but von Mises' convention is such that in this case, the relative frequency is not a probability.

3. Von Mises' definition is not the only one which establishes some connection between probability and relative frequency. In 2.4, and again in 2.6, we shall meet the *propensity interpretation*, which proceeds along rather different lines. 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 (2.2.3.1):

> Die Wahrscheinlichkeit, Sechs zu zeigen, ist *eine physikalische Eigenschaft* eines Würfels, von derselben Art, wie sein Gewicht, seine Wärmedurchlässigkeit, seine elektrische Leitfähigkeit usw [70,16].

but this aspect of probability does not figure in the definition.

To appreciate the strictness with which von Mises himself applied his doctrine, it is instructive to consider the case of attributes of probability zero. If computed in a finite Kollektiv, probability zero is of course equivalent to the non-occurrence of that attribute. But when our Kollektiv is infinite, as the precise version of the explicit definition of probability (2.2.3.1) requires, then probability zero of an attribute is compatible with the attribute occurring infinitely often. Although this idea is formulated in terms of infinite Kollektivs, it has consequences for observable events. If $x \in 2^{\omega}$ is a Kollektiv with probability distribution (1,0) and if we derive from x a Kollektiv $y \in (2^n)^{\omega}$ by selection and combination as is done in 2.4, then some of the y_j (which represent finite, observable populations) may contain 1's, although the probability of 1 is zero.

In the case of a continuous sample space, the idea that probability zero does not imply impossibility is universally accepted. But the application of this idea to a discrete sample space seemed too much to swallow, witness the following remark by Martin–Löf, when he contrasts his own approach to the definition of random sequences with that of von Mises:

> [...] an event with vanishing limit frequency is actually impossible. This contrasts sharply with the conception of von Mises, who explicitly stated that the opposite might occur. It seems as if he strained his seldom failing intuition on this point in order not to conflict with his somewhat arbitrary definition of randomness [62,619].

We shall see in 2.5–6 that this divergence of opinions, small as it may seem, actually points to irreconcilable intuitions as regards the principles which should govern the definition of Kollektivs. (And our conclusion will be that Martin-Löf's definition and its relatives are rather more arbitrary than that of von Mises.)

4. Another way to illustrate the strictness of strict frequentism, is to consider the role played by the laws of large numbers (and in fact all weak and strong limit laws of probability theory)

in von Mises' set–up; or rather, the role they do *not* play. We introduce some notation first.

2.2.3.2 Definition Let $p \in [0,1]$. The measure μ_p on 2^ω is defined to be the product measure $(1-p,p)^\omega$. We put

$$LLN(p) := \{x \in 2^\omega \mid \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k = p\}.$$

2.2.3.3 Theorem *Strong law of large numbers* : $\mu_p LLN(p) = 1$.

An influential interpretation of probability (influential because apparently unconsciously adopted by most mathematicians), the *propensity interpretation*, holds that probability should primarily be thought of as a physical characteristic. Now von Mises could concede this much (cf. the passage quoted on p.12) 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 e.g. the passage from Fréchet [28] cited in 2.6.) In other words, propensity theorists claim that it is possible to derive statements on relative frequencies from premisses which are (almost) probability–free.
We present the alleged derivation of the frequency interpretation in the form given in Popper's *Realism and the Aim of Science* [83]. This presentation might seem anachronistic. But expositions of the propensity interpretation which do show some awareness of its assumptions are rare (the reader may wish to compare Popper's version with that of Fréchet, quoted in 2.6.1). Since the propensity interpretation has inspired some of the work on random sequences in the literature, we have chosen to present it in its (for all its naiveté) most articulated form[1]. The derivation goes 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 μ_p–measure one. Now the auxiliary hypothesis comes in. After explaining why the weak law cannot be used in this context, Popper goes on to say:

> The case is different if we obtain a probability that is *exactly* equal to 1 (or 0, as in the case of measure zero). Admittedly, even in this case, "probability" has to mean something connected with frequency if we are to obtain the required result. But no precise connection need be assured – no limit axiom and no randomness axiom [the two conditions formally defining Kollektivs; see 2.3]; for these have been shown to be valid except for cases which have a probability (a measure) zero, and which therefore may be neglected. Thus 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* [84,380].

Stated like this, the argument is quite like the type of reasoning employed in the ergodic foundation of statistical mechanics. Here, one tries to justfy the auxiliary hypothesis on physical grounds:

> [...] one could have an invariant ensemble where every particle moves on the same straight line reflected at each end from a perfectly smooth parallel wall. The obviously exceptional character of this motion is reflected mathematically in the fact that this ensemble, though invariant, is confined to a region of zero "area" on *S* [a surface of constant energy] and therefore has no ensemble density. To set up such a motion would presumably be physically impossible because the slightest inaccuracy would rapidly destroy the perfect alignment (Lebowitz and Penrose [81,24]; for a variation on this argument, see Malament and Zabell [60]).

Von Mises 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:

> 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 [84, 381]).

In the same way, the ergodic theorem plus the auxiliary hypothesis are taken to explain the statistical behaviour of gases; and we may remark in passing that von Mises also declines such uses of ergodic theory (see the last chapter of [68]).

It is not our purpose here to judge between these two interpretations of probability, strict frequentism and propensity interpretation. We only note that the assumptions underlying the interpretations are of a rather different character:
– The auxiliary hypothesis of the propensity interpretation is of highly *theoretical* nature and badly in need of justification; indeed it is not clear what form a justification should take. In any case it seems more profitable to study concrete examples of its use, for instance in statistical mechanics.
– Von Mises starts from two brute *facts*, amply corroborated by experience, and makes no attempt to explain these facts.
Obviously, in order to turn Popper's *deduction* of the frequency interpretation into a true *explanation,* his premisses have to be analysed further. But since we shall show in the sequel that adherence to the propensity interpretation justifies requirements on the definition of

Kollektivs which are quite unjustified from a strict frequentist point of view, we ask the reader to be alive to both possibilities of interpretation.

2.2.4 Structure and task of probability theory. After all that has been said, it will come as no surprise that the outward appearance of von Mises' theory is rather different from that of Kolmogorov's. We now proceed to give a concise description of its structure; the mathematical details, in so far as they are relevant, will be given in 2.3 and 2.4.

Von Mises emphatically presents probability theory as an empirical theory, designed to transform data, in the form of probabilities, into predictions or explanations, again in the form of probabilities (we omit complications due to the fact that some relative frequencies, e.g. those in Markov processes, are not probabilities. See for these [68] and [70]). The theory should be judged solely on its empirical merits, its adequacy in predicting or explaining observable phenomena.

Since the data are probabilities, they are supplied in the form of relative frequencies in Kollektivs. It follows that probability theory must consist of rules transforming given Kollektivs into other Kollektivs. Accordingly, the axioms of the theory posit the validity of one type of transformations (so called *place selections*); the validity of the other necessary rules of transformation is derivable from these axioms. Consequences of the axioms include the Kolmogorov axioms (albeit with finite additivity only), the multiplicative property alluded to above (in 2.2.3.2) and the formula for conditional probability.

The axioms themselves are a translation into mathematical terms of the facts of experience mentioned in 2.2.2: approximate stability of relative frequencies in long series of trials and the impossibility of a successful gambling strategy. As such, these axioms exhibit a certain amount of idealisation; in particular, the Kollektivs, which in practice are finite, are represented by infinite sequences. This procedure is equally justified as concept formation in geometry: the ideal entities are introduced for their technical advantages, but their properties are studied only on so far as they are relevant to the prediction of observable, hence finite, phenomena. If the infinities can be eliminated, then so much the better.

It would, therefore, be a grave mistake to suppose that von Mises' theory is a *mathematical* theory of *infinite* Kollektivs, as is, for instance, the definition of random sequences proposed by Martin–Löf (for which see Chapter 3). Von Mises introduced infinite Kollektivs only for their technical advantages, not as autonomous objects of study [70,103-4]. We shall discuss Kolmogorov's attempt to define *finite* Kollektivs in 5.2.

2.3 Axiomatising Kollektivs. We now introduce a mathematical description of Kollektivs, essentially by expressing properties 2.2.2(i),(ii) in mathematical terms.

The formal set–up is as follows. 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 $x \in M^{\omega}$.

2.3.1 The axioms (as given by von Mises [67,57]).

2.3.1.1. Axiom A sequence $x \in M^{\omega}$ is called a *Kollektiv* if

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

(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^{ω}, by deleting all terms x_n which do not belong to either A or B. Now let Φ be an *admissible place selection*, i.e. a selection of a subsequence Φx' 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 wird."

Then $P'(A) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_A(\Phi x')_k$ and $P'(B) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_B(\Phi x')_k$ exist and

$\frac{P'(A)}{P'(B)} = \frac{P(A)}{P(B)}$ when $P(B) \neq 0$.

A few remarks on the above definition are in order.

1. The set of axioms 2.3.1.1 will alternatively be called a *definition* of Kollektivs. In Hilbertian jargon, 2.3.1.1 provides an implicit definition of Kollektivs (rather than of probability, as in the Kolmogorov axioms).
2.The quantifier "for all $A \subseteq M$" should not be taken too seriously. In the *Wahrscheinlichkeitsrechnung* [68,17] 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".
3. The function P defined in (i) is called *the probability distribution determined by the Kollektiv x*, in conformity with the slogan of 2.2.3. We shall occasionally use the phrase "$x \in M^{\omega}$ is a Kollektiv with respect to distribution P"; this phrase might suggest that the distribution is primary, but should be taken to mean only that P satisfies (i). In the same vein, the phrase "a fair coin" is used to designate a coin whose relative frequencies are approximately equal to ½. It will be clear from the discussion in 2.2.3 that no reference to the physical properties of the coin is intended.

4. We shall use the phrase "$x \in M^{\omega}$ is invariant under an admissible place selection Φ" to mean that the limiting relative frequency in the subsequence selected by Φ are the same as those in x. The notation " $x \in M^{\omega}$ " should be read to mean only that each term of x is an element from M; we do not imply that Kollektivs are elements of a universe described by Zermelo–Fraenkel set theory. Similarly, the notation "Φx" for the subsequence selected from x by the admissible place selection Φ should, until further notice (in 2.5) *not* be read as the application of a function $\Phi: M^{\omega} \to M^{\omega}$ to x, since at this stage it is not clear that an admissible place selection is indeed a function. The reasons for this caution will gradually become clear in the sequel. (Note also that the notation "Φx" is ambiguous: do we keep track of where the terms of the subsequence originate in x?)

5. 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 Φ is an admissible place selection, $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} (\Phi x)_k = P(\{1\})$.

As will be made clear in 2.4, the more elaborate condition is necessary in order to ensure the validity of the rule for conditional probabilities:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

It is interesting that the validity of this rule has to be built in blatantly into the axioms (thus emphasizing its empirical origin), especially in view of attempts such as Accardi's [1] to put the blame for the failure of classical probability theory in quantum mechanics upon this rule. (Wald [100, 41-2] claims that, also in the general case, condition (ii) can be reduced as for Kollektivs in 2^{ω}; but his proof uses evidently non–admissible place selections.)

6. The condition "$P(B) > 0$" is necessary for the ratios in (ii) to be well-defined. On the other hand, it is clear that in von Mises' set-up conditionalisation on a set B is possible if B occurs infinitely often in the Kollektiv, a strictly weaker requirement (cf. the discussion in 2.2.3). One could extend condition (ii) to incorporate B which occur infinitely often, but for which $P(B) = 0$ by means of non-standard analysis: if B occurs infinitely often, then, in any non-standard universe, P(B) is a *positive* infinitesimal, so the ratios in (ii) are well defined.

It will be noted that 2.3.1.1, and especially condition (ii) does not fully conform to present standards of mathematical rigour. In the sequel we shall review a number of attempts to make this condition precise; but let us first try to give an idea of what is meant by means of some examples.

2.3.1.2 Example Admissible place selections may be viewed as gambling strategies: if n is chosen, that means that a bet is placed on the outcome of the n^{th} trial; otherwise, the n^{th} trial is skipped. In the examples we consider the simplest case, cointossing; in other words, Kollektivs x in 2^{ω}.

