Published in: *Journal of Scientific Exploration* **13**, 579–613 (1999)

Basic elements and problems of probability theory

Hans Primas
ETH Zürich[1]

Abstract

After a brief review of ontic and epistemic descriptions, and of subjective, logical and statistical interpretations of probability, we summarize the traditional axiomatization of calculus of probability in terms of Boolean algebras and its set-theoretical realization in terms of Kolmogorov probability spaces. Since the axioms of mathematical probability theory say nothing about the conceptual meaning of "randomness" one considers probability as property of the *generating conditions* of a process so that one can relate randomness with predictability (or retrodictability). In the measure-theoretical codification of stochastic processes genuine chance processes can be defined rigorously as so-called regular processes which do not allow a long-term prediction. We stress that stochastic processes are *equivalence classes* of individual point functions so that they do not refer to individual processes but only to an ensemble of statistically equivalent individual processes.

Less popular but conceptually more important than statistical descriptions are individual descriptions which refer to individual chaotic processes. First, we review the individual description based on the generalized harmonic analysis by Norbert Wiener. It allows the definition of individual purely chaotic processes which can be interpreted as trajectories of regular statistical stochastic processes. Another individual description refers to algorithmic procedures which connect the intrinsic randomness of a finite sequence with the complexity of the shortest program necessary to produce the sequence.

Finally, we ask why there can be laws of chance. We argue that random events fulfill the laws of chance if and only if they can be reduced to (possibly hidden) deterministic events. This mathematical result may elucidate the fact that not all nonpredictable events can be grasped by the methods of mathematical probability theory.

1. Overview

1.1 Ontic and epistemic descriptions

One of the most important results of contemporary classical dynamics is the proof that the deterministic differential equations of some smooth classical Hamiltonian systems have solutions exhibiting irregular behavior. The classical view of physical determinism has been eloquently formulated by Pierre Simon Laplace. While Newton believed that the stability of the solar system could only be achieved with the help of God, Laplace "had no need of that hypothesis" [2] since he could explain the solar system by the deterministic Newtonian mechanics alone. Laplace discussed his doctrine of determinism in the introduction to his *Philosophical Essay on Probability*, in which he imaged a superhuman intelligence capable of grasping the initial conditions at any fixed time of all bodies and atoms of the universe, and all the forces acting upon it. For such a superhuman intelligence "nothing would be uncertain and the future, as the past, would be present to its eyes." [3] Laplace's reference to the future *and* the past implies that he refers to a fundamental theory with an unbroken time-reversal symmetry. His reference to a

---

1 Laboratory of Physical Chemistry, ETH-Zentrum, CH-8092 Zürich (Switzerland), Primas@phys.chem.ethz.ch.

2 Laplace's famous replay to Napoleon's remark that he did not mention God in his *Exposition du Système du Monde*.

3 Laplace (1814). Translation taken from the Dover edition, p. 4.

"superhuman intelligence" suggests that he is not referring to our possible knowledge of the world, but to things "as they really are". The manifest impossibility to ascertain experimentally exact initial conditions necessary for a description of things "as they really are" is what led Laplace to the introduction of a statistical description of the initial conditions in terms of probability theory. Later Josiah Willard Gibbs introduced the idea of an ensemble of a very large number of imaginary copies of mutually uncorrelated individual systems, all dynamically precisely defined but not necessarily starting from precisely the same individual states.[4] The fact that a statistical description in the sense of Gibbs *presupposes* the existence of a well-defined *individual* description demonstrates that *a coherent statistical interpretation in terms of an ensemble of individual systems requires an individual interpretation as a backing.*

The empirical inaccessibility of the precise initial states of most physical systems requires a distinction between epistemic and ontic interpretations.[5] *Epistemic interpretations* refer to our knowledge of the properties or modes of reactions of observed systems. On the other hand, *ontic interpretations* refer to intrinsic properties of hypothetical individual entities, regardless of whether we know them or not, and independently of observational arrangements. Albeit ontic interpretations do not refer to our knowledge, there is a meaningful sense in which it is natural to speak of theoretical entities "as they really are", since in good theories they supply the indispensable explanatory power.

States which refer to an epistemic interpretation are called *epistemic states*, they refer to our knowledge. If this knowledge of the properties or modes of reactions of systems is expressed by probabilities in the sense of relative frequencies in a statistical ensemble of independently repeated experiments, we speak of a statistical interpretation and of *statistical states*. States which refer to an ontic interpretation are called *ontic states*. Ontic states are assumed to give a description of a system "as it really is", that is, independently of any influences due to observations or measurements. They refer to individual systems and are assumed to give an exhaustive description of a system. Since an ontic description does not encompass any concept of observation, ontic states do not refer to predictions of what happens in experiments. At this stage it is left open to what extent ontic states are knowable. An adopted ontology of the intrinsic description induces an operationally meaningful epistemic interpretation for every epistemic description: *an epistemic state refers to our knowledge of an ontic state.*

1.2 Cryptodeterministic systems

In modern mathematical physics Laplacian determinism is rephrased as *Hadamard's principle of scientific determinism* according to which every initial ontic state of a physical system determines all future ontic states.[6] An ontically deterministic dynamical system which even in principle does not allow a precise forecast of its observable behavior in the remote future will be called *cryptodeterministic.*[7] Already, Antoine Augustine Cournot (1801–1877) and John Venn (1834–1923) recognized clearly that the dynamics of complex dynamical classical systems may depend in an extremely sensitive way on the initial and boundary conditions. Even if we can determine these conditions with arbitrary but finite accuracy, the individual outcome cannot be predicted; the resulting chaotic dynamics allows only an epistemic description in terms of

4 Gibbs (1902). A lucid review of Gibbs' statistical conception of physics can be found in Haas (1936), volume II, chapter R.

5 This distinction is due to Scheibe (1964), Scheibe (1973), pp. 50–51.

6 Compare Hille & Phillips (1957), p. 618.

7 In a slightly weaker form, this concept has been introduced by Edmund Whittaker (1943).

statistical frequencies.[8] The instability of such deterministic processes represents an *objective* feature of the corresponding probabilistic description. A typical experiment which demonstrates the *objective* probabilistic character of a cryptodeterministic mechanical system is Galton's desk.[9] Modern theory of deterministic chaos has shown how unpredictability can arise from the iteration of perfectly well-defined functions because of a sensitive dependence on initial conditions.[10] More precisely, the catchword "deterministic chaos" refers to *ontically* deterministic systems with a sensitive dependence on the ontic initial state such that no *measurement* on the systems allows a long-term prediction of the ontic state of the system.

Predictions refer to inferences of the observable future behavior of a system from empirically estimated initial states. While in some simple systems the ontic laws of motion may allow to forecast its observable behavior in the near future with great accuracy, *ontic determinism implies neither epistemic predictability nor epistemic retrodictability.* Laplace knew quite well that a perfect measurement of initial condition is impossible, and he never asserted that deterministic systems are empirically predictable. Nevertheless, many positivists tried to define determinism by predictability. For example Herbert Feigl:

> "The clarified (purified) concept of causation is defined in terms of predictability according to a law (or, more adequately, according to a set of laws)." [11]

Such attempts are based on a notorious category mistake. *Determinism does not deal with predictions.* Determinism refers to an *ontic* description. On the other hand, predictability is an *epistemic* concept. Yet, epistemic statements are often confused with ontic assertions. For example, Max Born has claimed that classical point mechanics is not deterministic since there are instable mechanical systems which are epistemically not predictable.[12] Similarly, it has been claimed that human behavior is not deterministic since it is not predictable.[13] A related mistaken claim is that "… an underlying deterministic mechanism would refute a probabilistic theory by contradicting the randomness which … is demanded by such a theory."[14] As emphasized by John Earman:

> "The history of philosophy is littered with examples where ontology and epistemology have been stirred together into a confused and confusing brew. … Producing an 'epistemological sense' of determinism is an abuse of language since we already have a perfectly adequate and more accurate term – prediction – and it also invites potentially misleading argumentation – e.g., in such-an-such a case prediction is not possible and, therefore, determinism fails." [15]

---

8 Compare Cournot (1843), §40; Venn (1866).

9 Galton's desk (after Francis Galton, 1822–1911) is a inclined plane provided with regularly arranged nails in n horizontal lines. A ball launched on the top will be diverted at every line either to left or to right. Under the last line of nails there are $n+1$ boxes (numbered from the left from $k=0$ to $k=n$) in which the balls are accumulated. In order to fall into the k-th box a ball has to be diverted k times to the right and $n-k$ times to the left. If at each nail the probability for the ball to go to left or to right is $\frac{1}{2}$, then the distribution of the balls is given by the binomial distribution $\binom{n}{k}(\frac{1}{2})^n$, which for large n approach a Gaussian distribution. Our ignorance of the precise initial and boundary does not allow us to predict individual events. Nevertheless, the experimental Gaussian distribution does in no way depend on our knowledge. In this sense, we may speak of *objective chance events.*

10 For an introduction into the theory of deterministic chaos, compare for example Schuster (1984).

11 Feigl (1953), p. 408.

12 Born (1955a), Born (1955b). For a critique of Born's view compare Von Laue (1955).

13 Scriven (1965). For a critique of Scriven's view compare Boyd (1972).

14 Gillies (1973), p. 135.

15 Earman (1986), pp. 6–7.

1.3 Kinds of probability

Often, probability theory is considered as the natural tool for an epistemic description of cryptodeterministic systems. However, this view is not as evident as is often thought. The virtue and the vice of modern probability theory are the split-up into a probability calculus and its conceptual foundation. Nowadays, mathematical probability theory is just a branch of pure mathematics, based on some axioms devoid of any interpretation. In this framework, the concepts "probability", "independence", etc. are conceptually unexplained notions, they have a purely mathematical meaning. While there is a widespread agreement concerning the essential features of the calculus of probability, there are widely diverging opinions what the *referent* of mathematical probability theory is. [16] While some author claim that probability refers exclusively to ensembles, there are important problems which require a discussion of single random events or of individual chaotic functions. Furthermore, it is noway evident that the calculus of axiomatic probability theory is appropriate for empirical science. In fact, "probability is one of the outstanding examples of the 'epistemological paradox' that we can successfully use our basic concepts without actually understanding them." [17]

Surprisingly often it is assumed that in a scientific context everybody means intuitively the same when speaking of "probability", and that the task of an interpretation only consists in exactly capturing this single intuitive idea. Even prominent thinkers could not free themselves from predilections which only can be understood from the historical development. For example, Friedrich Waismann[18] categorically maintains that there is no other motive for the introduction of probabilities than the incompleteness of our knowledge. Just as dogmatically, Richard von Mises[19] holds that, without exceptions, probabilities are empirical and that there is no possibility to reveal the values of probabilities with the aid of another science, e.g. mechanics. On the other hand, Harold Jeffreys maintains that "no 'objective' definition of probability in terms of actual or possible observations, or possible properties of the world, is admissible."[20] Leonard J. Savage claims that "personal, or subjective, probability is the only kind that makes reasonably rigorous sense." [21] However, despite many such statements to the contrary, we may state with some confidence that there is not just one single "correct" interpretation. *There are various valid possibilities to interpret mathematical probability theory.* Moreover, the various interpretations do not fall neatly into disjoint categories. As Bertrand Russell underlines,

> "in such circumstances, the simplest course is to enumerate the axioms from which the theory can be deduced, and to decide that any concept which satisfies these axioms has an equal right, from the mathematician's point of view, to be called 'probability'. … It must be understood that there is here no question of truth or falsehood. Any concept which satisfies the axioms may be taken to *be* mathematical probability. In fact, it might be desirable to adopt one interpretation in one context, and another in another." [22]

---

16 The definition and interpretation of probability has a long history. There exists an enormous literature on the conceptual problems of the classical probability calculus which cannot be summarized here. For a first orientation, compare the monographs by Fine (1973), Maistrov (1974), Von Plato (1994).