(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 $n-9^{th}$,......., $n-1^{st}$ 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^{th} toss with the second coin is 1.

Condition (ii) is intuitively satisfied in all three cases, although in (c) a heavy burden is put upon 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*.

Condition (ii) will usually be called the *axiom of randomness* (from *Regellosigkeitsaxiom*). Von Mises alternatively uses the designation *principle of the excluded gambling strategy* (from *Prinzip vom ausgeschlossenen Spielsystem*). Unfortunately, he uses the term "gambling strategy" in two different senses (which he evidently considers to be the same):

> Diese Unmöglichkeit, die Gewinstaussichte beim Spiel durch ein Auswahlsystem zu beeinflussen, die *Unmöglichkeit des Spielsystems*....[70,29]

> Daß sie nicht zum gewünschten Ziele führten, nämlich zu einer Verbesserung der Spielchancen, also zu einer Veränderung der relativen Häufigkeiten...[70,30].

Apparently, von Mises thinks that a gambling strategy making unlimited amounts of money can operate only by selecting a subsequence of trials in which relative frequencies are different. It was shown by Ville that this idea is mistaken: there exist gambling strategies (called *Martingales*)which cannot be represented as place selections. We shall come back to this point in 2.6.2 and in 3.4.

Put concisely, the definition of Kollektivs consists of two parts: global regularity (existence of limiting relative frequencies) and local irregularity (invariance under admissible place selections implies that a Kollektiv is unpredictable). Both separately and in conjunction, these parts have come under fire. The most pertinent objections will be reviewed in 2.6, except one, the charge that the theory is outright inconsistent. In view of its urgent character, this charge will be taken up in 2.3.3, after a brief review of some of the consequences of the axioms in 2.3.2.

2.3.2 Some consequences of the axioms The following propositions are literal translations of some of von Mises' *Sätze* in [67].

2.3.2.1 Proposition [67,57] Let $x \in M^{\omega}$ be a Kollektiv, P the probability distribution induced by x. Then P(M) = 1 and P is finitely additive.

This proposition might seem to be trivially true, but in fact its truth value is undetermined until it has been specified on which subsets of M P is defined. Wald [100,46] has shown that for continuous sample spaces M, there exists no non–atomic P defined on all subsets of M; but a finitely additive probability can be defined for all Peano–Jordan measurable subsets of M.

2.3.2.2 Proposition [67,58] An admissibly chosen subsequence of a Kollektiv x is again a Kollektiv, with the same distribution.

Proof Composition of two admissible place selections yields a new selection which still proceeds *ohne Benützung der Merkmalunterschiede*.
This proposition will be the starting point for some of our investigations in Chapters 4 and 5.

2.3.2.3 Proposition [67,59] A Kollektiv x is determined completely by its distribution; it is not possible to specify a function $n \to x_n$.

Proof Choose $A \subseteq M$ such that $0 < P(A) < 1$. If there were such a function we could use it to define an admissibly selected subsequence of x which consists of elements of A only. □

This consequence contains the essence of the new concept: a Kollektiv has no other regularities than frequency regularities. Von Mises adds the comment that 2.3.2.3 implies

> das 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
> [67,60].

In other words, Kollektivs are *new* mathematical objects, not constructible from previously defined objects. Hence in one place [68,15; see also 70,112] von Mises compares Kollektivs to Brouwer's free choice sequences, one extreme example of which is the sequence of outcomes produced by successive casts of a die[2]. In another place he contrasts his approach with that of Borel [8], in a way which makes clear that Kollektivs are not to be thought of as numbers, i.e. *known* objects:

> ...den von Borel u.a. untersuchten Fragen (z.B. über das Auftreten einzelner Ziffern in den unendlichen Dezimalbrüchen der irrationale Zahlen), wo das Erfülltsein oder Nicht–Erfülltsein der Forderung II [i.e. 2.3.2.1.(ii)] ohne Bedeutung ist [67,65].

The reference is to Borel's Strong Law of Normal Numbers, i.e. Theorem 2.3.2.3 for $p = \frac{1}{2}$ (or rather its analogue for sequences in 10^{ω})! To modern eyes, accustomed to set theory, von Mises' statement may look surprising: (dyadic) numbers and Kollektivs (as they arise in a coin tossing game) can both be thought of as elements of Cantor space. But it will be seen time and again that the set theoretic perspective is not very helpful in understanding von Mises' ideas and the debates to which they gave birth; because at that time, these set theoretic notions were still fresh and not part of the thinking habits of mathematicians. (Neither is set theory very helpful in understanding Borel's ideas on probability; see Novikoff and Barone [79] for a particularly disastrous example of prejudiced historiography. We shall come back to this point in 2.6.)

Digression Perhaps Borel wouldn't have disagreed with von Mises' comment. When he introduces the considerations which lead up to the strong law of normal numbers, he states [8,194–5]

> 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 égale a 1/q (dans le cas de la base q) de prendre chacun de ces valeurs possibles: 0,1,2,3,...., q–1.
> Il n'est pas besoin d'insister sur le charactère 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 précisement de se rendre compte de la mesure dans laquelle les choses se passent *comme si* cette hypothèse est vérifiée.

In this context it may be interesting to remark that, at the time when Borel proved his strong law (1909), it was by no means considered to be self–evident; in fact one expected the opposite result. Here is Hausdorff's comment [37,420]

> Dieser Satz ist merkwürdig. Auf der einen Seite erscheint er als plausibele 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* [94,286]. 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. It is perhaps useful to keep in mind that such was the intellectual climate in which von Mises first published his ideas.

2.3.3 Do Kollektivs exist? Objections to von Mises theory were not long in coming. Although his efforts met with sympathy, doubts were raised concerning the soundness of the foundation. In this respect the following comment is typical:

> Ich glaube nicht, daß Versuche, die von Misessche Theorie rein mathematisch zu fassen, zum Erfolg führen können, und glaube auch nicht daß solche Versuche dieser Theorie zum Nutzen gereichen. Es liegt hier offentsichtlich der sehr interessante Fall vor, daß ein praktisch durchaus sinnvoller Begriff – Auswahl ohne Berücksichtigung der Merkmalunterschiede – prinzipiell jede rein mathematische, auch axiomatische Festlegung ausschließt. Wohl aber wäre es wünschenswert, das sich diesem Sachverhalt, der vielleicht von grundlegender Bedeutung ist, das Interesse weiter mathematischen Kreise zuwendet (Tornier [96,320]).

A catalogue of objections (with their rebuttals) will be given in 2.6, but one simple objection, reiterated ad nauseam, will be dealt with rightaway. The objection states that the appeal to the "abstrakten logischen Existenz" in 2.3.2.3 is illusory, since it is easily shown that Kollektivs with respect to non–trivial distributions do not exist.

For 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. The above argument, purporting to show the inconsistency of 2.3.1.1 is translated almost literally from Kamke's report to the Deutsche Mathematiker Verein [41,23]. (It may not be entirely out of place to mention that Kamke is the author of a textbook on set theory.) The argument calls for several remarks.

1. It is obviously very insensitive to von Mises' intentions; in fact, it is almost verbally the same as the proof of 2.3.2.3, the proposition which states that a Kollektiv cannot be given by a function! Von Mises had no trouble in dismissing the argument: the set $\{n \mid x_n = 1\}$ does not define an admissible place selection since it uses *Merkmalunterschiede* in a most extreme way. The real problem is rather, to understand why the argument was considered to be convincing at all. It seems that this is one of those cases in which there was no common ground for discussion between von Mises and his adversaries. Kamke speaks as a set theorist: the set of all infinite binary sequences exists "out there", together with all its elements, some of which are Kollektivs. Hence the set $\{n \mid x_n = 1\}$ is available for admissible place selection

in much the same sense as is the set of primes (our example 2.3.1.2(a)).
Von Mises, on the other hand, considers Kollektivs to be *new* objects which, like choice sequences, are not pre–existent; hence $\{n \mid x_n = 1\}$ is *not* available. For him, $n \rightarrow x_n$ is not a legitimate mathematical function; functions are objects which have been constructed. (For evidence of von Mises' constructivist tendencies see, e.g., [71].)

2. Kamke's argument is somewhat beside the mark in that it fails to appreciate the purpose of von Mises' axiomatisation; namely, to provide a mathematical description for certain physical phenomena. The argument refers to what *could* happen, whereas von Mises' axioms are rooted in experience and refer to what *does* happen.
The empirical roots are twofold: in some cases (e.g. in example 2.3.1.2(c), where we use random selection) it is an empirical matter to decide whether a proposed place selection is admissible; and even if we have established to our satisfaction that a place selection is admissible (e.g. on a priori grounds, as for lawlike selections (examples 2.3.1.2(a,b)), the truth of the axiom is by no means self–evident, but at most a fact of experience.
An analogy may be helpful here. In various places (see for instance [70,30]) von Mises likens condition (ii) to the first law of thermodynamics. Both are statements of impossibility: condition (ii) is the principle of the excluded gambling strategy, while the first law (conservation of energy) is equivalent to the impossibility of a perpetuum mobile of the first kind.
It may be even more appropriate to compare condition (ii) to the second law of thermodynamics, the law of increase of entropy or the impossibility of a perpetuum mobile of the second kind, especially in view of Kamke's criticism. Indeed, Kamke's objection is reminiscent of Maxwell's celebrated demon, that "very observant and neat–fingered being", invented to show that entropy decreasing evolutions may occur. Maxwell's argument of course in no way detracts from the validity of the second law, but serves to highlight the fact that statistical mechanics cannot provide an absolute foundation for entropy increase, since it does not talk about what happens *actually*.

3. Another point completely overlooked by Kamke's argument is the *intensional* character of admissible selection, where we use "intensional" in Troelstra's sense:

> Whenever we are led to consider information on sets or sequences beyond their extensions or graphs, we shall speak loosely of "intensional aspects" [98,203].

Clearly, admissibility is not a property of the place selection itself; but, as can be seen from the definition ("Auswahl ohne Benützung der Merkmalunterschiede"), it also involves the consideration of the Kollektiv from which the choice is to be made, or perhaps the process generating that Kollektiv.

Only in the degenerate case where one is tempted to infer the admissibility of a place selection on a priori grounds (e.g. when the selection is lawlike) admissibility may be predicated of the place selection itself, but it must be kept in mind that this is an elliptical way of speaking only. It is not unusual for physical quantities to have an intensional character in the above sense. The notion of a disturbing measurement in quantum mechanics is intensional and likewise admits a degenerate case, namely the measurements which are disturbing because they destroy the system. In this example it is clear that the intensional element, the fact that "disturbing" is not a property of the observable representing the measurement, can be completely explained, using only extensional notions, in a more elaborate theory (via non–commuting operators etc.).

We do not, of course, mean to suggest that these considerations themselves suffice to instill precision in the phrase "Auswahl ohne Benützung der Merkmalunterschiede". But they do serve to show that Kamke has not grasped von Mises' point *and* to direct one's attention to possible formalisations of the enigmatic phrase.

Concluding this part of the discussion and having cleared the theory of the charge of outright inconsistency, we now take a closer look at its metamathematical status. Admittedly, the theory is not formalised, but then, formalisation is not an end in itself. One may expect to derive two benefits from formalisation: the possibility of mechanical checking of proofs, and a proof of consistency.
As can be guessed from the presence of *two* new primitive terms in 2.3.1.1, von Mises' theory is really two in one: *probability theory* , in which it is assumed that some Kollektiv is invariant under certain place selection; and an *explanation of invariance via admissibility*.
The structure of the first part is crystal clear: all notions can be defined in ordinary mathematical terms (even Kollektivs, as follows from results of Wald presented in 2.5) and proofs are just computations (as will be clear from the sample proofs given in 2.4).
Von Mises later came to regard this part as the essential mathematical part of the theory (see, for instance, [68] and [70]; we return to this point in 2.5); the verification that a certain Kollektiv is indeed invariant under a given set of place selections then had to proceed empirically. He considered admissibility to be the *intuitive* explanation of invariance under place selections, but admissibility as such dropped out of the theory [70,29].

The second part of the older theory (explanation of invariance via admissibility) is indeed less clear than the first part; but this does not mean that the notion of admissibility is completely unclear or even inconsistent. In particular, the notion is clear enough to show the validity of arguments like the proof of 2.3.2.2, which is of the form:

If Φ is an admissible place selection on x, then Ψ is an admissible selection on y.

The same type of argument occurs in 2.4, when it is shown that the Kollektivs are closed under certain operations (admissible place selections being a special case).

An axiomatisation of admissibility could proceed by *postulating* the validity of 2.3.2.2 and related propositions in 2.4, with an additional postulate which says that lawlike place selections are admissible. This is more or less the approach chosen by Dörge [22] and amounts to an implicit definition of admissibility.

We believe that the second part of the theory has enough physical plausibility to make further attempts at formalisation worthwhile. In 5.6 we present two different explicit definitions of admissibility, involving Kamae entropy and Kolmogorov complexity; we do not claim that these definitions exhaust the possible meanings of admissibility. Rather, these definitions should be viewed as different projections of the universe where Kollektivs "live", the formalisation of which still has to be found.

2.4 The use of Kollektivs In the previous section we examined the meaning of proposition 2.3.2.3 from the point of view of the foundations of mathematics.We saw that it laid the theory open to the (albeit unjustified) charge of inconsistency. Now we investigate its probabilistic meaning. On the face of it, proposition 2.3.2.3 seems to make von Mises' theory pointless: on the one hand a Kollektiv is completely determined by its distribution (in the sense that nothing more can be said about it), on the other hand, Kollektivs are deemed to be necessary for the interpretation of probability. Then a natural question arises: 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?

In what sense, then, do Kollektivs occur in computations, over and above their distribution? The answer, as we shall see, is that anybody who believes in the frequency interpretation and in the validity of the usual rules for probability, is bound to believe in Kollektivs. That is, not necessarily in the idealized, infinite Kollektivs as they occur in von Mises' axioms, but rather as finite approximations to these. In other words, *Kollektivs are a necessary consequence of the frequency interpretation.* This point is made by von Mises, when he states that

> Die Autoren, die die allgemeine Regellosigkeit "ablehnen" und durch eine beschränkte ersetzen, schließen entweder alle Fragen der Beantwortung aus, die nicht der von ihnen willkürlich gesetzten Beschränkung entsprechen; oder sie nehmen in jedem konkreten Fall die Regellosigkeit, die gerade gebraucht wird, als ein Datum der betreffenden Aufgabe an, was nur auf eine Änderung der Darstellungsform hinauslauft [70,128-9].

One of the main goals of this section is to establish the claim that Kollektivs are necessary for the frequency interpretation of probability (otherwise the reader might think that, von Mises' theory being superseded by Kolmogorov's, there is no use anymore in investigating Kollektivs). This will be done in 2.4.2. To do so, we need some facts concerning operations on Kollektivs, which will be presented in 2.4.1. There, we also have the opportunity to stress the differences in the treatment of *independence* in the theories of von Mises and Kolmogorov. In 2.4.3, we consider the role of the laws of large numbers in von Mises' theory, a subject already touched upon in 2.2.3.

2.4.1 The fundamental operations: definition and application.

2.4.1.1 Definition of the operations We indicate briefly how the usual rules of probability theory can be derived using 4 operations, which transform Kollektivs into Kollektivs. That is, we shall prove, using our intuitive understanding of admissibility, that these operations preserve *Kollektiv*hood. These proofs can be made fully rigorous if we start with a given set of place selections, in the spirit of von Mises' later ideas; alternatively, we may use the four operations to axiomatise admissibility.

1. *Place selection* This operation transforms a Kollektiv into a Kollektiv with respect to the same distribution; indeed, this is the content of proposition 2.3.2.2.
2. *Mixture* Let $x \in M^{\omega}$ be a Kollektiv with respect to a distribution P on M. Let N be a sample space and f: $M \rightarrow N$ a function (which, of course, must in some sense be constructive). Consider the sequence $y = (f(x_n))_n$ in N^{ω}. Obviously y induces the distribution Pf^{-1}. Moreover, y is a Kollektiv with respect to this distribution: since f is defined by a mathematical law, an admissible place selection operating on y can be transformed, using f, to an admissible place selection on x.
3. *Division* Let A be proper subset of M, $x \in M^{\omega}$ a Kollektiv with respect to P, and suppose that A occurs infintely often in x. Division allows one to define the conditional probability P(B|A) for $B \subseteq M$: we transform x into a sequence $x' \in A^{\omega}$ by retaining only those terms of x which belong to A. If we also suppose that P(A) > 0, then we may define (for $B \subseteq A$) P(B|A) := P(B)/P(A) and x' is a Kollektiv with respect to P(•|A). In fact, the whole point of the elaborate condition of randomness 2.3.1.1 (ii) is just to ensure that x' *is* Kollektiv (a point missed by Schnorr [88,18]). If A occurs infinitely often in x, but nonetheless P(A) = 0, we may use non-standard analysis as indicated in 2.3.1. If *P denotes the extension of P to the non-standard universe and st(•) the standard part map, the distribution in x' is given by P(B|A) = st(*P(B)/*P(A)). (Related ideas can be found in [104].)
4. *Combination* Let M,N be sample spaces, $x \in M^{\omega}$ a Kollektiv with respect to P, $y \in M^{\omega}$ a Kollektiv with respect to Q. Combining Kollektivs is the operation of forming the sequence $(\langle x_n,y_n \rangle)_n$ in $(M \times N)^{\omega}$. We then need to know conditions under which this sequence is again a Kollektiv and if so, with respect to which distribution. If we analyse the meaning of applying

an admissible place selection to a sequence $(<x_n,y_n>)_n$, we arrive at the following necessary and sufficient condition for this sequence to be a Kollektiv:

Independence Let x, y be as above. $(<x_n,y_n>)_n$ is a Kollektiv with respect to the distribution P×Q on M×N if x and y are *independent*[3] *Kollektivs*, i.e. if the following operation leads to a Kollektiv x'' in M^ω with distribution P:
Fix arbitrary $A \subseteq N$. Apply an admissible place selection to y, giving a subsequence $(n_k)_k$ of natural numbers and a sequence y' such that y'_k equals the n_kth term of y. Then select a subsequence x' from x as follows: the n_kth term of x is retained if $y'_k \in A$; and, lastly, apply an admissible place selection to x', giving x''. (It is not difficult to check that the relation of independence is symmetric. The last condition is necessary in order to ensure that x and y are themselves Kollektivs.)

Similarly, one may define independence of three Kollektivs: we say that x,y and z are independent Kollektivs if they are pairwise independent (in the above sense) and if each of them is independent (again in the sense introduced above) of the combination of the other two. The extension to n independent Kollektivs is routine.
In [70,58], von Mises calls the operation of selecting a subsequence x' from x as follows: the n_kth term of x is retained if $y'_k \in A$, *sampling*. We have met sampling already in example 2.3.1.2(c), as a special case of admissible place selection: if $x,y \in 2^\omega$ are Kollektivs supposedly generated by independent coins, choose those x_n for which $y_n = 1$. But note that in the above condition, sampling is used to *define* what it means for two Kollektivs to be independent.
The particular type of sampling displayed in example 2.3.1.2(c) will occur so often, that is denoted by a special symbol: for $x \in 2^\omega$, $y \in 2^\omega$, where y contains infinitely many ones, $^x/y$ is defined as:

$$(^x/y)_m = x_n \text{ if n is the index of the } m^{th} \text{ 1 in y.}$$

(This notation is slightly ambiguous; do we keep track of which n were chosen or not? We shall never need to.)

We now illustrate the condition of independence with two examples, one pertaining to two tosses with a single coin, the other to two coins, supposed to be physically independent. Whereas in Kolmogorov's theory these two cases are treated alike by *postulating* that probabilities multiply, in von Mises' theory the two cases are distinguishable in that in the first case independence, hence the product rule, is provable, while in the second case independence has to be assumed.

2.4.1.2 Examples

1. We are interested in the probability of obtaining two times heads with two tosses in succession of a fair coin. Let x be a Kollektiv with respect to distribution $(\frac{1}{2},\frac{1}{2})$. A new Kollektiv, representing the situation in which we are interested, is obtained as follows:
choose first those x_n for which n is odd, then those x_n for which n is even; then combine the two Kollektivs thus obtained, which gives $\xi = (<x_{2n-1},x_{2n}>)_{n\geq 1}$.
In this case it is *provable* that ξ is a Kollektiv with respect to the product distribution on $\{<0,0>,<0,1>,<1,0>,<1,1>\}$; in other words, it is provable that (x_{2n-1}), (x_{2n}) are independent Kollektivs. To calculate the distribution in ξ (e.g. the probability of $<1,1>$), we may proceed as follows: single out those odd n for which $x_n = 1$; this operation gives us a sequence f: $\omega \rightarrow \omega$ such that $x_{2f(k)-1} = 1$. For this particular f, consider $(x_{2f(k)})_k$. This sequence can be thought of as being chosen from x by the following admissible place selection: x_n is chosen if n is even and $x_{n-1} = 1$. Hence this sequence is a Kollektiv with distribution $(\frac{1}{2},\frac{1}{2})$. The computation is now a matter of bookkeeping:

if we put y $= (x_{2n-1})$, z is (x_{2n}), $Y(m) = \sum_{n=1}^{m} y_n$, then $(x_{2f(k)}) = {}^{z}/y$ and we may write

$$\frac{1}{m}\sum_{n=1}^{m} 1_{<1,1>}(<x_n,y_n>) = \frac{1}{m}Y(m)\cdot\frac{1}{Y(m)}\sum_{k=1}^{Y(m)} ({}^{z}/y)_k;$$

and the desired value $\frac{1}{4}$ is obtained by taking limits.
In the same way one proves that ξ is a Kollektiv. Let Φ be an admissible place selection operating on ξ. Φ determines an admissibly chosen subsequence $(x_{2g(i)-1})_i$, for some sequence g: $\omega \rightarrow \omega$; and also an admissibly chosen subsequence of (x_{2n}), the latter determined by the procedure: choose those n such that n = 2g(i) and $x_{2g(i)-1} = 1$. The computation now proceeds as above, with the sequences just defined replacing (x_{2n-1}) and $(x_{2f(k)})$.

We thus see that in von Mises' theory the product rule is part of the *meaning* of probability; it is provable from the properties of Kollektivs that the probability of the outcome of two tosses in succession is obtained by multiplying the probabilities of the single outcomes. The same holds for the probabilities for the outcomes of n tosses in succession.
A single Kollektiv x in 2^{ω} thus induces a product probability distribution on the binary words of length n, for each n. This example therefore illustrates the claim made earlier, that place selections are intended to capture the independence of successive tosses.

2. Now consider two tosses with two fair coins, supposed to be independent (in some physical sense). In this case we also expect the product rule to hold. But now its validity must be assumed; there is no way to deduce it from the theory. To be specific, if x and y are Kollektivs representing the two coins, we must assume that x and y are independent in the sense of the

condition given in 2.4.1.1.4. Once this assumption is made, a simple computation, exactly as in the previous example, shows that the probabilities of the outcomes <0,0>, <0,1>, <1,0> and <1,1> are given by the productrule:

if we put $Y(m) = \sum_{n=1}^{m} y_n$, then $\frac{1}{m}\sum_{n=1}^{m} 1_{<1,1>}(<x_n, y_n>) = \frac{1}{m}Y(m) \cdot \frac{1}{Y(m)}\sum_{k=1}^{Y(m)} (x/y)_k$;
since y is a Kollektiv with respect to $(\frac{1}{2}, \frac{1}{2})$, $\lim_{n\to\infty} \frac{1}{m} Y(m) = \frac{1}{2}$, so after taking limits the left hand side equals $\frac{1}{4}$.

The case of n independent coins (possibly with different distributions) is treated similarly; but note that we now need n independent Kollektivs to induce a product probability distribution on the binary words of length n, whereas in the previous example one Kollektiv sufficed for all binary words.