17 Von Weizsäcker (1973), p. 321.

18 Waismann (1930).

19 Von Mises (1928).

20 Jeffreys (1939).

21 Savage (1962), p. 102.

22 Russell (1948), pp. 356–357.

1.4 Subjective probability

A probability interpretation is called *objective* if the probabilities are assumed to be independent or dissected from any human considerations. *Subjective* interpretations consider probability as a rational measure of the personal belief that the event in question occurs. A more operationalistic view defines subjective probability as the *betting rate* on an event which is fair according to the opinion of a given subject. It is required that the assessments a rational person makes are *logically coherent* such that no logical contradictions exist among them. The postulate of coherence should make it impossible to set up a series of bets against a person obeying these requirements in such a manner that the person is sure to lose, regardless of the outcome of the events being wagered upon. Subjective probabilities depend on the degree of personal knowledge and ignorance concerning the events, objects or conditions under discussion. If the personal knowledge changes, the subjective probabilities change too. Often, it is claimed to be evident that subjective probabilities have no place in a physical theory. However, subjective probability cannot be disposed of quite that simply. It is astonishing, how many scientists uncompromisingly defend an objective interpretation without knowing any of the important contributions on subjective probability published in the last decades. Nowadays, there is a very considerable rational basis behind the concept of subjective probability.[23]

It is debatable how the pioneers would have interpreted probability, but their practice suggests that they dealt with some kind of "justified degree of belief". For example, in one of the first attempts to formulate mathematical "laws of chance", Jakob Bernoulli characterized 1713 his *Ars Conjectandi* probability as a *strength of expectation*.[24] For Pierre Simon Laplace probabilities represents a state of knowledge, he introduced *a priori* or g*eometric probabilities* as the ratio of favorable to "equally possible" cases[25] – a definition of historical interest which, however, is both conceptually and mathematically inadequate.

The early subjective interpretations are since long out of date, but practicing statisticians have always recognized that subjective judgments are inevitable. In 1937, Bruno de Finetti made a fresh start in the theory of subjective probability by introducing the essential new notion of *exchangeability*.[26] De Finetti's subjective probability is a betting rate and refers to single events. A set of n distinct events $E_1, E_2, \cdots, E_n$ are said to be *exchangeable* if any event depending on these events has the same subjective probability (de Finetti's betting rate) no matter how the E_j are chosen or labelled. Exchangeability is sufficient for the validity of the law of large numbers. The modern concept of subjective probability is not necessarily incompatible with that of objective probability. De Finetti's representation theorem gives convincing explanation how there can be wide intersubjective agreement about the values of subjective probabilities. According to Savage, a rational man behaves as if he used subjective probabilities.[27]

---

23 Compare for example Savage (1954), Savage (1962), Good (1965), Jeffrey (1965). For a convenient collection of the most important papers on the modern subjective interpretation, compare Kyburg & Smokler (1964).

24 Bernoulli (1713).

25 Compare Laplace (1814).

26 DeFinetti (1937). Compare also the collection of papers DeFinetti (1972) and the monographs DeFinetti (1974), DeFinetti (1975).

27 Savage (1954), Savage (1962).

1.5 Inductive probability

Inductive probability belongs to the field of scientific inductive inference. Induction is the problem of how to make inferences from observed to unobserved (especially future) cases. It is an empirical fact that we can learn from experience, but the problem is that nothing concerning the future can be *logically* inferred from past experience. It is the merit of the modern approaches to have recognized that induction has to be some sort of probabilistic inference, and that the induction problem belongs to a generalized logic. Logical probability is related to, but not identical with subjective probability. Subjective probability is taken to represent the extent to which a person believes a statement is true. The logical interpretation of probability theory is a generalization of the classical implication, it is not based on empirical facts but on the logical analysis of these. The inductive probability is the degree of confirmation of a hypothesis with reference to the available evidence in favor of this hypothesis.

The logic of probable inference and the logical probability concept goes back to the work of John Maynard Keynes who in 1921 defined probability as a "logical degree of belief".[28] This approach has been extended by Bernard Osgood Koopman[29] and especially by Rudolf Carnap to a comprehensive system of inductive logic.[30] Inductive probabilities occur in science mainly in connection with *judgements* of empirical results; they are always related to a *single case*, they are never to be interpreted as frequencies. The inductive probability is also called "nondemonstrative inference", "intuitive probability" (Koopman), "logical probability" or "probability$_1$" (Carnap). A hard nut to crack in probabilistic logic is the proper choice of a probability measure – it cannot be estimated empirically. Given a certain measure inductive logic works with a fixed set of rules so that all inferences can be effected automatically by a general computer. In this sense inductive probabilities are objective quantities.[31]

1.6 Statistical probability

Historically statistical probabilities have been interpreted as limits of frequencies, that is, as *empirical properties* of the system (or process) considered. But, statistical probabilities cannot be assigned to a single event. This is an old problem of the frequency interpretation of which already John Venn was aware. In 1866 Venn tried to define a probability explicitly in terms of relative frequencies of occurrence of events "in the long run". He added that "the run must be supposed to be very long, in fact never to stop."[32] Against this simpleminded frequency interpretation there is a grave objection: any empirical evidence concerning relative frequencies is necessarily restricted to a *finite* set of events. Yet, without additional assumptions *nothing can be inferred about the value of the limiting frequency of a finite segment, no matter how long it may be.* Therefore, the statistical interpretation of the calculus of probability has to be supplemented by a decision technique that allows to decide which probability statements we should accept. Satisfactory acceptance rules are notoriously difficult to formulate.

The simplest technique is the old maxim of Antoine Augustine Cournot: *if the probability of an event is sufficiently small, one should act in a way as if this event will not occur at a solitary*

28 Keynes (1921).

29 Koopman (1940a), Koopman (1940b), Koopman (1941).

30 Carnap (1950), Carnap (1952), Carnap & Jeffrey (1971). Carnap's concept of logical probabilities has been critized sharply by Watanabe (1969a).

31 For a critical evaluation of the view that statements of probability can be logically true, compare Ayer (1957), and the ensuing discussion, pp. 18–30.

32 Venn (1866), chapter VI, §35 , §36.

realization.[33] However, the theory gives no criterion for deciding what is "sufficiently small". A more elegant (but essentially equivalent) way out is the proposal by Carl Friedrich von Weizsäcker to consider probability as a *prediction* of a relative frequency, so that "the probability is only the expectation value of the relative frequency".[34] That is, we need in addition a judgment about a statement. This idea is in accordance with Carnap's view that two meanings of probability must be recognized: the *inductive probability* (his "probability$_1$"), and *statistical probability* (his "probability$_2$").[35] The logical probability is supposed to express a logical relation between a given evidence and a hypothesis. They "speak about statements of science; therefore, they do not belong to science proper but to the logic or methodology of science, formulated in the metalanguage." On the other hand, "the statements on statistical probability, both singular and general statements, e.g., probability laws in physics or in economics, are synthetic and serve for the description of general features of facts. Therefore these statements occur *within* science, for example, in the language of physics (taken as object language)." [36] That is, according to this view, inductive logic with its logical probabilities is a necessary completion of statistical probabilities: without inductive logic we cannot infer statistical probabilities from observed frequencies. The supplementation of the frequency interpretation by a subjective factor cannot be avoided by introduction of a new topology. For example, if one introduces the topology associated with the field of p-adic numbers[37], one has to select subjectively a finite prime number p. As emphasized by Wolfgang Pauli, no frequency interpretation can avoid a subjective factor:

> „An irgend einer Stelle [muss] eine Regel für die praktische Verhaltungsweise des Menschen oder spezieller des Naturforschers hinzugenommen werden, die auch dem subjektiven Faktor Rechnung trägt, nämlich: auch die einmalige Realisierung eines sehr unwahrscheinlichen Ereignisses wird von einem gewissen Punkt an als praktisch unmöglich angesehen. … An dieser Stelle stösst man schliesslich auf die prinzipielle Grenze der Durchführbarkeit des ursprünglichen Programmes der rationalen Objektivierung der einmaligen subjektiven Erwartung." [38]

Later Richard von Mises[39] tried to overcome this difficulty by introducing the notion of "irregular collectives", consisting of one infinite sequence in which the limit of the relative frequency of each possible outcome exists and is indifferent to a place selection. In this approach the value of this limit is called the probability of this outcome. The essential underlying idea was the "impossibility of a successful gambling system". While at first sight Mises' arguments seemed to be reasonable, he could not achieve a convincing success.[40] However, Mises' approach provided the crucial idea for the fruitful computational-complexity approach to random sequences, discussed in more detail below.

---

33 Cournot (1843). This working rule was still adopted by Kolmogoroff (1933), p. 4.

34 Von Weizsäcker (1973), p. 326. Compare also Von Weizsäcker (1985), pp. 100–118.

35 Carnap (1945), Carnap (1950).

36 Carnap (1963), p. 73.

37 Compare for example Khrennikov (1994), chapters VI and VII.

38 Pauli (1954), p. 114.

39 Von Mises (1919), Von Mises (1928), Von Mises (1931). The English edition Von Mises (1964) was edited and complemented by Hilda Geiringer; it is strongly influenced by the views of Erhard Tornier and does not necessarily reflect the views of Richard Mises.

40 The same is true for the important modifications of Mises' approach by Tornier (1933) and by Reichenbach (1994). Compare also the review by Martin-Löf (1969a).

2. Mathematical probability

2.1 Mathematical probability as a measure on a Boolean algebra

In the mathematical codification of probability theory a chance event is defined only implicitly by axiomatically characterized relations between events. These relations have a *logical* character so that one can assign to every event a proposition stating its occurrence. All codifications of classical mathematical probability theory are based on *Boolean classifications* or *Boolean logic*. That is, the algebraic structure of events is assumed to be a *Boolean algebra*, called the *algebra of events.* In 1854, George Boole introduced these algebras in order

> "to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolic language of a Calculus, and upon this foundation to establish the science of Logic and to construct its method; to make that method itself the basis of a general method for the application of the mathematical doctrine of Probabilities … "[41]

Mathematical probability is anything that satisfies the axioms of mathematical probability theory. As we will explain in the following in some more details, *mathematical probability theory is the study of a pair* $(\mathfrak{B}, p)$*, where the algebra of events is a* σ*-complete Boolean algebra* $\mathfrak{B}$*, and the map* $p : \mathfrak{B} \to [0,1]$ *is a* σ*-additive probability measure.*[42]

Pro memoria: Boolean algebras[43]

> A Boolean algebra is a non-empty set $\mathfrak{B}$ in which two binary operations $\vee$ (addition or disjunction) and $\wedge$ (multiplication or conjunction), and a unary operation $\perp$ (complementation or negation) with the following properties are defined
>
> - the operations $\vee$ and $\wedge$ are commutative and associative,
> - the operation $\vee$ is distributive with respect to $\wedge$, and vice versa,
> - for every $A \in \mathfrak{B}$ and every $B \in \mathfrak{B}$ we have $A \vee A^{\perp} = B \vee B^{\perp}$ and $A \wedge A^{\perp} = B \wedge B^{\perp}$,
> - $A \vee (A \wedge A^{\perp}) = A \wedge (A \vee A^{\perp}) = A$
>
> These axioms imply that in every Boolean algebra there are two distinguished elements $\mathbf{1}$ (called the unit of $\mathfrak{B}$) and $\mathbf{0}$ (called the zero of $\mathfrak{B}$), defined by $A \vee A^{\perp} = \mathbf{1}$, $A \wedge A^{\perp} = \mathbf{0}$ for every $A \in \mathfrak{B}$. With this it follows that $\mathbf{0}$ is the neutral element of the addition $\vee$, $A \vee \mathbf{0} = A$ for every $A \in \mathfrak{B}$, and that $\mathbf{1}$ is the neutral element of the multiplication $\wedge$, $A \wedge \mathbf{1} = A$ for every $A \in \mathfrak{B}$.