2.4.1.3 Comparison At this point, having seen some of the differences between von Mises' theory and that of Kolmogorov, the reader may well wonder how the results of the two theories are related. The answer is somewhat intricate. Recall the different definitions of probability in the two theories: Kolmogorov provides an implicit definition of probability as a positive measure with norm one, while in von Mises' theory, probability is basically a measure *together with* a Kollektiv which induces that measure.

It is clear from this description that not necessarily every theorem of the form "the probability of such-and-such is so-and-so" derived from the Kolmogorov axioms is derivable in von Mises' theory, *for the latter's interpretation of probability*.
In fact, we shall see in 2.6 that the law of the iterated logarithm, *when stated in this form*, is a counterexample. Roughly speaking, we may say that von Mises' theory can reproduce that part of Kolmogorov's theory (with the probability distribution interpreted in a Kollektiv), which makes no essential use of the σ-additivity of the measure. The first volume of Feller' treatise [25] gives a fair sample of problems which fall in this category, as do, of course, von Mises' own technical works on probability theory, *Wahrscheinlichkeitsrechnung* [68] and *Mathematical theory of probability and statistics* [74] (this is not to say that the books mentioned contain all that can be derived in von Mises' theory).
That part of Kolmogorov's theory which *does* use σ-additivity essentially, can be derived in von Mises' theory purely conventionally, as a statement concerning *measure*, which in some cases, but not in all, can also be interpreted as a statement concerning *probability*. The strong limit laws belong to this category, when stated in their usual form:

"The measure of the following set of infinite sequences {..|.....} is 1".

Nevertheless, as readers of Feller's [25] well know, the strong limit laws can also be stated in terms of finite sequences and in that form they *are* derivable in von Mises' theory. This holds for the finite version of the strong law of large numbers, briefly considered in 2.4.3, as well as for the finite version of law of the iterated logarithm (for which see Kolmogorov [45] and 2.4.3).

2.4.2 Necessity of Kollektivs A natural question suggested by the existence of these two different formalisms for probability theory is: Which formalism is to be preferred? The course of history has already provided some sort of an answer: no one uses von Mises' formalism anymore. Apparently, we must conclude from this fact that Kollektivs have no relevance for probability theory as such. They may perhaps be studied for their own sake, in some far-out corner of mathematics; but, to use Poincaré's famous distinction, as a problem that one poses, not as a problem that poses itself. This conclusion, however, is mistaken.

Indeed, it is quite trivial to show that anyone who interprets probability as relative frequency and accepts the Kolmogorov axioms plus the product rule for (physically) independent events, also has to believe in Kollektivs. (If the sample space has cardinality greater than 2, the rule for conditional probabilities must be added to this list.)

In practice, we have to operate with relative frequencies in finite sequences, so strictly speaking one can't deduce the existence of infinite Kollektivs. However, for simplicity we shall assume that probability is interpreted as limiting relative frequency, in which case the existence of infinite Kollektivs *can* be deduced. With suitable approximations the argument works as well for finite sequences. (In fact, Kolmogorov's later conviction that his axioms needed to be supplemented by a precise form of the frequency interpretation, led him to the first satisfactory definition of randomness for *finite* sequences; see 5.2)

We shall now give the argument, which consists essentially only in inverting the examples in 2.4.1.2.

Referring to the first example, we claim the following. Consider an infinite sequence of tosses with a fair coin; if the probability of heads is identified with its limiting relative frequency in the sequence (in this case $\frac{1}{2}$), and if this probability satisfies the usual rules plus the product rule for two consecutive tosses, then the sequence must be invariant under the place selections which occur in the proof of the product rule.

To prove the claim, recall that three place selections occuured in example 1: if x denotes a sequences of tosses with a fair coin, we select from x

(i) x_m with m odd

(ii) x_m with m even

(iii) x_m with m even and $x_{m-1} = 1$.

We show that in each of the selected subsequences, the limiting relative frequency of 1 is $\frac{1}{2}$. We assume the frequency interpretation and the product rule: the probability of each of the outcomes <0,0>, <0,1>, <1,0>, <1,1> in $(<x_n,y_n>)_n$ is $\frac{1}{4}$. The computation goes as follows.

$$\text{(i)} \lim_{m\to\infty} \frac{1}{m}\sum_{n=1}^{m} x_{2n-1} = \lim_{m\to\infty} \frac{1}{m}\sum_{n=1}^{m} 1_{<1,0>}(<x_{2n-1},x_{2n}>) + \lim_{m\to\infty} \frac{1}{m}\sum_{n=1}^{m} 1_{<1,1>}(<x_{2n-1},x_{2n}>)$$

$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

(ii) is treated analogously.

(iii) If we put $y = (x_{2n-1})$, $z = (x_{2n})$, $Y(m) = \sum_{n=1}^{m} y_n$, then the selected subsequence can be written z/y and we have

$$\lim_{n\to\infty} \frac{1}{Y(m)} \sum_{k=1}^{m} (z/y)_k = \frac{\lim_{m\to\infty} \frac{1}{m}\sum_{n=1}^{m} 1_{<1,1>}(<z_n,y_n>)}{\lim_{m\to\infty} \frac{1}{m} Y(m)};$$

by (i), $\lim_{n\to\infty} \frac{1}{m} Y(m) = \frac{1}{2}$, so the right hand side equals $\frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$.

The same trivial argument can be applied to the second example, to show that the sequence x of outcomes of the tosses with the first coin must be invariant under a place selection defined by the second coin: choose those x_n for which $y_n = 1$ (for example).

Summarizing: interpreting probability as limiting relative frequency and applying the deductions of probability theory to a sequence x entails assuming that x is a Kollektiv, or at least that it has the Kollektiv-properties required for the particular deduction at hand (and one is tempted to argue: since we could have chosen to perform a different calculation, e.g. that of the probability of n times heads on n consecutive tosses, x must in fact be a Kollektiv, invariant under all admissible place selections).
Part of probability theory is adequately represented by Kolmogorov's axioms, but as soon as it comes to interpreting the results (as results on relative frequency), one necessarily has to consider Kollektivs. And to say precsiely what the frequency interpretation is, one has to give a precise definition of Kollektivs.

2.4.3 Strong limit laws Twice already, strong limit laws were mentioned in connection with von Mises' theory and both times we stressed a negative aspect. In 2.2.3 it was said that the existence of limiting relative frequencies in a Kollektiv cannot be inferred from the strong law of large numbers (which states that these limiting relative frequencies exist in "almost all" sequences). Rather, they were assumed to exist because that is a reasonable idealisation of experience. In 2.4.1.3 we remarked that the law of the iterated logarithm, when stated in its usual form (that is, for infinite sequences), is not derivable in von Mises' system. Given the central role of the strong limit laws in probability theory, it is natural to inquire into their status in von Mises' theory.

Von Mises devoted a chapter of *Wahrscheinlichkeit, Statistik und Wahrheit* [70,129-163] to this problem; and elsewhere in this book, in a description of the contents of "das schöne und sehr lesenswehrte Büchlein von A. Kolmogoroff" [70,124], the *Grundbegriffe*, he indicated in what sense the law of the iterated logarithm is derivable in his system ([70,125]; a passage which has apparently gone unnoticed).

We shall follow von Mises' description of the strong law of large numbers; after that, little need be added to clarify the status of the law of the iterated logarithm.

Let x be a Kollektiv in 2^ω with respect to distribution (1–p,p). Fix $n,m \in \omega$ with $m < n$ and let $\varepsilon \in (0,1)$. From x a Kollektiv y in $(2^n)^\omega$ is derived as in example 1 of 2.4.1.2: y is a *combination* (in the sense of the fourth operation discussed in 2.4.1.1) of the n Kollektivs $(x_{kn+i})_k$ for $1 \le i \le n$. As in the example, one shows that y is a Kollektiv with respect to the product distribution on the binary words of length n. From y we derive by *mixing* a Kollektiv z in 2^ω as follows (recall that each y_j is an n–tuple):

$$z_j = \begin{cases} 1 & \text{if for some k, } m \le k \le n, \ \left|\frac{1}{k}\sum_{i=1}^{k}(y_j)_i - p\right| > \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

As von Mises presents it, the strong law of large numbers then says that the limiting frequency of 1 in z, i.e. of the event:

$$\exists k(m \le k \le n \ \& \ \left|\frac{1}{k}\sum_{i=1}^{k}(y_j)_i - p\right| > \varepsilon) \text{ in y,}$$

is less than $\varepsilon^{-2}\cdot m^{-1}$, independent of the values of n and p. What is the relation of this form of the strong law of large numbers to the form stated as Theorem 2.3.3?

Put

$$A_{mn}(\varepsilon) := \{w \in 2^n \mid \forall k \ (m \le k \le n \ \rightarrow \left|\frac{1}{k}\sum_{i=1}^{k} w_i - p\right| \le \varepsilon)\}.$$

Let P_n be the probability distribution on 2^n induced by x (via y). Von Mises' version of the strong law then implies:

$$\forall \varepsilon > 0 \ \forall \delta > 0 \ \exists m \ \forall n \ge m \ P_n(A_{mn}(\varepsilon)) > 1\text{–}\delta.$$

Now P_n may *formally* be regarded as the restriction of the measure $\mu_p = (1\text{–}p,p)^\omega$ on 2^ω to 2^n. We may then write equivalently (*we just use a different notation*):

$$\forall\varepsilon > 0 \forall\delta > 0 \exists m \forall n \geq m \ \mu_p\{x \in 2^\omega | \forall k\ (m \leq k \leq n \rightarrow |\frac{1}{k}\sum_{i=1}^{k} x_i - p| \leq \varepsilon)\} > 1-\delta.$$

This statement is, *using the σ-additivity of* μ_p, equivalent to

$$\mu_p\{x \in 2^\omega | \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k = p\} = 1.$$

In other words, the usual version of the strong law can be derived from the version acceptable to von Mises if we take the collection of *probability distributions* (P_n), induced by the Kollektiv $x \in 2^\omega$, to define a single σ-additive *measure* μ_p on 2^ω. From the standpoint of von Mises, however, the extension of the collection (P_n) to μ_p is a purely conventional matter, bereft of probabilistic significance.

It is perhaps not superfluous to recall that Kolmogorov was of the same opinion; in fact, von Mises credits his presentation of the strong law to Kolmogorov [45], the paper which contains the general form of the law of the iterated logarithm for independent random variables. In this article, Kolmogorov emphatically states that the only meaningful form of the law pertains to *finite* sequences. Since we do not at present need the result in full generality, we state it for i.i.d. two-valued random variables. Modulo this simplification, Kolmogorov's version reads as follows:

(a) $\forall\varepsilon > 0 \forall\delta > 0 \exists m \forall n \geq m \ \mu_p\{x \in 2^\omega | \exists k: m \leq k \leq n \ \& \sum_{j=1}^{k} x_j > (1+\delta)\sqrt{2k\cdot p(1-p)\log\log n}\ \} < \varepsilon$

(b) $\forall\varepsilon > 0 \forall\delta > 0 \exists m \forall n \geq m \ \mu_p\{x \in 2^\omega | \forall k: m \leq k \leq n \ \& \sum_{j=1}^{k} x_j < (1-\delta)\sqrt{2k\cdot p(1-p)\log\log n}\ \} < \varepsilon;$

with analogous conditions for the lower bound on the relative frequency. (For notational convenience we have used μ_p instead of the P_n; but it will be clear that we refer in fact to finite sequences only.)

As with the strong law of large numbers, *this* form is derivable from von Mises' axioms, the extension to the version for infinite sequences then being purely conventional. Ville's theorem, discussed in 2.6 and improved upon in 4.6, will in fact show that there is no straightforward frequency interpretation for the infinite version.

In conclusion we emphasize again that, for von Mises, the limiting relative frequencies in a Kollektiv do not owe their existence to the strong law of large numbers. Rather, it is the other way around, as the above derivation should have made clear: only because our x satisfies the two conditions on Kollektivs, it allows us to deduce the strong law, as a statement on the relative frequency of a particular event.

2.5 Making Kollektivs respectable: 1919 – 1940 For a while, from 1919 to 1933, the only explicit, more or less rigorous, axiomatisation of probability theory (von Mises') made use of Kollektivs, hence the imperative need to make this objects mathematically acceptable. Two principal lines of attack can be distinguished.

1. Restricting *a priori* the class of admissible place selections and trying to construct explicitly a Kollektiv with respect to the class so obtained (Reichenbach, Popper, Copeland; 2.5.1);
2. Showing that von Mises' theory is consistent *in context*, that is, showing that in each specific application we may assume the existence of a Kollektiv with respect to the place selections required for the application (von Mises, Wald; 2.5.2).

After the appearance of Kolmogorov's *Grundbegriffe* in 1933, and especially after the Geneva conference in 1937, at which strict frequentists and the proponents of an implicit definition of probability came into head-on collision (see 2.6), attempts to define Kollektivs petered out, with Church's [16] (1940) as a notable exception. Only in 1963, with the publication of Kolmogorov's [47], hostilities were resumed. We now discuss attempts 1. and 2.; for simplicity, we consider Kollektivs in 2^{ω} only.

2.5.1 Lawlike selections Common to all attempts which fall under the heading 1. above, is the conviction that "admissible place selection" should mean "place selection given by a mathematical law", as in the first two examples illustrating the definition of admissible place selection (choose the n^{th} term if n is prime; choose the n^{th} term if it is preceded by 10 1's). We comment on this interpretation later, but let us first consider some representative examples of this approach.
Various authors (e.g. Popper [83], Reichenbach [85], Copeland [17]) independently arrived at a class of place selections which is a generalisation of the second example (2.3.1.2(b)): the so-called *Bernoulli selections*. They can be described as follows: let x be a Kollektiv; fix a binary word w and choose all x_n such that w is a final segment (or *suffix*) of x(n–1).
Note that this selection chooses an infinite subsequence of x if x contains infinitely many occurrences of w (which is for instance the case if x is a Kollektiv with respect to (1–p,p), for $0 < p < 1$).
We henceforth treat place selections as partial functions $\Phi: 2^{\omega} \to 2^{\omega}$, where Φx is the infinite subsequence selected from x by Φ. This identification is not unproblematic. It has the technical disadvantage that it does not keep track of where the n^{th} selected term occurred in the original sequence. Its main philosophical disadvantage is, that it is most appropriate for place selections which are judged admissible on a priori grounds. It is considerably less so for

place selections which are admissible for a given Kollektiv, the general case of admissibility (cf. the discussion in 2.3.3). Since we are concerned in this section with place selections which are, for various reasons, judged admissible on a priori grounds, the identification is harmless here.

The *domain* of a place selection Φ will be the set of those x such that Φ operating on x produces an infinite subsequence of x. Intuitively, a place selection Φ is completely determined by a function $\phi: 2^{<\omega} \to \{0,1\}$, when we interpret the statement "$\phi(w) = 1$" as: choose the $|w|+1^{th}$ term, and "$\phi(w) = 0$" as: skip the $|w|+1^{th}$ term. To bridge the gap between ϕ and Φ it is convenient to use a place selection Φ' which operates on finite sequences. We formalize these remarks in the following definition; we first introduce a general definition of place selection, and then specialize to Bernoulli selections, as introduced informally above.

2.5.1.1 Definition Let $\phi: 2^{<\omega} \to \{0,1\}$ be any function. ϕ determines a place selection Φ in two steps:

(i) $\Phi' : 2^{<\omega} \to 2^{<\omega}$ is given by $\Phi'(uj) = \begin{cases} \Phi'(u)j & \text{if } \phi(u) = 1 \\ \Phi'(u) & \text{if } \phi(u) = 0 \end{cases}$ where $j \in \{0,1\}$

(ii) a partial function $\Phi: 2^{\omega} \to 2^{\omega}$ is defined by

(a) $\mathrm{dom}\,\Phi = \{x \in 2^{\omega} | \forall n \exists k \geq n\ \phi(x(k)) = 1\}$

(b) $x \in \mathrm{dom}\,\Phi$ implies $\Phi(x) = \bigcap_n [\Phi'(x(n))]$

2.5.1.2 Definition Let $w \in 2^{<\omega}$ and $\phi_w: 2^{<\omega} \to \{0,1\}$ defined by

$$\phi_w(u) = \begin{cases} 1 & \text{if w is a final segment of u} \\ 0 & \text{otherwise} \end{cases}$$

$\Phi_w: 2^{\omega} \to 2^{\omega}$ is a *Bernoulli selection* if it results from ϕ_w by application of (i) and (ii) of 2.5.1.1.

Recall that for $p \in [0,1]$, the set LLN(p) was defined as (2.2.3.2):

$$\mathrm{LLN}(p) = \{x \in 2^{\omega} | \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k = p\}.$$

2.5.1.3 Definition Let $p \in [0,1]$. $x \in 2^{\omega}$ is called a *Bernoulli sequence* with parameter p (notation: $x \in B(p)$) if for all w: $x \in \mathrm{dom}\,\Phi_w$ implies $\Phi_w(x) \in \mathrm{LLN}(p)$.

It is not difficult to show that, if x is a Bernoulli sequence with parameter p, for each word w the limiting relative frequency of w in x equals $\mu_p[w]$.

2.5.1.4 Lemma Let $p \in [0,1]$. Then

$$x \in B(p) \text{ iff } \forall w \in 2^{<\omega}: \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_{[w]}(T^k x) = \mu_p[w],$$

where $T: 2^\omega \to 2^\omega$ is the left shift and $\mu_p = (1-p,p)^\omega$.

Proof See, e.g., Schnorr [88,22]. ∎

Remark The preceding lemma has as a consequence that, at least for $1 > p > 0$, $x \in B(p)$ implies for all words w: $x \in \text{dom}\,\Phi_w$. The "implies" in definition 2.5.1.3 might therefore have been replaced by "and".

In the special case $p = \frac{1}{2}$, Bernoulli sequences are commonly called, *normal numbers*. Now, although Kollektivs were not supposed to be constructible (cf. proposition 2.3.2.3), Bernoulli sequences can be constructed explicitly. E.g.

2.5.1.5 Lemma There exists a recursive normal number.

Proof (Champernowne [15]) Let x = 0100011011000........, i.e. the set of all finite binary words written in lexicographic order. For the construction of Bernoulli sequences for arbitrary p, see von Mises [69]. ∎

In one sense, normal numbers and, more generally, Bernoulli sequences, are clearly not satisfactory models of Kollektivs, if only because problems involving two coins (say), cannot be treated in the way von Mises intended[4]. On the other hand, the beautiful work of Kamae [40], which is described in 5.6, shows that there are really many more place selections Φ such that $x \in B(p)$ implies $\Phi x \in B(p)$; in fact an uncountable set and what's more, with an appealing physical description.

Bernoulli selections are examples of lawlike selections, but by no means the only ones; e.g. our second example: choose x_n if n is prime, is not of this form. The apparently most general characterisation of lawlike place selections is due to Church [16] (the article is from 1940, a time when von Mises' theory was no longer a hot issue).

2.5.1.6 Definition A function $\Phi: 2^\omega \to 2^\omega$ is called a *recursive place selection* if it is generated by a total recursive $\phi: 2^{<\omega} \to \{0,1\}$ according to (i) and (ii) of 2.5.1.1.

2.5.1.7 Definition Let $p \in [0,1]$. x is *Church random* with parameter p (notation: $x \in C(p)$) if for all recursive place selections Φ: $x \in \text{dom}\,\Phi$ implies $\Phi x \in LLN(p)$.

Remark Unlike the situation for Bernoulli sequences, in this case the "implies" cannot be replaced by "and". In other words, while

$$B(p) \subseteq \bigcap_w \text{dom}\,\Phi_w$$

we do not have

$$C(p) \subseteq \bigcap \{\text{dom}\,\Phi \mid \Phi \text{ recursive}\}.$$

In 2.6.2 we shall meet an example of a place selection Φ such that $C(p) \not\subset \text{dom}\,\Phi$. This observation implies that, with the above definition of Church randomness, some Bernoulli sequences are Church random for fairly trivial reasons. Note that, from the point of view of von Mises' theory, it would be natural to require that a Kollektiv belongs to the domain of the place selections needed to solve a particular problem, since the theory consists essentially of transformations of (infinite) Kollektivs into (infinite) Kollektivs. Also, the wording of the definition of Kollektivs (2.3.1.1; originally [67,57]) suggests that it is *assumed* that admissible place selections select infinite subsequences. However, it is customary in the literature to use the implication in 2.5.1.7 (see e.g. Schnorr [88,22]) and for good reason, since there exist (recursive!) place selections with disjoint domains.

We now discuss the merits of the identification of "admissible place selection" with "lawlike place selection".

1. It is an illusion to suppose that one can restrict oneself to the existence of lawlike place selections only. As the paragraph on combination in 2.4.1.1 shows, a lawlike selection on $(\langle x_n, y_n \rangle)$ factors as a lawlike selection on y and a *random* selection on x. Hence, by the argument given in 2.4.2, it follows that an application of the theory, even to such a simple problem as that of the probability of two coins coming up heads, assumes that x and y satisfy stronger properties of randomness than just being Church random. And if it is maintained that the admissibility of lawlike place selections can be recognized a priori, this has a consequence that the admissibility of the above random selection on y is also an priori; a consequence which should perhaps instill some caution in the use of the a priori in this context.

2. The recursive analogue of proposition 2.3.2.2:

An admissibly chosen subsequence of a Kollektiv is again a Kollektiv, with the same distribution, is

If $x \in C(p)$, *then for every recursive place selection* Φ, $x \in \text{dom}\,\Phi$ *implies* $\Phi x \in C(p)$.

If the admissible place selections were identified with the recursive place selections, C(p) would be the set of Kollektivs with distribution C(p); so if $x \in C(p)$, we have by the above analogue of 2.3.2.2 at least countably many subsequences of x which are also Kollektivs with respect to (1–p,p). Now it seems very implausible that, for a satisfactory definition of Kollektivs, *only* countably many subsequences of a Kollektiv are themselves a Kollektiv (with the same distribution).

On the contrary, we shall prove the following *principle of homogeneity*, which can be read as a quantitative version of proposition 2.3.2.2:

If x *is a Kollektiv with respect to* (1–p,p), *so is almost every subsequence of* x.

To turn this rather vague principle into a precise mathematical statement requires some effort; this will be done in Chapters 3 and 4 and involves, perhaps somewhat surprisingly, a study of modern definitions of randomness. But to give the reader already at this stage an impression of the formal version of the principle, we state it in semi-formal terms (where / denotes the operation of *sampling* intoduced in 2.4.1.1):

If x *is a Kollektiv with respect to distribution* (1–p,p), *then* $\mu_p\{y|\ x/y$ Kollektiv with respect to $(1–p.p)\} = 1$.

Already from this form of the principle, which is considerably weaker that the version that will be proved in 4.5, it is clear that the content of proposition 2.3.2.2 is not likely to be exhausted by its recursive analogue stated above. In other words, the principle of homogeneity, which is in itself a purely quantitative statement not mentioning admissibility, suggests that there are many more admissible place selections than just those which are recursive.

3. The recursive place selections owe their appeal to the circumstance that they are a priori admissible. But there might be many more such selections, even disregarding possible wider interpretations of the term "lawlike". We shall not consider these wider interpretations (such as hyperarithmetical, constructible), since, although the admissibility of selections thus defined is a priori, the *truth* of the axiom that Kollektivs are invariant under these admissible place selections is by no means a priori; and our experience with constructible, non-recursive, place selections is restricted, to say the least. In fact, one might also argue that the class of recursive place selections is already much too large.

Physical processes are a possible source of *a priori* selections, that is to say, if these processes are in some sense physically independent of the process which generates the Kollektiv from which is to be selected. Another source is the human mind (but perhaps this example can be subsumed under the previous one): a choice sequence seems no less an admissible place selection than e.g., the sequence of primes (at least if the mind generating the sequence has no prognostic or telepathic abilities). The trouble with these examples is, that they do not lead to a well defined class of admissible place selections, considered as functions on the infinite binary sequences. If we select from the Kollektiv produced by a coin using the outcomes of the tosses of a second coin, all we can say a priori is that the second coin will produce a sequence in 2^{ω}.