An algebra of events is a Boolean algebra $(\mathfrak{B}, \wedge, \vee, \perp)$. If an element $A \in \mathfrak{B}$ is an event, then $A^{\perp}$ is the event that A does not take place. The element $A \vee B$ is the event which occurs when at least one of the events A and B occurs, while $A \wedge B$ is the event when both events A and B occur. The unit element $\mathbf{1}$ represents the sure event while the zero element $\mathbf{0}$ represents the impossible element. If A and B are any two elements of the Boolean algebra $\mathfrak{B}$ which satisfies the relation $A \vee B = B$ (or the equivalent relation $A \wedge B = A$) we say that "A is smaller than B", or that "A implies B", and write $A \le B$.

Probability is defined as a norm $p : \mathfrak{B} \to [0,1]$ on a Boolean algebra $\mathfrak{B}$ of events. That is, to every event $A \in \mathfrak{B}$ there is associated a probability $p(A)$ for the occurrence of the event A. The following properties are required for $p(A)$:

- p is strictly positive, i.e. $p(A) \ge 0$ for every $A \in \mathfrak{B}$ and $p(A) = 0$ if and only if $A = 0$, where 0 is the zero of $\mathfrak{B}$,

---

41 Boole (1854), p. 1.

42 Compare Halmos (1944), Kolmogoroff (1948), Łoš, (1955). A detailed study of the purely lattice-theoretical ("point-free") approach to classical probability can be found in the monograph by Kappos (1969).

43 For more details, compare Sikorski (1969).

- p is normed, i.e. $p(\mathbf{1}) = 1$, where $\mathbf{1}$ is the unit of $\mathfrak{B}$,
- p is additive, i.e. $p(A \vee B) = p(A) + p(B)$ if A and B are disjoint, that is if $A \wedge B = 0$.

It follows that $0 \le p(A) \le 1$ for every $A \in \mathfrak{B}$, and $A \le B \Rightarrow p(A) \le p(B)$.

In contrast to a Kolmogorov probability measure, the measure p is *strictly* positive. That is, $p(B) = 0$ implies that B is the *unique* smallest element of the Boolean algebra $\mathfrak{B}$ of events.

In probability theory it is necessary to consider also countably infinitely many events so that one needs in addition some continuity requirements. By a Boolean σ-algebra one understands a Boolean algebra where the addition and multiplication operations are performable on each countable sequence of events. That is, in a Boolean σ-algebra $\mathfrak{B}$ there is for every infinite sequence A_1, A_2, $A_3, \cdots$ of elements of $\mathfrak{B}$ a smallest element $A_1 \vee A_2 \vee A_3 \vee \cdots \in \mathfrak{B}$. The continuity required for the probability p is then the so-called σ-additivity:

- a measure p on a σ-algebra is σ-additive if $p\{\vee_{k=1}^{\infty} A_k\} = \sum_{k=1}^{\infty} p(A_k)$ whenever $\{A_k\}$ is a sequence of pairwise disjoint events, $A_j \wedge A_k = \mathbf{0}$ for all $j \ne k$.

Since not every Boolean algebra is a σ-algebra, the property of *countable additivity* is an essential restriction.

2.2 Set-theoretical probability theory

It is almost universally accepted that mathematical probability theory consists of the study of Boolean σ-algebras. For reasons of mathematical convenience, one usually represents the Boolean algebra of events by a Boolean algebra of subsets of some set. Using this representation one can go back to a well-established integration theory, to the theory of product measures, and to the Radon–Nikodým-theorem for the definition of conditional probabilities. According to a fundamental representation theorem by Marshall Harvey Stone every ordinary Boolean algebra with no further condition is isomorphic to the algebra $(\mathfrak{P}(\boldsymbol{\Omega}), \cap, \cup, ')$ of all subsets of some point set $\boldsymbol{\Omega}$.[44] Here $\mathfrak{B}$ corresponds to the power set $\mathfrak{P}(\boldsymbol{\Omega})$ of all subsets of the set $\boldsymbol{\Omega}$, the conjunction $\wedge$ corresponds to the set-theoretical intersection $\cap$, the disjunction $\vee$ corresponds to the set-theoretical union $\cup$, and the negation $\perp$ corresponds to the set-theoretical complementation $'$. The multiplicative neutral element $\mathbf{1}$ corresponds to the set $\boldsymbol{\Omega}$, while the additive neutral element corresponds to the empty set $\emptyset$. However, a σ-complete Boolean algebra is in general not σ-isomorphic to a σ-complete Boolean algebra of point sets. Yet, every σ-complete Boolean algebra is σ-isomorphic to a σ-complete Boolean algebra of point sets modulo a σ-ideal in that algebra.[45]

Conceptually, this result is the starting point for the axiomatic foundation by Andrei Nikolaevich Kolmogorov of 1933 which reduces mathematical probability theory to classical measure theory.[46] It is based on a so-called probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ consisting of a non-empty set $\boldsymbol{\Omega}$ (called *sample space*) of points, a class $\boldsymbol{\Sigma}$ of subsets of $\boldsymbol{\Omega}$ which is a σ-algebra (i.e. is closed with respect to the set-theoretical operations executed a countable number of times), and a probability measure $\boldsymbol{\mu}$ on $\boldsymbol{\Sigma}$. Sets that belong to $\boldsymbol{\Sigma}$ are called *$\boldsymbol{\Sigma}$-measurable* (or just *measurable* if $\boldsymbol{\Sigma}$ is understood). The pair $(\boldsymbol{\Omega}, \boldsymbol{\Sigma})$ is called a *measurable space*. A probability measure $\boldsymbol{\mu}$ on $(\boldsymbol{\Omega}, \boldsymbol{\Sigma})$ is a function $\boldsymbol{\mu} : \boldsymbol{\Sigma} \to [0,1]$ satisfying $\boldsymbol{\mu}(\emptyset) = 0$, $\boldsymbol{\mu}(\boldsymbol{\Omega}) = 1$, and the condition of

44 Stone (1936).

45 Loomis (1947).

46 Kolmogoroff (1933). There are many excellent texts on Kolmogorov's mathematical probability theory. Compare for example: Breiman (1968), Prohorov & Rozanov (1969), Laha & Rohatgi (1979), Rényi (1970a), Rényi (1970b). Recommendable introductions to measure theory are, for example: Cohn (1980) , Nielsen (1997).

countable additivity (that is, $\boldsymbol{\mu}\{\cup_{n=1}^{\infty}\mathcal{B}_n\} = \sum_{n=1}^{\infty}\boldsymbol{\mu}(\mathcal{B}_n)$ whenever $\{\mathcal{B}_n\}$ is a sequence of members of $\boldsymbol{\Sigma}$ which are pairwise disjoint subsets in $\boldsymbol{\Omega}$). The points of $\boldsymbol{\Omega}$ are called elementary events. The subsets of $\boldsymbol{\Omega}$ belonging to $\boldsymbol{\Sigma}$ are referred to as events. The nonnegative number $\boldsymbol{\mu}(\mathcal{B})$ is called the *probability* of the event $\mathcal{B}\in\boldsymbol{\Sigma}$.

In most applications the sample space $\boldsymbol{\Omega}$ contains uncountably many points. In this case, there exist non-empty Borel sets in $\boldsymbol{\Sigma}$ of measure zero, so that there is no strictly positive σ-additive measure on $\boldsymbol{\Sigma}$. But it is possible to eliminate the sets of measure zero by using the σ-complete Boolean algebra $\mathfrak{B} = \boldsymbol{\Sigma}/\boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is the σ-ideal of Borel sets of $\boldsymbol{\mu}$-measure zero. With this, every Kolmogorov probability space $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{\mu})$ generates probability algebra with the σ-complete Boolean algebra $\mathfrak{B} = \boldsymbol{\Sigma}/\boldsymbol{\Delta}$ and the restriction of $\boldsymbol{\mu}$ to $\mathfrak{B}$ is a strictly positive measure p. Conversely, every probability algebra $(\mathfrak{B}, p)$ can be realized by some Kolmogorov probability space $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{\mu})$ with $\mathfrak{B} \sim \boldsymbol{\Sigma}/\boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is the σ-ideal of Borel sets of $\boldsymbol{\mu}$-measure zero.

One usually formulates the set-theoretical version of probability theory directly in terms of the conceptually less transparent triple $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{\mu})$, and not in terms of the probabilistically relevant Boolean algebra $\mathfrak{B} = \boldsymbol{\Sigma}/\boldsymbol{\Delta}$. Since there exist non-empty Borel sets in $\boldsymbol{\Sigma}$ (i. e., events different from the impossible event) of measure zero, one has to use the "almost everywhere" terminology. A statement is said to be true "almost everywhere" or "for almost all $\boldsymbol{\omega}$" if it is true for all $\boldsymbol{\omega}\in\boldsymbol{\Omega}$ except, may be, in a set $\mathcal{N}\in\boldsymbol{\Sigma}$ of measure zero, $\boldsymbol{\mu}(\mathcal{N}) = 0$. If the sample space $\boldsymbol{\Omega}$ contains uncountably many points, elementary events do not exist in the operationally relevant version in terms of the *atom-free* Boolean algebra $\boldsymbol{\Sigma}/\boldsymbol{\Delta}$. JOHANN VON NEUMANN has argued convincingly that the finest events which are empirically accessible are given by Borel sets of nonvanishing Lebesgue measure, and not by the much larger class of *all* subsets of $\boldsymbol{\Omega}$.[47]

This setting is almost universally accepted, either explicitly or implicitly. However, some paradoxical situations do arise unless further restrictions are placed on the triple $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{\mu})$. The requirement that a probability measure has to be a *perfect* measure avoids many difficulties.[48] Furthermore, in all physical applications there are natural additional regularity conditions. In most examples the sample space $\boldsymbol{\Omega}$ is polish (i.e. separable and metrisable), the σ-algebra $\boldsymbol{\Sigma}$ is taken as the σ-algebra of Borel sets[49] and $\boldsymbol{\mu}$ is a regular Radon measure.[50] Moreover, there are some practically important problems which require the use of unbounded measures, a feature which does not fit into Kolmogorov's theory. A modification, based on *conditional* probability spaces (which contains Kolmogorov's theory as a special case), has been developed by ALFRÉD RÉNYI.[51]

---

47 VON NEUMANN (1932a), pp. 595–598. Compare also BIRKHOFF & NEUMANN (1936), p. 825.

48 Compare GNEDENKO & KOLMOGOROV (1954), §3.

49 If $\boldsymbol{\Omega}$ is a topological space, then the smallest σ-algebra with respect to which all continuous complex-valued functions on $\boldsymbol{\Omega}$ are measurable, is called the *Baire σ-algebra* of $\boldsymbol{\Omega}$. The smallest σ-algebra containing all open sets of $\boldsymbol{\Omega}$ is called the *Borel σ-algebra* of $\boldsymbol{\Omega}$. In general, the Baire σ-algebra is contained in the Borel σ-algebra. If $\boldsymbol{\Omega}$ is metrisable, then the Baire and Borel σ-algebras coincide. Compare Bauer (1974), theorem 40.4, p. 198.

50 A polish space is a separable topological space that can be metrized by means of a complete metric; compare COHN (1980), chapt. 8. For a review of probability theory on complete separable metric spaces, compare PARTHASARATHY (1967). For a discussion of Radon measures on arbitrary topological spaces, compare SCHWARTZ (1973). For a critical review of Kolmogorov's axioms, compare FORTET (1958), LORENZEN (1978).

51 RÉNYI (1955). Compare also chapter 2 in the excellent textbook by RÉNYI (1970b).

2.3 Random variables in the sense of Kolmogorov

In probability theory observable quantities of a statistical experiment are called *statistical observables*. In Kolmogorov's mathematical probability theory statistical observables are represented by $\boldsymbol{\Sigma}$-measurable functions on the sample space $\boldsymbol{\Omega}$. The more precise formulation goes as follows. The Borel σ-algebra $\Sigma_{\mathbb{R}}$ of subsets of $\mathbb{R}$ is the σ-algebra generated by the open subsets of $\mathbb{R}$. The members of $\Sigma_{\mathbb{R}}$ are called Borel subsets of $\mathbb{R}$. In Kolmogorov's set-theoretical formulation, a *statistical observable* is a σ-homomorphism $\xi : \Sigma_{\mathbb{R}} \to \boldsymbol{\Sigma}/\boldsymbol{\Delta}$. In this formulation, every observable ξ can be induced by a real-valued Borel function $x : \boldsymbol{\Omega} \to \mathbb{R}$ via the inverse map[52]

$$\xi(\mathcal{R}) := x^{-1}(\mathcal{R}) := \{\boldsymbol{\omega} \in \boldsymbol{\Omega} | x(\boldsymbol{\omega}) \in \mathcal{R}\} \quad , \quad \mathcal{R} \in \Sigma_{\mathbb{R}} \quad .$$

In mathematical probability theory a real-valued Borel function x defined on $\boldsymbol{\Omega}$ is said to be a real-valued *random variable*.[53] Every statistical observable is induced by a random variable, but an observable (that is, a σ-homomorphism) defines only an *equivalence class* of random variables which induce this homomorphism. Two random variables x and y are said to be equivalent if they are equal $\boldsymbol{\mu}$-almost everywhere,[54]

$$x(\boldsymbol{\omega}) \sim y(\boldsymbol{\omega}) \;\Leftrightarrow\; \boldsymbol{\mu}\{\boldsymbol{\omega} \in \boldsymbol{\Omega} | x(\boldsymbol{\omega}) \neq y(\boldsymbol{\omega})\} = 0 \quad .$$

That is, for a statistical description is not necessary to know the point function $\boldsymbol{\omega} \mapsto x(\boldsymbol{\omega})$, it is sufficient to know the observable ξ, or in other words, the equivalence class $[x(\boldsymbol{\omega})]$ of the point functions, which induce the corresponding σ-homomorphism,

$$\xi \;\Leftrightarrow\; [x(\boldsymbol{\omega})] := \{y(\boldsymbol{\omega}) | y(\boldsymbol{\omega}) \sim x(\boldsymbol{\omega})\} \quad .$$

The description of a physical system in terms of an individual function $\boldsymbol{\omega} \mapsto f(\boldsymbol{\omega})$ distinguishes between different points $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ and corresponds to an *individual description* (maybe in terms of hidden variables). In contrast, a description in terms of equivalence classes of random variables does not distinguish between different points and corresponds to a *statistical ensemble description*.

If $\boldsymbol{\omega} \mapsto x(\boldsymbol{\omega})$ is a random variable on $\boldsymbol{\Omega}$, and if $\boldsymbol{\omega} \mapsto x(\boldsymbol{\omega})$ is integrable over $\boldsymbol{\Omega}$ with respect to $\boldsymbol{\mu}$, we say that the expectation of x with respect to $\boldsymbol{\mu}$ exists, and we write

$$\mathcal{E}(x) := \int_{\boldsymbol{\Omega}} x(\boldsymbol{\omega})\, \boldsymbol{\mu}(d\boldsymbol{\omega}) \quad ,$$

and call $\mathcal{E}(x)$ the *expectation value* of x. Every Borel-measurable complex-valued function $x \mapsto f(x)$ of a random variable $\boldsymbol{\omega} \mapsto x(\boldsymbol{\omega})$ on $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ is also a complex-valued random variable on $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$. If the expectation of the random variable $\boldsymbol{\omega} \mapsto f\{x(\boldsymbol{\omega})\}$ exists, then

$$\mathcal{E}(f) = \int_{\boldsymbol{\Omega}} f\{x(\boldsymbol{\omega})\}\, \boldsymbol{\mu}(d\boldsymbol{\omega}) \quad .$$

A real-valued random variable $\boldsymbol{\omega} \mapsto x(\boldsymbol{\omega})$ on a probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ induces a probability measure $\boldsymbol{\mu}_x : \Sigma_{\mathbb{R}} \to [0,1]$ on the state space $(\Sigma_{\mathbb{R}}, \mathbb{R})$ by

$$\boldsymbol{\mu}_x(\mathcal{R}) := \boldsymbol{\mu}\{x^{-1}(\mathcal{R})\} = \boldsymbol{\mu}\{\boldsymbol{\omega} \in \boldsymbol{\Omega} | x(\boldsymbol{\omega}) \in \mathcal{R}\} \quad , \quad \mathcal{R} \in \Sigma_{\mathbb{R}} \quad ,$$

so that

$$\mathcal{E}(f) = \int_{\mathbb{R}} f(x)\, \boldsymbol{\mu}_x(dx) \quad .$$

---

52 Sikorski (1949).

53 A usual but rather ill-chosen name since a "random variable" is neither a variable nor random.

54 While the equivalence of two *continuous* functions on a closed interval implies their equality, this is not true for arbitrary measurable (that is, in general, discontinuous) functions. Compare, for example Kolmogorov & Fomin (1961), p. 41.

2.4 Stochastic processes

The success of Kolmogorov's axiomatization is largely due to the fact that *it does not busy itself with chance.*[55] Probability has become a branch of pure mathematics. Mathematical probability theory is supposed to provide a model for situations involving random phenomena, but we are never told what exactly "random" conceptually means besides the fact that random events cannot be predicted exactly. Even if we have only a rough idea what we mean by "random", it is plain that Kolmogorov's axiomatization does not give sufficient conditions for characterizing random events. However, if we adopt the view proposed by Friedrich Waismann[56] and consider probability not as a property of a *given sequence* of events but as property of the *generating conditions* of a sequence then we can relate randomness with predictability and retrodictability.

A family $\{\xi(t)\,|\,t \in \mathbb{R}\}$ of statistical observables $\xi(t)$ indexed by a time parameter t is called a *stochastic process*. In the framework of Kolmogorov's probability theory a stochastic process is represented by a family $\{[x(t|\boldsymbol{\omega})]\,|\,t \in \mathbb{R}\}$ of equivalence classes $[x(t|\boldsymbol{\omega})]$ of random variables $x(t|\boldsymbol{\omega})$ on a common probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$,

$$[x(t|\boldsymbol{\omega})] := \{y(t|\boldsymbol{\omega})\,|\,y(t|\boldsymbol{\omega}) \sim x(t|\boldsymbol{\omega})\} \quad .$$

Two individual point functions $(t, \boldsymbol{\omega}) \mapsto x(t|\boldsymbol{\omega})$ and $(t, \boldsymbol{\omega}) \mapsto y(t|\boldsymbol{\omega})$ on a common probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ are said to be *statistically equivalent* (in the narrow sense), if and only if

$$\boldsymbol{\mu}\{\boldsymbol{\omega} \in \boldsymbol{\Omega}\,|\,x(t|\boldsymbol{\omega}) \neq x(t|\boldsymbol{\omega})\} = 0 \quad \text{for all } t \in \mathbb{R} \quad .$$

Some authors find it convenient to use the same symbol for functions and equivalences classes of functions. We avoid this identification, since it muddles individual and statistical descriptions. A stochastic process is not an individual function but an indexed family of σ-homomorphism $\xi(t) : \Sigma_{\mathbb{R}} \to \boldsymbol{\Sigma}/\boldsymbol{\Delta}$ which can be represented by an indexed family of *equivalence classes* of random variables. For fixed $t \in \mathbb{R}$ the function $\boldsymbol{\omega} \mapsto x(t|\boldsymbol{\omega})$ is a random variable. The point function $t \mapsto x(t|\boldsymbol{\omega})$ obtained by fixing $\boldsymbol{\omega}$ is called a *realization*, or a *sample path*, or a *trajectory* of the stochastic process $t \mapsto [x(t|\boldsymbol{\omega})]$. The description of a physical system in terms of an individual *trajectory* $t \mapsto x(t|\boldsymbol{\omega})$ ($\boldsymbol{\omega}$ fixed) of a stochastic process $\{[x(t|\boldsymbol{\omega})]\,|\,t \in \mathbb{R}\}$ corresponds to a *point dynamics*, while a description in terms of equivalence classes of trajectories and an associated probability measure corresponds to an *ensemble dynamics*.

Kolmogorov's characterization of stochastic processes as collections equivalence classes of random variables is much too general for science. Some additional regularity requirements like separability or continuity are necessary in order that the process has "nice trajectories" and does not disintegrate into an uncountable number of events. We will only discuss stochastic processes with some regularity properties, so that we can ignore the mathematical existence of nonseparable versions.

Furthermore, the traditional terminology is somewhat misleading since according to Kolmogorov's definition precisely predictable processes also are stochastic processes. However, the theory of stochastic processes provides a conceptually sound and mathematically workable distinction between so-called *singular processes* that allow a perfect prediction of any future value from a knowledge of the past values of the process, and so-called *regular processes* for which long-

---

55 Compare Aristotle's criticism in *Metaphysica*, 1064b 15: "Evidently, none of the traditional sciences busies itself about the accidental." Quoted from Ross (1924).

56 Waismann (1930).

term predictions are impossible.[57] For simplicity, we discuss here only the important special case of *stationary* processes.

A stochastic process is called *strictly stationary* if all its joint distribution functions are invariant under time translation, so that they depend only on time differences. For many applications this is too strict a definition, often it is enough to require that the mean and the covariance are time-translation invariant. A stochastic process $\{[x(t|\boldsymbol{\omega})]\,|\,t\in\mathbb{R}\}$ is said to be *weakly stationary* (or: *stationary in the wide sense*) if

- $\mathcal{E}\{x(t|\boldsymbol{\omega})^2\} < \infty$ for every $t\in\mathbb{R}$,
- $\mathcal{E}\{x(t+\tau|\cdot)\} = \mathcal{E}\{x(t|\cdot)\}$ for all $t,\tau\in\mathbb{R}$,
- $\mathcal{E}\{x(t+\tau|\cdot)\,x(t'+\tau|\cdot)\} = \mathcal{E}\{x(t|\cdot)\,x(t'|\cdot)\}$ for all $t,t',\tau\in\mathbb{R}$.

Since the covariance function of a weakly stationary stochastic process is positive definite, Bochner's theorem[58] implies Khintchin's spectral decomposition of the covariance:[59] A complex-valued function $R:\mathbb{R}\to\mathbb{C}$ which is continuous at the origin is the covariance function of a complex-valued second-order, weakly stationary and continuous (in the quadratic mean) stochastic process if and only if it can be represented in the form

$$R(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\,d\hat{R}(\lambda) \quad ,$$

where $\hat{R}:\mathbb{R}\to\mathbb{R}$ is a real, never decreasing and bounded function, called the *spectral distribution function of the stochastic process.*

Lebesgue's decomposition theorem says that every distribution function $\hat{R}:\mathbb{R}\to\mathbb{R}$ can be decomposed *uniquely* according to

$$\hat{R} = c^{\mathrm{d}}\hat{R}^{\mathrm{d}} + c^{\mathrm{s}}\hat{R}^{\mathrm{s}} + c^{\mathrm{ac}}\hat{R}^{\mathrm{ac}} \quad , \quad c^{\mathrm{d}}\ge 0 \quad , \quad c^{\mathrm{s}}\ge 0 \quad , \quad c^{\mathrm{ac}}\ge 0 \quad , \quad c^{\mathrm{d}}+c^{\mathrm{s}}+c^{\mathrm{ac}} = 1 \quad .$$

where $\hat{R}^{\mathrm{d}}, \hat{R}^{\mathrm{s}}$ and $\hat{R}^{\mathrm{ac}}$ are normalized spectral distribution functions. The function $\hat{R}^{\mathrm{d}}$ is a step function. Both functions $\hat{R}^{\mathrm{s}}$ and $\hat{R}^{\mathrm{ac}}$ are continuous, $\hat{R}^{\mathrm{s}}$ is singular and $\hat{R}^{\mathrm{ac}}$ is absolutely continuous. The absolute continuous part has a derivative almost everywhere, it is called the spectral density function $\lambda\mapsto d\hat{R}^{\mathrm{ac}}(\lambda)/d\lambda$. The Lebesgue decomposition of spectral distribution of a covariance function $t\mapsto R(t)$ induces an additive decomposition of the covariance function into a discrete distribution function $t\mapsto R^{\mathrm{d}}(t)$, a singular distribution function $t\mapsto R^{\mathrm{s}}(t)$, and an absolutely continuous distribution function $t\mapsto R^{\mathrm{ac}}(t)$. The discrete part R^{d} is almost periodic in the sense of Harald Bohr, so that its asymptotic behavior is characterized by $\limsup_{|t|\to\infty}|R^{\mathrm{d}}(t)|=1$. For the singular part the limit $\limsup_{|t|\to\infty}|R^{\mathrm{s}}(t)|$ may be any number between 0 and 1. The Riemann-Lebesgue lemma implies that for the absolutely continuous part R^{ac}, we have $\lim_{|t|\to\infty}|R^{\mathrm{ac}}(t)|=0$.