Of course we trust that it will produce a sequence which is independent of the first sequence and hence an admissible selection for that sequence. But to describe this situation, we must widen our framework and consider, not only a priori admissibility, but also admissibility with respect to a given Kollektiv, in conformity with the intensional character of admissibility mentioned in 2.3.3.

However, there exist situations in which the old framework (i.e. admissibility as a priori property) suffices and which nevertheless give rise to continuously many admissible place selections: the special case of independence discussed under the name *disjointness* by Furstenberg [30], is a case in point. The place selections obtained in this way are defined in 5.6.

4. The remarks in 3. point toward a general conclusion: lawlikeness is not as fundamental as may seem at first sight. What *is* fundamental is a relation of physical independence between the process generating the Kollektiv and the process determining the selection. A lawlike selection rule is (as far as we know!) indeed independent of coin tossing in this sense; but there are many other such selection procedures. The physical roots of probability theory, emphasized by von Mises, are obscured rather than illuminated by Church' definition.

2.5.2 The contextual solution We have noted already that von Mises' later presentations of the theory differs slightly from the version given in 2.3 (which dates from 1919). The new version is best described as being contextual: in each specific application of the theory it is assumed that the Kollektiv under consideration is invariant under the place selections needed for that application. This assumption of course has to be justified, and in the process of justification notions such as admissibility or independence may come into play; but they do not form part of the theory.

> Die Festsetzung daß in einem Kollektiv jede Stellenauswahl die Grenzhäufigkeit unverändert läßt, besagt nichts anderes als dieses: Wir verabreden daß, wenn in einer konkreten Aufgabe ein Kollektiv einer bestimmten Stellenauswahl unterworfen wird, wir annehmen wollen, diese Stellenauswahl ändere nichts an den Grenzwerten der relativen Häufigkeiten. Nichts darüber hinaus enthält mein Regellosigkeitsaxiom [i.e. 2.3.1.1(ii)].
> Da nun in einer bestimmten Aufgabe niemals "alle" Auswahlen in Frage kommen, sondern deren nur wenige, so das man jedesmal mit einer eingeschränkten, ad hoc zugeschnittenen Regellosigkeit das Auslangen finden könnte, so kann tatsächlich nichts von dem eintreten, was ängstliche Gemüter befürchten [namely, inconsistency] [70,119].

As an instrumentalist position, von Mises' position is no more absurd than, say, the complementarity interpretation of quantum mechanics. But, if taken to be the whole truth, it leads to the same type of objection, known as "counterfactual definiteness": the real, physical Kollektiv does not know which computation we are going to perform; we could have chosen to perform a computation different from the one we in fact performed; hence the real Kollektiv must be invariant under "all" place selections. In other words, although for computational purposes an instrumentalist reading of the randomness axiom, *with its abandonment of a definition of Kollektivs*, suffices, explaining the applicability of probability seems to require more (recall that the older theory had both these aims).

The consistency of the contextual version of the theory was settled by Wald [100]. (Note that von Mises wrote the passage quoted just now before Wald's results became known.)

2.5.2.1 Theorem Let $p \in [0,1]$ and let $\mathcal{H}$ be a countable set of place selections. Put $C(\mathcal{H},p) := \{x \mid \forall\Phi \in \mathcal{H}\ (x \in \text{dom}\ \Phi \to x \in LLN(p))\}$. Then $C(\mathcal{H},p)$ has the cardinality of the continuum.

This theorem provides for the existence of many Bernoulli sequences or Church random sequences; but its applicability is of course not so restricted. Von Mises was perfectly satisfied with this result [75,92], since any specific application of the theory never involves more than countably many place selections.

We now give a proofsketch of a measure theoretic version of the above theorem, a proofsketch which will at the same time illustrate von Mises' stand on the laws of large numbers.

2.5.2.2 Lemma (Doob [20], Feller [24]) Let $p \in (0,1)$ and let $\Phi: 2^\omega \to 2^\omega$ be a place selection. Then for all Borel sets $A \subseteq 2^\omega$: $\mu_p\Phi^{-1}A \leq \mu_p A$. If $\mu_p \text{dom}\ \Phi = 1$, we have equality for all A.

Proof See Schnorr [88,23]. □

As a consequence, we have

2.5.2.3 Theorem Let $p \in (0,1)$ and let $\mathcal{H}$ be a countable set of place selections. Then $\mu_p C(\mathcal{H},p) = 1$.

Proof Let $\Phi \in \mathcal{H}$. Since $\mu_p LLN(p)^c = 0$ (theorem 2.2.3.3) we get $\mu_p\Phi^{-1}LLN(p)^c = 0$, by the preceding lemma. □

The theorem is of course most interesting for those $\mathcal{H}$ which contain only place selections whose domain has full measure (an assumption which is usually made). Note that we have surreptitiously changed the condition "$p \in [0,1]$" in theorem 2.5.2.1 to "$p \in (0,1)$" in 2.5.2.3, for the simple reason that for p = 0,1, the measure μ_p is concentrated at one point. It is possible to give a measure theoretic proof of theorem 2.5.2.1 for the extremal values of p, but in that case one has to use the techniques of 4.6.

The correct interpretation of theorem 2.5.2.3 (from von Mises' point of view) is *not* given by the following quotation from Feller [25,204]:

> Taken in conjunction with our theorem on the impossibility of gambling systems, the law of large numbers implies the existence of the limit [relative frequency] not only for the original sequence of trials but also for all subsequences obtained in accordance with the rules of selection [i.e. admissible place selections]. Thus the two theorems together describe the fundamental properties of randomness which are inherent in the intuitive notion of probability and whose importance was stressed with special emphasis by von Mises.

Feller's remark fits in with the propensity interpretation, which allows one to say that theorem 2.5.2.3 *explains* the impossibility of gambling strategies; but, as we know by now, this is not von Mises' interpretation of probability.
For him, theorem 2.5.2.3 has significance as an existence result only, since μ_p is a measure, not a probability distribution (cf. the careful discussion in [74,41-2]). The theorem shows that the concept of Kollektiv is free of contradiction (in context), but does not thereby render superfluous the empirically motivated axioms for Kollektivs.

2.6. The Geneva conference: Fréchet's objections In 1937, the Université de Genève organized a conference on the theory of probability theory, part of which was devoted to foundational problems (the proceedings of this part have been published as [35]). The focal point of the discussion was von Mises' theory, and especially its relation to the newly published axiomatisation of probability theory by Kolmogorov. 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 [35,23-55]. Von Mises himself was absent, but his rebuttals of the objections were published in the proceedings [35,57-66]. To no avail: the same objections were reiterated in Fréchet's [103]; 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.

2.6.1 Fréchet's philosophical position In view of the persistent controversy between von Mises and his critics, with arguments seemingly having little or no effect, it seems worthwhile to investigate why the participants in the debate had so little common ground for discussion. As stated in 2.1, we shall adopt as working hypothesis that the lack of mutual comprehension is due to widely differing views on the foundations of mathematics as well as on the foundations of probability.

The first difference comes out clearly when Fréchet advances the usual "proof of inconsistency" against von Mises. Although the argument itself is identical to that of Kamke reported in 2.3.3, it is worth quoting since it shows the extent of the mutual incomprehension.

> Or la deuxième condition [i.e. 2.3.1.1(ii)] n'imposait aucune limitation au choix de la selection des épreuves après laquelle la fréquence totale devait garder la même valeur. On pouvait donc conclure: ou bien qu'en faisant intervenir la totalité des selections imaginables, elle faisait intervenir un ensemble sans signification concrète precise, ou bien que si l'on considère cet ensemble de selections comme bien défini, il contient la selection S_1 qui retient seulement la suite des épreuves ou l'évenement considéré E s'est produit – ou aura lieu – et la selection S_2 qui ne retient que les autres. L'une au moins de ces suites partielles est infini; si c'est S_1, la fréquence totale de E y est égale à 1; si c'est S_2, elle est égale à zéro. Il n'existe donc pas de collectif où la probabilité d'un évenement soit supérieure à zéro et inférieur à l'unité. Cette observation évidente ayant été faite depuis longtemps de diverses côtés, il nous est difficile de comprendre ce qu'entend M. de Misès, en écrivant que jamais on n'a pu signaler un cas concret de contradictions qui pourraient se produire dans l'application de la notion de collectif [28,29-30].

Cette observation évidente...... it is astonishing to see that Fréchet has not grasped any of the subtle properties of Kollektivs: the intensional character of admissible place selections and the fact that Kollektivs have to be considered as new mathematical objects, so that the above selections S_1 and S_2 cannot be elements of a collection of place selections "bien défini".

Like Kamke, Fréchet reveals himself in this passage as one who believes that all mathematical objects are equally accessible; a view clearly not shared by von Mises (cf. his comparison of Kollektivs with choice sequences)[5].

So far, we have been concerned with different viewpoints on the foundations of mathematics. We now turn to the foundations of probability. We shall assume as working hypothesis that Fréchet is an adherent of the propensity interpretation. This hypothesis will explain at least in part why Fréchet thought that Ville' theorem dealt such a devastating blow to von Mises program. But part of Fréchet's conviction also results from plain confusion.

We shall now compile some passages from Fréchet [28,45-7] 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 mésurons 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 évenement fortuit dans un

grand nombre d'épreuves se comporte comme la mesure d'une constante physique attachée à cette évenement dans une certaine catégorie d'épreuves, constante qu'on peut appeler probabilité, on en deduit, par des raisonnements dont la rigueur n'est pas absolue, les lois des probabilités totales et composées et on verifie 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, lesquelles ne donnent pas – contrairement à ceux de M. de Misès –, 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 géneral de la façon suivante: on adopte comme mesures expérimentales de certaines probabilités p, p',.... les fréquences f, f',..., correspondantes dans les groupes dépreuves nombreuses. Certains théorèmes de la théorie déductive établissent les expressions de certaines autres probabilités, P,P',..., en fonction de p, p',... . Ayant calculé P,P',... au moyen de ces expressions où l'on a remplacé approximativement p,p',.. par f, f',.., la vérification consistera à s'assurer que les valeurs approchées ainsi obtenus pour P, P',... sont aussi approchées des fréquences F, F',... qui sont les mesures expérimentales directes de P, P',...

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 évenement de probabilité p, diffère de p de plus de ε, alors d'après le théorème de Bernoulli, P_n converge vers zéro avec 1/n. Si donc on se content de vérifier expérimentalement qu'un évenement de probabilité assez petite est pratiquement très rare et même qu'une évenement 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 ε, 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 long 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 2.2.3 (although Fréchet seems to be much less aware of his assumptions than e.g., Popper!). It is evident from [28] and [103] that Fréchet considers the propensity interpretation to be much simpler than the strict frequency interpretation. Superficially, this is indeed so: much of that which 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" [28,45]. 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.

But, although the outward appearance of the propensity interpretation is indeed simple, it is so only because it takes so much for granted. The rules of probability theory are valid for certain phenomena because these phenomena are Kollektivs (2.4.2) and Fréchet's use of the weak law supposes either a large amount of randomness (2.4.3) or some highly theoretical assumption 2.2.3; but even then...). Pragmatic solutions indeed look simple, but a pragmatic attitude does not contribute much toward an understanding of foundations.

2.6.2 Formal objections Above we considered Fréchet's methodological objections. We now discuss the objections which concern the formal structure of von Mises' theory.

2.6.2.1 Inconsistency Since Fréchet, as we have seen, advances the same "proof of inconsistency" as the one discussed at length in 2.3.3, we need not dwell upon it here. Let us recall only that this objection eventually led Wald to prove the consistency of von Mises' theory in context, on the assumption that each specific computation employs at most countably many place selections.

Fréchet objects that the revision by Wald causes the theory to lose much of its primordial simplicity and elegance. It is hard to make sense of this objection, since Wald's theorem is *metamathematical* in character and shows only that the ordinary deductions can be performed without fear of contradiction. The deductions themselves are in no way affected by the consistency proof.