---

57 Compare for example Doob (1953), p. 564; Pinsker (1964), section 5.2; Rozanov (1967), sections II.2 and III.2. Sometimes, singular processes are called *deterministic*, and regular processes are called *purely nondeterministic*. We will not use this terminology since determinism refers to an ontic description, while the singularity or the regularity refers to epistemic predictability of the process.

58 Bochner (1932), §20. The representation theorem by Bochner (1932), §19 and §20, refers to continuous positive-definite functions. Later, Cramér (1939) showed that the continuity assumption is dispensable. Compare also Cramér & Leadbetter (1967), sect. 7.4.

59 Khintchine (1934). Often this result is called the Wiener–Khintchin-theorem but this terminology should be avoided since Khintchin's theorem relates the *ensemble averages* of the covariance and the spectral functions while the theorem by Wiener (1930), chapt. II.3, relates the autocorrelation function of a *single* function with a spectral function of a *single* function.

A strictly stationary stochastic process $\{[x(t|\boldsymbol{\omega})]\,|\,t \in \mathbb{R}\}$ is called *singular* if a knowledge of its past $\{[x(t|\boldsymbol{\omega})]\,|\,t < 0\}$ allows an error-free prediction. A stochastic process is called *regular* if it is not singular and if the conditional expectation is the best forecast. The remote past of a singular process contains already all information necessary for the exact prediction of its future behavior, while a regular process contains no components that can be predicted exactly from an arbitrary long past record. The optimal prediction of a stochastic process is in general nonlinear.[60] Up to now, there is no general workable algorithm for nonlinear prediction.[61] Most results refer to linear prediction of weakly stationary second-order processes. The famous *Wold decomposition* says that every weakly stationary stochastic process is the sum of a uniquely determined *linearly singular* and a uniquely determined *linearly regular* process.[62] A weakly stationary stochastic process $\{[x(t|\boldsymbol{\omega})]\,|\,t \in \mathbb{R}\}$ is called *linearly singular* if the optimal *linear* predictor in terms of the past $\{[x(t|\boldsymbol{\omega})]\,|\,t < 0\}$ allows an error-free prediction. If a weakly stationary stochastic process does not contain a linearly singular part, it is called *linearly regular*.

There is an important analytic criterion for the dichotomy between linearly singular and linearly regular processes, the so-called *Wiener–Krein criterion*[63]: A weakly stationary stochastic process $\{[x(t|\boldsymbol{\omega})]\,|\,t \in \mathbb{R}\}$ with mean value $\mathcal{E}\{x(t|\cdot)\} = 0$ and the spectral distribution function $\lambda \mapsto \hat{R}(\lambda)$ is linearly regular if and only if its spectral distribution function is absolutely continuous and if

$$\int_{-\infty}^{\infty} \frac{\ln\{d\hat{R}(\lambda)/d\lambda\}}{1+\lambda^2}\, d\lambda > -\infty \quad .$$

Note that for a linearly regular process the spectral distribution function $\lambda \mapsto \hat{R}(\lambda)$ is necessarily absolutely continuous so that the covariance function $t \mapsto R(t)$ vanishes for $t \to \infty$. However, there are exactly predictable stochastic processes with an asymptotically vanishing covariance function, so that *an asymptotically vanishing covariance function is not sufficient for a regular behavior.*

There is a close relationship between regular stochastic processes and the *irreversibility* of physical systems.[64] A characterization of genuine irreversibility of classical linear input–output system can be based on the entropy-free non-equilibrium thermodynamics with the notion of *lost energy* as central concept.[65] Such a system is called *irreversible* if the lost energy is strictly positive. According to a theorem by König and Tobergte[66] a linear input–output system behaves irreversible if and only if the associated distribution function fulfills the Wiener–Krein criterion for the spectral density of a linearly regular stochastic process.

---

60 Compare for example Rosenblatt (1971), section VI.2.

61 Compare also the review by Kallianpur (1961).

62 This decomposition is due to Wold (1938) for the special case of discrete-time weakly stationary processes, and to Hanner (1950) for the case of continuous-time processes. The general decomposition theorem is due to Cramér (1939).

63 Wiener (1942), republished as Wiener (1949); Krein (1945), Krein (1945). Compare also Doob (1953), p. 584.

64 Compare also Lindblad (1993).

65 Meixner (1961), Meixner (1965).

66 König & Tobergte (1963).

2.5 Birkhoff's individual ergodic theorem

A stochastic process on the probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ is called *ergodic* if its associated measure-preserving transformation $\boldsymbol{\tau}_t$ is ergodic for every $t \geq 0$ (that is, if every σ-algebra of sets in $\boldsymbol{\Sigma}$, invariant under the measure-preserving semiflow associated with the process, is trivial). According to a theorem by Wiener and Akutowicz[67] a strictly stationary stochastic process with an absolutely continuous spectral distribution function is weakly mixing, hence ergodic. Therefore, *every regular process is ergodic* so that the so-called ergodic theorems apply. Ergodic theorems provide conditions for the equality of time averages and ensemble averages. Of crucial importance for the interpretation of probability theory is the *individual* (or *pointwise*) *ergodic theorem* by George David Birkhoff.[68] The discrete version of the pointwise ergodic theorem is a generalization of the strong law of large numbers. In terms of harmonic analysis of stationary stochastic processes, this theorem can be formulated as follows.[69] Consider a strictly stationary zero-mean stochastic process $\{[x(t|\boldsymbol{\omega})] \,|\, t \in \mathbb{R}\}$ over the probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$, and let $\boldsymbol{\omega} \mapsto x(t|\boldsymbol{\omega})$ be quadratically integrable with respect to the measure $\boldsymbol{\mu}$. Then for $\boldsymbol{\mu}$-almost all $\boldsymbol{\omega}$ in $\boldsymbol{\Omega}$, every trajectory $t \mapsto x(t|\boldsymbol{\omega})$ the individual autocorrelation function $t \mapsto C(t|\boldsymbol{\omega})$,

$$C(t|\boldsymbol{\omega}) := \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} x(\tau|\boldsymbol{\omega})^* \, x(t+\tau|\boldsymbol{\omega}) \, d\tau \quad , \quad t \in \mathbb{R} \quad , \quad \boldsymbol{\omega} \text{ fixed} \quad ,$$

exists and is continuous on $\mathbb{R}$. Moreover, the autocorrelation function $t \mapsto C(t|\boldsymbol{\omega})$ equals for $\boldsymbol{\mu}$-almost all $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ the covariance function $t \mapsto R(t)$,

$$C(t|\boldsymbol{\omega}) = R(t) \quad \text{for } \boldsymbol{\mu}\text{-almost all } \boldsymbol{\omega} \in \boldsymbol{\Omega} \quad ,$$

$$R(t) := \int_{\boldsymbol{\Omega}} x(t|\boldsymbol{\omega}) \, x(0|\boldsymbol{\omega}) \, \boldsymbol{\mu}(d\boldsymbol{\omega}) \quad .$$

The importance of this relation lies in the fact that in most applications we see only a single individual trajectory, that is, a particular realization of the stochastic process. Since Kolmogorov's theory of stochastic processes refer to equivalence classes of functions Birkhoff's individual ergodic theorem provides a crucial link between the ensemble description and the individual description of chaotic phenomena. In the next chapter we will sketch two different direct approaches for the description of chaotic phenomena which avoid the use of ensembles.

---

67 Wiener & Akutowicz (1957), theorem 4.

68 Using the linearization of a classical dynamical system to Hilbert-space description introduced by Bernard Osgood Koopman (1931), Johann von Neumann (1932b) (communicated December 10, 1931, published 1932) was the first to establish a theorem bearing to the quasiergodic hypothesis: the *mean ergodic theorem* which refers to L^2-convergence. Stimulated by these ideas, one month later George David Birkhoff (1931) (communicated December 1, 1931, published 1931) obtained the even more fundamental *individual* (or *pointwise*) *ergodic theorem* which refers to pointwise convergence. As Birkhoff & Koopman (1932) explain, von Neumann communicated his results to them on October 22, 1931, and „raised at once the important question as to whether or not ordinary time means exist along the individual path-curves excepting for a possible set of Lebesgue measure zero". Shortly thereafter Birkhoff proved his individual ergodic theorem.

69 This formulation has been taken from Masani (1990), p. 139–140.

3. Individual descriptions of chaotic processes

3.1 Deterministic chaotic processes in the sense of Wiener

More than a decade before Kolmogorov's axiomatization of mathematical probability theory, Norbert Wiener invented a possibly deeper paradigm for chaotic phenomena: his mathematically rigorous analytic construction of an *individual* trajectory of Einstein's idealized Brownian motion[70] nowadays called a *Wiener process*.[71] In Wiener's mathematical model chaotic changes in direction of the Brownian path take place constantly. All trajectories of a Wiener process are almost certainly continuous but nowhere differentiable, just as conjectured by Jean Baptiste Perrin for the Brownian motion.[72] Wiener's constructions and proof are much closer to physics than Kolmogorov's abstract model, but also very intricate so that for a long time Kolmogorov's approach has been favored. Nowadays, Wiener's result can be derived in a much simpler way. The generalized derivative of the Wiener process is called "white noise" since according to the Einstein–Wiener theorem its spectral measure equals the Lebesgue measure $d(\lambda)/2\pi$. It turned out that white noise is the paradigm for a nonpredictable regular process; it serves to construct other more complicated stochastic structures.

Wiener's characterization of individual chaotic processes is founded on his basic paper "*Generalized harmonic analysis*".[73] The purpose of Wiener's generalized harmonic analysis is to give an account of phenomena which can neither be described by Fourier analysis nor by almost periodic functions. Instead of *equivalence class* of Lebesgue square summable functions, Wiener focused his harmonic analysis on *individual* Borel measurable functions $t \mapsto x(t)$ for which the *individual autocorrelation function*

$$C(t) := \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t')\, x(t+t')\, dt' \quad , \quad t \in \mathbb{R} \quad ,$$

exists and is continuous for all t. Wiener's generalized harmonic analysis of an *individual* trajectory $t \mapsto x(t)$ is in an essential way based on the spectral representation of the autocorrelation function. The Bochner–Cramér representation theorem implies that there exists a non-decreasing bounded function $\lambda \mapsto \hat{C}(\lambda)$, called the *spectral distribution function* of the individual function $t \mapsto x(t)$,

$$C(t) = \int_{-\infty}^{\infty} e^{i\lambda t} d\hat{C}(\lambda) \quad .$$

This relation is usually known under the name *individual Wiener–Khintchin theorem*.[74] However, this name is misleading. Khintchin's theorem[75] relates the covariance function and the spectral function in terms of *ensemble averages*. In contrast, Wiener's theorem[76] refers to *individual* functions. This result was already known to Albert Einstein long before.[77] The

70 Einstein (1905), Einstein (1906).
71 Wiener (1923), Wiener (1924).
72 Perrin (1906). A rigorous proof of Perrin's conjecture is due to Paley, Wiener & Zygmund (1933).
73 Wiener (1930). In his valuable commentary Pesi P. Masani (1979) stresses the importance of role of generalized harmonic analysis for the quest for randomness.
74 Compare for example Middleton (1960), p. 151.
75 Khintchine (1934).
76 Wiener (1930), chapt. II.3.
77 Einstein (1914a), Einstein (1914b).

terminology "Wiener–Khintchin theorem" caused many confusions[78] and should therefore be avoided. Here, we refer to the individual theorem as the *Einstein–Wiener theorem.* For many applications it is crucial to distinguish between the Einstein–Wiener theorem which refer to *individual functions*, and the statistical Khintchin theorem which refers to *equivalence classes of functions* as used in Kolmogorov's probability theory. *The Einstein–Wiener theorem is in no way probabilistic.* It refers to *well-defined single functions* rather than to an ensemble of functions.

If an individual function $t \mapsto x(t)$ has a pure point spectrum, it is almost periodic in the sense of Besicovitch, $x(t) \sim \sum_{j=1}^{\infty} \hat{x}_j \exp(i\lambda_j t)$. In a physical context an almost-periodic time function $x : \mathbb{R} \to \mathbb{C}$ may be considered as *predictable* since its future $\{x(t) | t > 0\}$ is completely determined by its past $\{x(t) | t \le 0\}$. If an individual function has an absolutely continuous spectral distribution, then the autocorrelation function vanishes in the limit as $t \to \infty$. The autocorrelation function $t \mapsto C(t)$ provides a measure of the memory: if the individual function $t \mapsto x(t)$ has a particular value at one moment, its autocorrelation tells us the degree to which we can guess that it will have about the same value some time later. In 1932, Koopman and von Neumann conjectured that an absolutely continuous spectral distribution function is the crucial property for the epistemically chaotic behavior of an ontic deterministic dynamical system.[79] In the modern terminology, Koopman and von Neumann refer to the so-called "mixing property". However, a rapid decay of correlations is not sufficient as a criterion for the absence of any regularity. Genuine chaotic behavior requires stronger instability properties than just mixing. If we know the past $\{x(t) | t \le 0\}$ of an individual function $t \mapsto x(t)$, then the future $\{x(t) | t > 0\}$ is completely determined if and only if the following Szegö condition for perfect linear predictability is fulfilled,[80]

$$\int_{-\infty}^{\infty} \frac{\ln\{d\hat{C}^{\mathrm{ac}}(\lambda)/d\lambda\}}{1+\lambda^2}\, d\lambda = -\infty \quad ,$$

where $\hat{C}^{\mathrm{ac}}$ is the absolutely continuous part of the spectral distribution function of the autocorrelation function of the individual function $t \mapsto x(t)$. Every individual function $t \mapsto x(t)$ with an absolutely continuous spectral distribution $\hat{C}$ fulfilling the Paley–Wiener criterion

$$\int_{-\infty}^{\infty} \frac{|\ln d\hat{C}(\lambda)/d\lambda)|}{1+\lambda^2}\, d\lambda < \infty \quad ,$$

will be called a *chaotic function in the sense of Wiener.*

Wiener's work initiated the mathematical theory of stochastic processes and functional integration. It was a precursor of the general probability measures as defined by Kolmogorov. However, it would be mistaken to believe that the theory of stochastic processes in the sense of Kolmogorov has superseded Wiener's ideas. Wiener's approach has been criticized as unnecessarily cumbersome[81] since it was based on individual functions $t \mapsto x(t)$, and not on Kolmogorov's more effortless definition of measure-theoretical stochastic processes (that is, equivalence

78 Compare for example the controversy by Brennan (1957), Brennan (1958) and Beutler (1958), Beutler (1958), with a final remark by Norbert Wiener (1958).

79 Koopman & Neumann (1932), p. 261.

80 Compare fore example Dym & McKean (1976), p. 84. Note that there are processes which are singular in the linear sense but allow a perfect *nonlinear* prediction. An example can be found in Scarpellini (1979), p. 295.

81 For example by Kakutani (1950).

classes $t \mapsto [x(t|\boldsymbol{\omega})]$). It has to be emphasized that for many practical problems *only Wiener's approach is conceptually sound*. For example, for weather prediction or anti-aircraft fire control there is no ensemble of trajectories but just a single individual trajectory from whose past behavior one would like to predict something about its future behavior.

The basic link between Wiener's individual and Kolmogorov's statistical approach is Birkhoff's individual ergodic theorem. Birkhoff's theorem implies that $\boldsymbol{\mu}$-almost every trajectory of an ergodic stochastic process on a Kolmogorov probability space $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{\mu})$ spends an amount of time in the measurable set $\mathscr{B} \in \boldsymbol{\Sigma}$ which is proportional to $\boldsymbol{\mu}(\mathscr{B})$. For $\boldsymbol{\mu}$-almost all points $\boldsymbol{\omega} \in \boldsymbol{\Omega}$, the trajectory $t \mapsto x(t|\boldsymbol{\omega})$ (with a precisely fixed $\boldsymbol{\omega} \in \boldsymbol{\Omega}$) of an ergodic regular stochastic process $t \mapsto [x(t|\boldsymbol{\omega})]$ is an individual chaotic function in the sense of Wiener. This result implies that one can switch from an *ensemble description* in terms of a Kolmogorov probability space $(\boldsymbol{\Omega},\boldsymbol{\Sigma},\boldsymbol{\mu})$ to an individual chaotic deterministic description in the sense of Wiener, and vice versa. Moreover, Birkhoff's individual ergodic theorem implies the equality

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T}^{0} x(\tau|\boldsymbol{\omega})\, x(t+\tau|\boldsymbol{\omega})\, d\tau \;=\; \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(\tau|\boldsymbol{\omega})\, x(t+\tau|\boldsymbol{\omega})\, d\tau$$

so that for ergodic processes the autocorrelation function can be evaluated in principle from observations of the *past* $\{x(t|\boldsymbol{\omega})\,|\, t \le 0\}$ of a *single* trajectory $t \mapsto x(t|\boldsymbol{\omega})$, a result of crucial importance for the prediction theory of individual chaotic processes.

3.2 Algorithmic characterization of randomness

The roots of an algorithmic definition of a random sequence can be traced to the pioneering work by Richard von Mises who proposed in 1919 his *principle of the excluded gambling system.*[82] The use of a precise concept of an algorithm has made it possible to overcome the inadequacies of the Mises' formulations. Von Mises wanted to exclude "all" gambling systems but he did not properly specify what he meant by "all". Alonzo Church pointed out that a gambling system which is not effectively calculable is of no practical use.[83] Accordingly, a gambling system has to be represented mathematically not by an arbitrary function but as an effective algorithm for the calculation of the values of a function. In accordance with von Mises' intuitive ideas and Church's refinement a sequence is called random if no player who calculates his pool by effective methods can raise his fortune indefinitely when playing on this sequence.

An adequate formalization of the notion of effective computable function was given in 1936 by Emil Leon Post and independently by Alan Mathison Turing by introducing the concept of an ideal computer nowadays called *Turing machine.*[84] A *Turing machine* is essentially a computer having an infinitely expandable memory; it is an abstract prototype of a universal digital computer and can be taken as a precise definition of the concept of an algorithm. The so-called Church–Turing thesis states that every functions computable in any intuitive sense can be computed by a by a Turing machine.[85] No example of a function intuitively considered as computable but not Turing-computable is known. According to the Church–Turing thesis a Turing machine represents the limit of computational power.

---

82 Von Mises (1919). Compare also his later books Von Mises (1928), Von Mises (1931), Von Mises (1964).

83 Church (1940).

84 Post (1936), Turing (1936).

85 Church (1936)

The idea, that the computational complexity of a mathematical object reflects the difficulty of its computation, allows to give a simple, intuitively appealing and mathematically rigorous definition of the notion of randomness of sequence. Unlike most mathematicians, Kolmogorov himself has never forgotten that the *conceptual* foundation of probability theory is wanting. He was not completely satisfied with his measure-theoretical formulation. Particularly, the exact relation between the probability measures $\boldsymbol{\mu}$ in the basic probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ and real statistical experiments remained open. Kolmogorov emphasized that

> "the application of probability theory … is always a matter of consequences of hypotheses about the impossibility of reducing in one way or another the complexity of the description of the objects in question".[86]

In 1963, Kolmogorov again took up the concept of randomness. He retracted his earlier view that "the frequency concept … does not admit a rigorous formal exposition within the framework of pure mathematics", and stated that he came "to realize that the concept of random distribution of a property in a large finite population can have a strict formal mathematical exposition."[87] He proposed a measure of complexity based on the "size of a program" which, when processed by a suitable universal computing machine, yields the desired object.[88] In 1968, Kolmogorov sketched how information theory can be founded without recourse to probability theory and in such a way that the concepts of entropy and mutual information are applicable to *individual* events (rather than to equivalence classes of random variables or ensembles). In this approach the "quantity of information" is defined in terms of storing and processing signals. It is sufficient to consider binary strings, that is, strings of bits, of zeros and ones.

The concept of algorithmic complexity allows to rephrase the old idea that "randomness consists in a lack of regularity" in a mathematically acceptable way. Moreover, a complexity measure and hence algorithmic probability refers to an *individual* object. Loosely speaking the complexity $K(x)$ of a binary string x is the size in bits of the shortest program for calculating it. If the complexity of x is not smaller than its length $\ell(x)$ then there is no simpler way to write a program for x then to write it out. In this case the string x shows no periodicity and no pattern. Kolmogorov and independently Solomonoff and Chaitin suggested that patternless finite sequences should be considered as random sequences.[89] That is, *complexity is a measure of irregularity in the sense that maximal complexity means randomness.* Therefore, it seems natural to call a binary string random if the shortest program for generating it is as long as the string itself. Since $K(x)$ is not computable, *it is not decidable whether a string is random.*

This definition of random sequences turned out not to be quite satisfactory. Using ideas of Kolmogorov, Per Martin-Löf succeeded in giving an adequate precise definition of random sequences.[90] Particularly, Martin-Löf proposed to define random sequences as those which withstand certain universal tests of randomness, defined as recursive sequential tests. Martin-Löf's random sequences fulfill all stochastic laws as the laws of large numbers, and the law of the iterated logarithm. A weakness of this definition is that Martin-Löf requires also stochastic properties that cannot be considered as physically meaningful in the sense that they cannot be tested by *computable* functions.

---

86 Kolmogorov (1983a), p. 39.

87 Kolmogorov (1963), p. 369.

88 Compare also Kolmogorov (1968a), Kolmogorov (1968b), Kolmogorov (1983a), Kolmogorov (1983b), Kolmogorov & Uspenskii (1988). For a review, compare Zvonkin & Levin (1970).

89 Compare Solomonoff (1964), Chaitin (1966), Chaitin (1969), Chaitin (1970).

90 Martin-Löf (1966), Martin-Löf (1969b).

A slightly different but more powerful variant is due to CLAUS-PETER SCHNORR.[91] He argues that a candidate for randomness must be rejected if there is an effective procedure to do so. A sequence such that no effective process can show its non-randomness must be considered as operationally random. He considers the nullsets of Martin-Löf's sequential tests in the sense of Brower (i.e. nullsets that are effectively computable) and defines a sequence to be random if it is not contained in any such nullset. Schnorr requires the stochasticity tests to be *computable* instead of being merely constructive. While the Kolmogorov-Martin-Löf approach is nonconstructive, the tests considered by Schnorr are constructive to such an extent that it is possible to approximate infinite random sequences to an arbitrary degree of accuracy by computable sequences of high complexity (pseudo-random sequences). By that, the approximation will be the better, the greater the effort required to reject the pseudo-random sequence as being truly random. The fact that the behaviour of Schnorr's random sequences can be approximated by constructive methods is of outstanding conceptual and practical importance. Random sequences in the sense of Martin-Löf do not have this approximation property, but non-approximable random sequences exist only by virtue of the axiom of choice.