A really forceful objection, which brings out clearly the underlying difference in the interpretation of probability, is provided by:

2.6.2.2 Ville's construction To understand this objection, we have to go back to the law of the iterated logarithm. In 2.4.3 we stated this law for finite sequences. This time, we state it for infinite sequences, since this is the form used in Fréchet's objection.

Law of the iterated logarithm Let $p \in (0,1)$.

(a) For $\alpha>1$, $\mu_p \{x \in 2^\omega | \exists k \forall n \geq k \, |\sum_{j=1}^{n} x_j - np| < \alpha\sqrt{2p\cdot(1-p)n\log\log n} \} = 1$

(b) For $\alpha<1$, $\mu_p \{x \in 2^\omega | \forall k \exists n \geq k \, (\sum_{j=1}^{n} x_j - np) > \alpha\sqrt{2p\cdot(1-p)n\log\log n} \} = 1$ and

for $\alpha<1$, $\mu_p \{x \in 2^\omega | \forall k \exists n \geq k \, (np - \sum_{j=1}^{n} x_j) > \alpha\sqrt{2p\cdot(1-p)n\log\log n} \} = 1$.

Part (b) in particular shows that the quantities

$$\sum_{j=1}^{n} x_j - np \, , \; np - \sum_{j=1}^{n} x_j$$

exhibit fairly large oscillations. This observation provides the starting point for Ville's construction [99,55-69], 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 Φ: $2^\omega \rightarrow 2^\omega$, Ville is able to construct a sequence $x \in 2^\omega$ with the following properties (we assume the identity is in $\mathcal{H}$):

(i) $x \in C(\mathcal{H}, \frac{1}{2})$ (for $C(\mathcal{H},p)$, see definition 2.5.2.1)

(ii) $\forall n \; \frac{1}{n}\sum_{k=1}^{n} x_k \geq \frac{1}{2}$.

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. A very much stronger form of (i) and (ii) will be proven in 4.6.

2. In the second stage of the construction, Ville temporarily adopts von Mises' viewpoint and interprets probability measures on 2^ω as in effect being induced by Kollektivs $\xi \in (2^\omega)^\omega$; so that $\mu_{\frac{1}{2}}A = 1$ must mean:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_A(\xi_\kappa) = 1$$

So far we have considered only Kollektivs in 2^ω; in particular, we have not defined what place selections Ψ: $(2^\omega)^\omega \rightarrow (2^\omega)^\omega$ 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 [99,63-67]). Now put

$$A := \left\{ x \in 2^\omega | \; \forall n \exists k \geq n \left(np - \sum_{j=1}^{n} x_j\right) > \tfrac{1}{2}\sqrt{\tfrac{1}{2} n \log\log n} \right\}.$$

Then Ville shows the following, using 1. :

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

(iii) ξ induces $\mu_{\frac{1}{2}}$ and is a Kollektiv with respect to $\mathcal{H}$

(iv) for A as defined above, $\lim_{n\to\infty} \frac{1}{m}\sum_{j=1}^{m} 1_A(\xi_\kappa) = 0$.

Remark The reader may well wonder what "induces" in (iii) means in view of (iv), since we defined "ξ induces P" to mean:

for all $B \subseteq 2^\omega$, $P(B) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_B(\xi_k)$;

but since $P(A) = 0$ (by (iv)), the induced measure P cannot be equal to $\mu_{\frac{1}{2}}$ as claimed by (iii). Therefore (iii) should be understood as follows. A σ-additive measure on 2^ω is determined completely by its values on the cylinders [w], for finite binary words w; and we do have for the ξ constructed by Ville:

$$\text{for all w,} \quad 2^{-|w|} = \mu_{\frac{1}{2}}[w] = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_{[w]}(\xi_k).$$

Ville's construction is thus a very interesting case of the phenomenon that relative frequency is not a σ-additive measure; since if P were σ-additive, it would coincide with $\mu_{\frac{1}{2}}$.

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 so called *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.

For those who are accustomed to see Ville's construction as the deathblow to the theory of Kollektivs, its cavalier dismissal by von Mises may come as a surprise: "J'accepte ce théorème, mais je n'y vois pas une objection" [72,66]. In fact, von Mises to some extent anticipated Ville's construction in his discussion of the meaning of probability zero [70,38]. As we have seen (in 2.2.3), 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 must now try to understand why von Mises could remain unmoved, when apparently the foundations of his work lay shattered. We believe that objections (a) and (b) are either untenable or based on an interpretation of probability which was not his. Objection (c) is justified, but of no consequence. Before we go deeper into the objections, however, we discuss in more detail the formal structure of Ville's argument.

We simplify a suggestion of Wald[6] to show that Ville's theorem is appreciably less general than may seem at first sight. Consider a countable set $\mathcal{H}$ of place selections, as in 1. Obviously (i) would be *trivially* true if the x constructed did not belong to the domains of the place selections contained in $\mathcal{H}$; and the construction would seem to be less interesting in that case. Unfortunately, such cases do occur. For we may define a countable set $\mathcal{H}$ of recursive place selections as follows:

$\mathcal{H} := \{\Phi^-_\alpha \mid \alpha \in (0,1)\cap\mathbb{Q}\} \cup \{\Phi^+_\alpha \mid \alpha \in (0,1)\cap\mathbb{Q}\}$; Φ^-_α (Φ^+_α) is generated by ϕ^-_α (ϕ^+_α) as in definition 2.5.1.1; ϕ^-_α is determined by $\phi^-_\alpha(x(n)) = 1$ iff $\left(\frac{n}{2} - \sum_{j=1}^{n} x_j\right) > \alpha\sqrt{\tfrac{1}{2}n\log\log n}$ and similarly $\phi^+_\alpha(x(n)) = 1$ iff $\left(\sum_{j=1}^{n} x_j - \frac{n}{2}\right) > \alpha\sqrt{\tfrac{1}{2}n\log\log n}$.

Obviously

$$x \in \operatorname{dom} \Phi^-_\alpha \text{ iff } \forall k \exists n \geq k \left(\frac{n}{2} - \sum_{j=1}^{n} x_j\right) > \alpha\sqrt{\tfrac{1}{2}n\log\log n}$$

and similarly for the Φ^+_α.

Hence, if a sequence x belongs to the domains of the place selections in $\mathcal{H}$, it must exhibit the oscillations prescribed by the law of the iterated logarithm. This means that, when Ville's construction is applied to the set of recursive place selections (say), the constructed sequence x is Church–random partly for trivial reasons. An analogous statement holds for the strengthened form of Ville's theorem proved in Chapter 4. It is then of interest to ask to which countable sets of place selections Ville's construction can be applied non-trivially. The advantage of the measure theoretic proof given in Chapter 4 is, that a furnishes a characterisation of sets of place selections to which the construction is non-trivially applicable:

Ville's theorem applies non-trivially to a collection of place selections if for each Φ *in the collection and for each product measure* $\mu = \Pi_n(1-p_n,p_n)$ *such that* p_n *converges to* $\tfrac{1}{2}$, $\mu(\operatorname{dom} \Phi) = 1$.

The Φ_w satisfy this condition, but the Φ_α don't. Roughly speaking, the theorem applies to place selections which do not have too much "memory".

These considerations show that Ville's theorem is somewhat restricted in scope. One might even go further and argue that sequences such as constructed by Ville are not Kollektivs at all, even on von Mises' definition; for this it suffices to replace the "implies" in definition 2.5.2.1 by "and". When we discussed this question in 2.5, we remarked that von Mises' *use* of Kollektivs seemed to make such a convention natural: Kollektivs are useful in a particular calculation only if the place selections needed for that application select an infinite subsequence from the Kollektiv. On the other hand, in a Church–style definition of randomness it is clearly impossible to demand that a random sequence belong to the domain of *all* recursive place selections: just consider place selections based on the law of the iterated logarithm for $p \neq \tfrac{1}{2}$. Fortunately we need not consider the merits of such a modification of the definition of randomness in detail, since there are weightier arguments which show that the above objections are unjustified. So let us state the import of Ville's theorem in the following way: place selections with "limited memory" do not enforce satisfaction of the law of the iterated logarithm. We now investigate the consequences of this result upon von Mises' theory.

Objection (a) is easiest to dispose of; in fact we have done so already in 2.4.3, 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.
But the latter statement does not mean that von Mises is not able to derive the law as stated in 2.6.2.2, only that this theorem does not have a frequency interpretation (in the space of *infinite* binary sequences).

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 σ-additivity:

> Wenn man die Mengen (Ereignisse) A aus $\mathbb{E}$ [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($\mathbb{E}$) [the σ-algebra generated by $\mathbb{E}$] eine solche Deutung als reelle beobachtbare Erscheinungen vernünftiger Weise gestatten. Es kann also vorkommen, daß das Wahrscheinlichkeitsfeld ($\mathbb{E}$,P) als ein (vielleicht idealisiertes) Bild reeller zufälliger Erscheinungen betrachtet werden kann, während das erweiterte Wahrscheinlichkeitsfeld (B($\mathbb{E}$),P) eine reine mathematische Konstruktion ist [44,16].

Objection (b) raises questions which go to the heart of the foundations of probability. It consists of two parts:
(b_1) Kollektivs are not satisfactory models of random phenomena, since a unilateral approach of the limit will not occur in practice;
(b_2) 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.

Ad (b_1). "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 whatsoever 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. Now objection (b_2) claims that in fact there *does* exist another criterion: satisfaction of asymptotic laws derived by measure theoretic methods. So let us now consider the second part of objection (b).

Ad (b_2). As we have seen in 2.2.3, this objection does not make sense on the strict frequency interpretation of probability, i.e. von Mises' own interpretation. 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 (lemma 2.5.2.2). 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_2) makes sense, although in that case it is less clear at whom the objection is directed, since infinite Kollektivs then have no role to play in the theory of probability.
An adherent of the propensity interpretation may study Kollektivs for their own sake, as models for the deductions of probability theory, but to give a "good" definition becomes a fairly hopeless task: since one can't have satisfaction of all properties of probability one, it is necessary to choose, but what are the guiding principles for such a choice?
Note that, although von Mises' theory might seem to be plagued by the same problem (which set of place selections do we choose to define Kollektivs?) it is in reality less vulnerable: you need assume only that amount of invariance which allows you to perform a (successful) computation and if the computation fails to produce the right answer, you know the assumption of invariance was wrong.
No such empirical check exists for definitions of random sequences based on the propensity interpretation, such as those of Martin-Löf and Schnorr considered in the next chapter.

Another way to state von Mises' viewpoint on the relationship between Kollektivs in 2^ω and strong limit laws (considered as subsets of 2^ω) is the following.
If μ_p is considered as just a measure, there is no relationship at all. If μ_p is a veritable probability distribution, then there exists some Kollektiv $\xi \in (2^\omega)^\omega$ such that P_ξ defined by

$$P_\xi(A) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} 1_A(\xi_k)$$

conincides with μ_p on some reasonably large algebra $\mathbb{E}$ of events $A \subseteq 2^\omega$.(Von Mises briefly considered this set-up in [75,101]. Interestingly, he attributes it to Doob [20], although it is doubtful whether Doob would have been happy with this attribution[7].) Now even if $P_\xi(A) = 1$, this statement has no immediate bearing on Kollektivs in 2^ω; it tells us only that "most" ξ_k are in A. For reasonable definitions of Kollektivs in $(2^\omega)^\omega$, the ξ_k are themselves Kollektivs in 2^ω; but we see that there is no reason whatsoever why *all* Kollektivs in the sequence $\xi = (\xi_k)$, much less all Kollektivs in 2^ω, should satisfy A .

If μ_p is considerd as a probability measure, it describes the situation of picking points from 2^ω at random; a situation which is very different from that of picking zeros and ones at random to generate a sequence in 2^ω .

The latter procedure is evidently more constructive; and this was clearly one of the reasons why Borel preferred his own theory of "probabilités dénombrables", based on assumptions 1 and 2 as cited in 2.3.2, to measure theoretic probability [8,195], thus perhaps for the first time introducing free choice sequences (see Troelstra's survey of the history of choice sequences [97]).