A useful characterization of random sequences can be given in terms of games of chance. According to Mises' intuitive ideas and Church's refinement a sequence is called random if and only if no player who calculates his pool by effective methods can raise his fortune indefinitely when playing on this sequence. For simplicity, we restrict our discussion to, the practically important case of random sequences of the exponential type. A gambling rule implies a capital function C from the set $\mathfrak{F}$ of all finite sequences to the set $\mathbb{R}$ of all real numbers. In order that a gambler actually can use a rule, it is crucial that this rule is given algorithmically. That is, the capital function C cannot be any function $\mathfrak{F} \rightarrow \mathbb{R}$, but has to be a *computable function.*[92] If we assume that the gambler's pool is finite, and that debts are allowed, we get the following simple but rigorous characterization of a random sequence:

> A sequence $\{x_1, x_2, x_3 \cdots\}$ is a random sequence (of the exponential type) if and only if every computable capital function $C : \mathfrak{F} \rightarrow \mathbb{R}$ of bounded difference fulfills the relation $\lim_{n \to \infty} n^{-1} C\{x_1, \cdots, x_n\} = 0$.

According to Schnorr a universal test for randomness cannot exist. A sequence fails to be random if and only if there is an effective process in which this failure becomes evident. Therefore, *one can refer to randomness only with respect to a well-specified particular test.*

The algorithmic concept of random sequences can be used to derive a model for Kolmogorov's axioms (in their constructive version) of mathematical probability theory.[93] It turns out that the measurable sets form an σ algebra (in the sense of constructive set theory). This result shows the amazing insight Kolmogorov had in creating his axiomatic system.

---

91 SCHNORR (1969), SCHNORR (1970a), SCHNORR (1970b), SCHNORR (1971a), SCHNORR (1971b), SCHNORR (1973).

92 A function $C : \mathfrak{F} \rightarrow \mathbb{R}$ is called computable if there is a recursive function R such that $|R(n, w) - C(w)| < 2^{-n}$ for all $w \in \mathfrak{F}$ and all $n \in \{1, 2, 3, \cdots\}$. Recursive functions are functions computable with the aid of a Turing machine.

93 For, a review of modern algorithmic probability theory, compare SCHNORR (1971b).

4. Why are there "laws of chance"?

4.1 Laws of chance and determinism

It would be a logical mistake to assume that arbitrary chance events can be grasped by the statistical methods of mathematical probability theory. Probability theory has a rich mathematical structure so we have to ask *under what conditions* the usual "laws of chance" are valid. The modern concept of subjective probabilities presupposes a *coherent* rational behavior based on Boolean logic. That is, it is postulated that a rational man acts *as if* he had a deterministic model compatible with his preknowledge. Since also in many physical examples the appropriateness of the laws of probability can be traced back to an underlying deterministic ontic description, it is tempting to *presume* that chance events which satisfy the axioms of classical mathematical probability theory result *always* from the deterministic behavior of an underlying physical system. Such a claim cannot be demonstrated.

What *can* be proven is the weaker statement that every probabilistic system which fulfills the axioms of classical mathematical probability theory can be *embedded* into a larger deterministic system. A classical system is said to be *deterministic* if there exists a complete set of dispersionfree states such that Hadamard's principle of scientific determinism is fulfilled. Here, a state is said to be dispersionfree if every observable has a definite dispersionfree value with respect to this state. For such a deterministic system statistical states are given by mean values of dispersionfree states. A probabilistic system is said to allow *hidden variables* if it is possible to find a hypothetical larger system such that every statistical state of the probabilistic system is a mean value of dispersionfree states of the enlarged system. Since the logic of classical probability theory is a Boolean σ-algebra we can use the well-known result that a classical dynamical system is deterministic if and only if the underlying Boolean algebra is atomic.[94] As proved by Franz Kamber, every classical system characterized by a Boolean algebra allows the introduction of hidden variables such that every statistical state is a mean value of dispersionfree states.[95] This theorem implies that *random events fulfill the laws of chance if and only if they can formally be reduced to hidden deterministic events*. Such a deterministic embedding is never unique but often there is a unique minimal dilation of a probabilistic dynamical system to a deterministic one.[96] Note that the deterministic embedding is usually not constructive and that nothing is claimed about a possible ontic interpretation of hidden variables of the enlarged deterministic system.

Kolmogorov's probability theory can be viewed as a hidden variable representation of the basic abstract point-free theory. Consider the usual case where the Boolean algebra $\mathfrak{B}$ of mathematical probability theory contains no atoms. Every classical probability system $(\mathfrak{B}, p)$ can be represented in terms of some (not uniquely given) Kolmogorov space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ as a σ-complete Boolean algebra $\mathfrak{B} = \boldsymbol{\Sigma}/\boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is the σ-ideal of Borel sets of $\boldsymbol{\mu}$-measure zero. The points $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ of the set $\boldsymbol{\Omega}$ correspond to two-valued individual states (the so-called atomic or pure states) of the fictitious embedding atomic Boolean algebra $\mathfrak{P}(\boldsymbol{\Omega})$ of all subsets of the point set $\boldsymbol{\Omega}$. If (as usual) the set $\boldsymbol{\Omega}$ is not countable, the atomic states are epistemically inaccessible. Measure-theoretically, an atomic state corresponding to a point $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ is represented by the Dirac measure $\boldsymbol{\delta}_{\boldsymbol{\omega}}$ at the point $\boldsymbol{\omega} \in \boldsymbol{\Omega}$, defined for every subset $\mathcal{B}$ of $\boldsymbol{\Omega}$ by $\boldsymbol{\delta}_{\boldsymbol{\omega}}(\mathcal{B}) = 1$ if $\boldsymbol{\omega} \in \mathcal{B}$ and $\boldsymbol{\delta}_{\boldsymbol{\omega}}(\mathcal{B}) = 0$ if $\boldsymbol{\omega} \notin \mathcal{B}$.

---

94 Compare for example Kronfli (1971).

95 Kamber (1964), §7, and Kamber (1965), §14.

96 For the Hilbert-space theory of such minimal dilations in Hilbert space, compare Sz.-Nagy & Foiaş (1970). More generally, Antoniou & Gustafson (1997) have shown that an arbitrary Markov chain can be dilated to a unique minimal deterministic dynamical system.

Every epistemically accessible state can be described by a probability density $f \in L^1(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ which can be represented as an *average* of epistemically inaccessible atomic states,

$$f(\boldsymbol{\omega}) = \int_{\boldsymbol{\Omega}} f(\boldsymbol{\omega}')\, \boldsymbol{\delta}_{\boldsymbol{\omega}}(d\boldsymbol{\omega}') \quad .$$

The set-theoretical representation of the basic Boolean algebra $\mathfrak{B}$ in terms of a Kolmogorov probability space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ is mathematically convenient since it allows to relate an epistemic dynamics $t \mapsto f_t$ in terms of a probability density $f_t \in L^1(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ to a fictitious *deterministic* dynamics for the points $t \mapsto \boldsymbol{\omega}_t \in \boldsymbol{\Omega}$ by $f_t(\boldsymbol{\omega}) = f(\boldsymbol{\omega}_{-t})$.[97] It is also physically interesting since *all known context-independent physical laws are deterministic and formulated in terms of pure states.* In contrast, *every statistical dynamical law depends on some phenomenological constants* (like the half-time constants[98] in the exponential decay law for the spontaneous decay of a radioactive nucleus). That is, we can formulate *context-independent* laws only if we introduce atomic states.

4.2 Quantum mechanics does not imply an ontological indeterminism

Although it is in general impossible to predict an individual quantum event, in an ontic description the most fundamental law-statements of quantum theory are deterministic. Yet, probability is an essential element in every *epistemic* description of quantum events, but does not indicate an incompleteness of our knowledge. The context-independent laws of quantum mechanics (which necessarily have to be formulated in an *ontic* interpretation) are strictly deterministic but refer to a *non-Boolean* logical structure of reality. On the other hand, every experiment ever performed in physics, chemistry and biology has a *Boolean* operational description. The reason for this situation is enforced by the necessity to communicate about facts in an unequivocal language.

The epistemically irreducible probabilistic structure of quantum theory is induced by the interaction of the quantum object system with an external classical observing system. Quantum mechanical probabilities do not refer to the object system but to the state transition induced by the interaction of the object system with the measuring apparatus. The nonpredictable outcome of a quantum experiment is related to the projection of the atomic non-Boolean lattice of the ontic description of the deterministic reality to the atomfree Boolean algebra of the epistemic description of a particular experiment. The restriction of an ontic atomic state (which gives a complete description of the non-Boolean reality) to a Boolean context is no longer atomic but is given by a probability measure. The measure generated in this way is a conditional probability which refers to the state transition induced by the interaction. Such quantum-theoretical probabilities cannot be attributed to the object system alone; they are conditional probabilities where the condition is given by experimental arrangement. The epistemic probabilities depend on the experimental arrangement but, for a fixed context, they are objective since the underlying ontic structure is deterministic. Since a quantum-theoretical probability refers to a singled out classical experimental context, it corresponds exactly to the mathematical probabilities of

---

97 For example, every continuous *regular* Gaussian stochastic processes can be generated by an *deterministic* conservative and reversible linear Hamiltonian system with an infinite-dimensional phase space. For an explicit construction, compare for instance Picci (1986), Picci (1988).

98 These decay constants are not "an invariable property of the nucleus, unchangeable by any external influences", (as claimed by Max Born (1949), p. 172), but depend for example on the degree of ionization of the atom.

Kolmogorov's set-theoretical probability theory.[99] Therefore, a non-Boolean generalization of probability theory is not necessary since all these measures refer to a Boolean context. The various theorems which show that it is impossible in quantum theory to introduce hidden variables only say that it is impossible to embed quantum theory into a deterministic Boolean theory.[100]

4.3 Chance events for which the traditional "laws of chance" do not apply

Conceptually, quantum theory does not require a generalization of the traditional Boolean probability theory. Nevertheless, mathematicians created a non-Boolean probability theory by introducing a measure on the orthomodular lattice of projection operators on the Hilbert space of quantum theory.[101] The various variants of a *non-Boolean probability theory* are of no conceptual importance for quantum theory, but they show that genuine and interesting generalizations of traditional probability theory are possible.[102] At present there are few applications. But, if we find empirical chance phenomena with a nonclassical statistical behavior, the relevance of a non-Boolean theory should be considered. Worth mentioning are the non-Boolean pattern recognition methods[103], the attempt to develop a non-Boolean information theory[104], and speculations on the mind-body relation in terms of non-Boolean logic[105].

From a logical point of view the existence of irreproducible unique events cannot be excluded. For example, if we deny a strict determinism on the ontological level of a Boolean or non-Boolean reality, then there are no reasons to expect that every chance event is governed by statistical laws of any kind. Wolfgang Pauli made the inspiring proposal to characterize unique events by the absence of any type of statistical regularity:

> „Die von [Jung] betrachteten Synchronizitätsphänomene … entziehen sich der Einfangung in Natur-*'Gesetze'*, da sie nicht reproduzierbar, d.h. einmalig sind und durch die Statistik grosser Zahlen *verwischt* werden. In der Physik dagegen sind die 'Akausalitäten' gerade durch statistische Gesetze (grosse Zahlen) *erfassbar*." [106]

Acknowledgement

Thanks are due to Harald Atmanspacher and Werner Ehm for clarifying discussions and a careful reading of a draft of this paper.

---

99 In quantum theory, a Boolean context is described by a commutative W*-algebra which can be generated by a single selfadjoint operator, called the observable of the experiment. The expectation value of the operator-valued spectral measure of this observable is exactly the probability measure for the statistical description of the experiment in terms of a classical Kolmogorov probability space.

100 The claim by Hans Reichenbach (1949) (p. 15), "dass das Kausalprinzip in keiner Weise mit der Physik der Quanta verträglich ist", is valid only if one restricts arbitrarily the domain of the causality principle to Boolean logic.

101 For an introduction, compare Jauch (1974) and Beltrametti & Cassinelli (1981), chapters 11 and 26.

102 Compare for example Gudder & Hudson (1978).

103 Compare Watanabe (1967), Watanabe (1969b), Schadach (1973). For a concrete application of non-Boolean pattern recognition for medical diagnosis, compare Schadach (1973).