Digression Another reason for Borel's preference was his conviction that the practical continuum (consisting of elements which can really be defined) is countable. So he states in the introduction to [8] that "dénombrable" refers to the cardinality of the sample space. Curiously, later authors, including Fréchet [28,53], thought that "dénombrable" refers to σ-additivity, in spite of Borel's statements to the contrary! Now Borel's conviction necessitated a new approach to probability theory, not based on measure theory, since an approach based on the latter seemed to require that the continuum be uncountable. The only measures *he* could think of were (what came to be called:) Lebesque-measure and measures defined from Lebesgue measure via densities; and all of these assign measure zero to countable sets. This point has been completely overlooked by Novikoff and Barone [79], who keep wondering about the "curious oversight" of Borel not to notice that probability theory *is* measure theory. This is not to say that Borel's reasoning is free of muddles; it is possible to do measure theory in a countable continuum, as Bishop [5] has shown. (End of digression.)

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 so-called Martingale) which may win unlimited amounts of money on the type of sequence constructed in 1. For the present discussion, one need not know precisely what a Martingale is; suffice it to say that it is given by a function $V: 2^{<\omega} \to \mathbb{R}^+$, where V(w) denotes the capital which the gambler , having played according to the strategy, possesses after w has occurred. The full definition will be given in Chapter 3. Ville exhibits a Martingale V such that for the sequence x constructed in 1., $\limsup_{n\to\infty} V(x(n)) = \infty$; but, obviously, since x is a Kollektiv, no gambling strategy in the sense of von Mises can win unlimited amounts of money on x. This objection is undoubtedly correct, but not very serious.

The purpose of von Mises' axioms is not to formalise the concept of an infinite sequence for which no successful gambling strategy exists. Rather, the purpose of the axioms is to lay down properties which allow the derivation of probabilistic laws. These properties are indeed justified by an appeal to the (empirical) "principle of the excluded gambling strategy" and perhaps this principle sanctions stronger axioms. For instance, in Chapter 3 we shall study definitions of randomness which take this principle as basic. But stronger axioms are necessary only if the given axioms do not suffice for the derivation of probabilistic laws.
At first sight it might seem that von Mises' theory cannot derive the characteristic properties of Martingales, e.g. the following:

$$(*) \text{ if } V \text{ is a Martingale (w.r.t. } \lambda), \lambda\{x \in 2^{\omega} | \limsup_{n\to\infty} V(x(n)) = \infty\} = 0.$$

But the situation here is completely analogous to that of the law of the iterated logarithm. There is no trouble in deriving the properties of Martingales in so far as they pertain to finite sequences (e.g. the Martingale inequality, from which (*) can be derived). The extension to infinite sequences is then, again, a matter of convention.
Conversely, we know by now that the derivation of (*) does not justify the requirement that for each Kollektiv y, $\limsup_{n\to\infty} V(x(n)) < \infty$.

But, one might argue, although Kollektivs such as x do not imperil the derivability of probabilistic laws, they may lead to wrong predictions. The following story illustrates what may go wrong and is at the same time an informal exposition of the results that will be obtained in 4.6.

Consider a casino, in which bets are placed on the outcomes of coin tosses. If the outcome is 1, the casino wins, otherwise the gambler wins. Beginning with the foundation of the establishment, the house issues each day a new coin with which the games have to be played. The management of the house, however, is thoroughly corrupt and issues coins which are false: the coin issued on the n^{th} day is such that the probability of heads on this day is $p_n = \frac{1}{2}(1 + (n+1)^{-\frac{1}{2}})$ (so that $p_n > \frac{1}{2}$, but $\lim_{n\to\infty} p_n = \frac{1}{2}$). The reason behind this devious procedure is the following.
A state inspector checkes the honesty of the casino by tossing a coin once a day, jotting down the outcome and testing at the end of the year (say) whether the sequence so obtained is Church random. The management of the house knows that, with the above choice of the p_n, there is a very large probability that the sequence in the inspector's notebook is indeed Church random (lemma 4.6.2). One day, however, the inspector learns of the definition of randomness given by Martin-Löf (Chapter 3), which is a (at least extensionally) a refinement of that of Church, and decides to check, after a year, whether the sequence of outcomes is Martin-Löf

random. Unfortunately for the management, there is also a very high probability that this sequence is not Martin-Löf random (theorem 4.6.1). However, after consulting the relevant literature (corollary 4.6.5), they change the value of p_n to $\frac{1}{2}(1+ (n+1)^{-1})$. To his satisfaction, the inspector notes that the sequences produced are (approximately) Martin-Löf random. The management is also satisfied, since no definition of randomness, however strong, can force them to change the value of p_n to a value which is less advantageous to them.

The moral of this tale is that, for each w, the inspector's prediction for the relative frequency of the occurrence of w on a specific day is false, regardless of whether a Church- or a Martin-Löf random sequence is used for the prediction (and that is the reason why the establishment is so profitable to its owners). Doesn't it follow that von Mises' theory fails in this case? No; the inspector could, on the basis of his data, only predict the relative frequencies of the outcomes of the experiment which consists of grouping (say) n days together and tossing a coin each day. The data are not relevant for the experiment which consists of taking a single day and grouping together the outcomes of n tosses with the coin isssued that day.

This concludes our review of the objections brought forward by Fréchet. These objections do not necessitate a revision either of strict frequentism or of the definition of Kollektivs; but we do not, of course, wish to claim that such objections are logically impossible.

2.7 Conclusions Two themes have occupied us in the preceding pages: the interpretation of probability and the definition of Kollektivs.

1. The great merit of von Mises' theory lies in the rigorous version of the frequency interpretation it presents. This interpretation, strict frequentism, is perhaps not the ultimate truth; but its main rival among the objective interpretations of probability, the propensity interpretation, has not yet arrived at a comparable stage of development, no one having investigated its consequences and assumptions as thoroughly as von Mises did for strict frequentism.

This is not to say that henceforth measure theoretic probability theory should be abandoned in favour of von Mises' theory. We view the relation between the first and the latter much as the relation between classical and constructive mathematics; there is nothing objectionable in doing classical mathematics, but if you really want to know what your results mean, you have to translate them in constructive terms, a translation which is sometimes impossible. Similarly, a deduction in measure theoretic probability theory should ideally be accompanied by a translation in terms of frequencies and Kollektivs; and this translation is not always trivial, as was demonstrated using the law of the iterated logarithm.

2. Von Mises' theory shows very clearly the assumptions that underlie any application of probability theory, in particular the necessity of the assumption that the mass phenomena to which probability theory is applicable be Kollektivs.

The older theory consists of two parts: invariance under place selections as an instrument for deductions and an explanation of invariance via admissibility.
The explanatory part has strong intuitive appeal, but is rather difficult to formalize; although the formalisation implicitly adopted in the alleged proof of inconsistency is blatantly not the one intended by von Mises.
We could distinguish two approaches toward formalisation: identifying admissible selections with lawlike selections and a contextual approach. For various reasons the identification of lawlikeness and admissibility leads to a much too restricted notion of the latter, and in particular leaves out the physical aspects.
Von Mises himself favoured the contextual approach, which means renouncing the attempt to define Kollektivs, but assuming in each specific instance the amount of invariance needed. To justify invariance, one may appeal to admissibility, but it does not occur anymore in the theory.
However, to study the question why probability theory is applicable to certain phenomena it seems best to follow the lines of the older theory and to make precise its basic idea: probabilistic computations are successful when they correspond to admissible place selections.
In subsequent chapters we present a piecemeal approach to this problem: different formalisations of admissibility which embody different aspects.
Lastly, we saw that, on the strict frequency interpretation, it suffices to define Kollektivs using place selections only. The demand that truly random sequences satisfy all strong limit laws proved by probability theory stems from a misinterpretation of the condition that limiting relative frequencies in a Kollektiv exist; such a demand can be justified at most on the propensity interpretation of probability.

Nevertheless, the objections voiced by Fréchet were almost universally accepted. Attempts to define Kollektivs became rare. A renewal of interest in the subject occurred only after Kolmogorov emphasized the necessity of Kollektivs for the frequency interpretation. For technical reasons, however, we start, not with Kolmogorov's own proposal, but with a later development: Martin-Löf's definition.

Notes to Chapter 2

1. Kolmogorov's *Grundbegriffe* contains a paragraph on "Das Verhältnis zur Erfahrungswelt" in which he says

> In de Darstellung der notwendigen Voraussetzungen für die Anwendbarkeit der Wahrscheilichkeitsrechnung auf die Welt der reellen Geschehnisse folgt der Verfasser im hohen Maße den Ausführungen von Herrn von Mises [44,3].

But his *condition B* is slightly awkward from a strict frequentist point of view:

> B. Ist P(A) sehr klein, so kann man praktisch sicher sein, daß bei einer einmaligen Realisation der Bedingungen [which determine the occurrence of A or its complement] das Ereignis A nicht stattfindet [44,4].

This condition contains a vestige of the propensity interpretation and does not harmonize very well with von Mises' views on the meaning of probability zero. However, even in von Mises-Geiringer [74,110] we read:

> Hence we assume that *in certain known fields of application the frequency limits are approached fairly rapidly*. We also assume that certain "privileged" sequences (to be expected by the law of large numbers) appear right from the beginning and not only after a million of trials.

Apparently, the part of [74] where this passage occurs was not written by von Mises (see the preface to [74]); I know of no comparable passage in von Mises' own works ([67] to [73]).

2. But note that von Mises' axioms for Kollektivs go much further and attempt to capture the *independence* of the successive casts, using asymptotic properties in a way which is anathema to the intuitionist. See also note 5.

3. For simplicity, we call *independent* what von Mises calls *independent and combinable* [74,31].

4. We saw in 2.4 that lawlike selections do not suffice for this purpose.

5. In our overview of the history of Kollektivs, we did not consider objections inspired by various forms of constructivism. But it will be clear that, for those who hold that the mathematical universe consists of lawlike objects only, Kollektivs are equally impossible. For in this case, if x is a purported Kollektiv, the set $\{n \mid x_n = 1\}$ is itself lawlike (see Reichenbach [85]). Other objections were based on the conviction that the convergence of the relative frequency postulated of Kollektivs had to be uniform; see, e.g., the lecture notes "Grondslagen der Waarschijnlijkheidsrekening" [Foundations of Probability] by D. van Dantzig (library of the Mathematical Institute, University of Amsterdam).

6. Wald's suggestion occurs in [101,98]. He defines a place selection to be *singular* (with respect to Lebesgue measure) if its domain has Lebesgue measure zero. A sequence x is a *Kollektiv in the strong sense* (with respect to $(\frac{1}{2},\frac{1}{2})$ and some countable set of place selections $\mathcal{H}$) if it is a Kollektiv in the old sense and is, moreover, not contained in the domain of a singular place selection in $\mathcal{H}$. Now given any countable set of probabilistic laws (with respect to Lebesgue measure) one can construct a set of place selections $\mathcal{H}$, such that a Kollektiv in the strong sense with respect to $\mathcal{H}$ satsfies these laws. For by the regularity of Lebesgue measure, the set of sequences not satisfying a probabilistic law is contained in a G_δ set. However, the domain of a place selection is also a G_δ set and it is easy to construct a place

selection whose domain is a given G_δ. Since the complement of a probabilistic law has measure zero, place selections so constructed are ipso facto singular.

Because that part of the law of the iterated logarithm which is of interest to us, is itself a G_δ set, we could use a simpler construction.

7. As will be clear from the discussion of the meaning of independence in 2.4, the measure $\mu_{\frac{1}{2}}$ refers to the following experimental set-up: each time you want to toss a coin, you take a *new* fair coin. In von Mises' theory this situation is to be distinguished from that of repeatedly tossing the same coin: in this case the productrule is provable. Apparently, von Mises considered the possibility of dropping this feature: see his references to the "Tornier-Doob frequency theory" in [75,101]. Tornier's theory is explained in Feller [23], von Mises-Geiringer [74] and Martin-Löf [63].