104 Compare for example Watanabe (1969a), chapter 9.

105 Watanabe (1961).

106 Letter of June 3, 1952, by Wolfgang Pauli to Markus Fierz, quoted from Von Meyenn (1996), p. 634.

References

Antoniou, I. & Gustafson, K. (1997): *From irreversible Markov semigroups to chaotic dynamics.* Physica **A 236**, 296–308.

Ayer, A. J. (1957): *The conception of probability as a logical relation.* In: S. Körner (ed.): *Observation and Interpretation in the Philosophy of Physics*. New York: Dover Publications. Pp.12–17.

Beltrametti, E. G. & Cassinelli, G. (1981): *The Logic of Quantum Mechanics.* London: Addison-Wesley.

Bernoulli, J. (1713): *Ars conjectandi.* Basel. German translation by R. Haussner under the title of *Wahrscheinlichkeitsrechnung*, Leipzig: Engelmann, 1899.

Beutler, F.J. (1958): *A further note on differentiability of autocorrelation functions.* Proceedings of the Institute of Radio Engineers **45**, 1740.

Beutler, F.J. (1958): *A further note on differentiability of autocorrelation functions. Author's comments.* Proceedings of the Institute of Radio Engineers **46**, 1759–1760.

Birkhoff, G. & Neumann, J. von (1936): *The logic of quantum mechanics.* Annals of Mathematics **37**, 823–843.

Birkhoff, G. D. (1931): *Proof of the ergodic theorem.* Proceedings of the National Academy of Sciences of the United States of America **17**, 656–660.

Birkhoff, G. D. & Koopman, B. O. (1932): *Recent contributions to the ergodic theory.* Proceedings of the National Academy of Sciences of the United States of America **18**, 279 – 282.

Bochner, S. (1932): *Vorlesungen über Fouriersche Integrale.* Leipzig: Akademische Verlagsgesellschaft.

Boole, G. (1854): *An Investigation of the Laws of Thought.* London: Macmillan. Reprint: Dover Publication, New York, 1958.

Born, M. (1949): *Einstein's statistical theories.* In: P. A. Schilpp (ed.): *Albert Einstein: Philosopher–Scientist.* Evanston, Illinois: Library of Living Philosophers. Pp.163–177.

Born, M. (1955a): *Ist die klassische Mechanik wirklich deterministisch?* Physikalische Blätter **11**, 49–54.

Born, M. (1955b): *Continuity, determinism and reality.* Danske Videnskabernes Selskab Mathematisk Fysiske Meddelelser **30**, No. 2, pp. 1–26.

Boyd, R. (1972): *Determinism, laws, and predictability in principle.* Philosophy of Science **39**, 431–450.

Breiman, L. (1968): *Probability.* Reading Massachusetts: Addison-Wesley.

Brennan, D.G. (1957): *Smooth random functions need not have smooth correlation functions.* Proceedings of the Institute of Radio Engineers **45**, 1016–1017.

Brennan, D.G. (1958): *A further note on differentiability of autocorrelation functions.* Proceedings of the Institute of Radio Engineers **46**, 1758–1759.

Carnap, R. (1945): *The two concepts of probability.* Philosophy and Phenomenological Research **5**, 513–532.

Carnap, R. (1950): *Logical Foundations of Probability.* Chicago: University of Chicago Press. 2nd ed. 1962.

Carnap, R. (1952): *The Continuum of Inductive Methods.* Chicago: University of Chicago Press.

Carnap, R. (1963): *Intellectual autobiography.* In: P. A. Schilpp (ed.): *The Philosophy of Rudolf Carnap*. La Salle, Illinois: Open Court. Pp.1–84.

Carnap, R. & Jeffrey, R. C. (1971): *Studies in Inductive Logic and Probability.* Volume I. Berkeley: University of California Press.

Chaitin, G. (1966): *On the length of programs for computing finite binary sequences.* Journal of the Association for Computing Machinery **13**, 547–569.

Chaitin, G. (1969): *On the length of programs for computing finite binary sequences: Statistical considerations.* Journal of the Association for Computing Machinery **16**, 143–159.

Chaitin, G. (1970): *On the difficulty of computations.* IEEE Transactions on Information Theory **IT-16**, 5–9.

Church, A. (1936): *An unsolvable problem of elementary number theory.* The American Journal of Mathematics **58**, 345–363.

Church, A. (1940): *On the concept of a random sequence.* Bulletin of the American Mathematical Society **46**, 130–135.

Cohn, D. L. (1980): *Measure Theory.* Boston: Birkhäuser.

Cournot, A. A. (1843): *Exposition de la théorie des chances et des probabilitiés.* Paris.

Cramér, H. (1939): *On the representation of a function by certain Fourier integrals.* Transactions of the American Mathematical Society **46**, 191–201.

Doob, J. L. (1953): *Stochastic Processes.* New York: Wiley.

Dym, H. & McKean, H. P. (1976): *Gaussian Processes, Function Theory, and the Inverse Spectral Problem.* New York: Academic Press.

Earman, J. (1986): *A Primer on Determinism.* Dordrecht: Reidel.

Einstein, A. (1905): *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierter Teilchen.* Annalen der Physik **17**, 549–560.

Einstein, A. (1906): *Zur Theorie der Brownschen Bewegung.* Annalen der Physik **19**, 371–381.

Einstein, A. (1914a): *Méthode pour la détermination de valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières.* Archives des sciences physiques et naturelles **37**, 254–256.

Einstein, A. (1914b): *Eine Methode zur statistischen Verwertung von Beobachtungen scheinbar unregelmässig quasiperiodisch verlaufender Vorgänge.* Unpublished Manuscript. Reprinted in: M. J. Klein, A. J. Kox, J. Renn & R. Schulmann (eds.). *The Collected Papers of Albert Einstein. Volume 4. The Swiss Years, 1912–1914*. Princeton: Princeton University Press. 1995. Pp.603–607.

Feigl, H. (1953): *Readings in the Philosophy of Science.* In: H. Feigl & M. Brodbeck (eds.): *Notes on causality*. New York: Appleton-Century-Crofts.

Fine, T. L. (1973): *Theories of Probability. An Examination of Foundations.* New York: Academic Press.
Finetti, B. de (1937): *La prévision: ses lois logiques, ses sources subjectives.* Annales de l'Institut Henri Poincaré 7, 1–68.
Finetti, B. de (1972): *Probability, Induction and Statistics. The Art of Guessing.* London: Wiley.
Finetti, B. de (1974): *Theory of Probability. A Critical Introductory Treatment.* Volume 1. London: Wiley.
Finetti, B. de (1975): *Theory of Probability. A Critical Introductory Treatment.* Volume 2. London: Wiley.
Fortet, R. (1958): *Recent advances in probability theory. Surveys in Applied Mathematics. IV. Some Aspects of Analysis and Probability*. New York: Wiley. Pp.169 – 240.
Gibbs, J. W. (1902): *Elementary Principles in Statistical Mechanics.* New Haven: Yale University Press.
Gillies, D. A. (1973): *An Objective Theory of Probability.* London: Methuen.
Gnedenko, B. V. & Kolmogorov, A. N. (1954): *Limit Distributions for Sums of Independent Random Variables.* Reading, Massacusetts: Addision-Wesley.
Good, I. J. (1965): *The Estimation of Probabilities. An Essay on Modern Bayesian Methods.* Cambridge, Massachusetts: MIT Press.
Gudder, S. P. & Hudson, R. L. (1978): *A noncommutative probability theory.* Transactions of the American Mathematical Society **245**, 1–41.
Haas, A. (1936): *Commentary of the Scientific Writings of J. Willard Gibbs.* New Haven: Yale University Press.
Halmos, P. R. (1944): *The foundations of probability.* American Mathematical Monthly **51**, 493–510.
Hanner, O. (1950): *Deterministic and non-deterministic stationary random processes.* Arkiv för Matematik **1**, 161–177.
Hille, E. & Phillips, R. S. (1957): *Functional Analysis and Semi-groups.* Providence, Rhode Island: American Mathematical Society.
Jauch, J. M. (1974): *The quantum probability calculus.* Synthese **29**, 131–154.
Jeffrey, R. C. (1965): *The Logic of Decision.* New York: McGraw-Hill.
Jeffreys, H. (1939): *Theory of Probability.* Oxford: Clarendon Press. 2nd ed. 1948, 3rd ed. 1961.
Kakutani, S. (1950): *Review of „Extrapolation, Interpolation and Smoothing of Stationary Time Series" by Norbert Wiener.* Bulletin of the American Mathematical Society **56**, 378–381.
Kallianpur, G. (1961): *Some ramifications of Wiener's ideas on nonlinear prediction.* In: P. Masani (ed.): *Norbert Wiener. Collected Works with Commentaries. Volume III.* Cambridge, Massachusetts: MIT Press. Pp.402–424.
Kamber, F. (1964): *Die Struktur des Aussagenkalküls in einer physikalischen Theorie.* Nachrichten der Akademie der Wissenschaften, Göttingen. Mathematisch Physikalische Klasse **10**, 103–124.
Kamber, F. (1965): *Zweiwertige Wahrscheinlichkeitsfunktionen auf orthokomplementären Verbänden.* Mathematische Annalen **158**, 158–196.
Kappos, D. A. (1969): *Probability Algebras and Stochastic Spaces.* New York: Academic Press.
Keynes, J. M. (1921): *A Treatise on the Principles of Probability.* London: Macmillan.
Khintchine, A. (1934): *Korrelationstheorie der stationären stochastischen Prozesse.* Mathematische Annalen **109**, 604–615.
Khrennikov, A. (1994): *p-Adic Valued Distributions in Mathematical Physics.* Dordrecht: Kluwer.
Kolmogoroff, A. (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: Springer.
Kolmogoroff, A. (1948): *Algèbres de Boole métriques complètes.* VI. Zjazd Matematyków Polskich. Annales de la Societe Polonaise de Mathematique **20**, 21–30.
Kolmogorov, A. N. (1963): *On tables of random numbers.* Sankhyá. The Indian Journal of Statistics **A 25**, 369–376.
Kolmogorov, A. N. (1968a): *Three approaches to the quantitative definition of information.* International Journal of Computer Mathematics **2**, 157-168. Russian original in: Problemy Peredachy Informatsii **1**, 3–11 (1965).
Kolmogorov, A. N. (1968b): *Logical basis for information theory and probability theoy.* IEEE Transactions on Information Theory **IT-14**, 662–664.
Kolmogorov, A. N. (1983a): *Combinatorial foundations of information theory and the calculus of probability.* Russian Mathematical Surveys **38:4**, 29–40.
Kolmogorov, A. N. (1983b): *On logical foundations of probability theory. Probability Theory and Mathematical Statistics. Lecture Notes in Mathematics*. Berlin: Springer. Pp.1–5.
Kolmogorov, A. N. & Fomin, S. V. (1961): *Elements of the Theory of Functions and Functional Analysis. Volume 2. Measure. The Lebesgue Integral. Hilbert Space.* Albany: Graylock Press.
Kolmogorov, A. N. & Uspenskii, V. A. (1988): *Algorithms and randomness.* Theory of Probability and its Applications **32**, 389–412.
König, H. & Tobergte, J. (1963): *Reversibilität und Irreversibilität von linearen dissipativen Systemen.* Journal für die reine und angewandte Mathematik **212**, 104–108.
Koopman, B. O. (1931): *Hamiltonian systems and transformations in Hilbert space.* Proceedings of the National Academy of Sciences of the United States of America **17**, 315–318.
Koopman, B. O. (1940a): *The bases of probability.* Bulletin of the American Mathematical Society **46**, 763–774.
Koopman, B. O. (1940b): *The axioms and algebra of intuitive probability.* Annals of Mathematics **41**, 269–292.
Koopman, B. O. (1941): *Intuitive probabilities and sequences.* Annals of Mathematics **42**, 169–187.
Koopman, B. O. & Neumann, J. von (1932): *Dynamical systems of continuous spectra.* Proceedings of the National Academy of Sciences of the United States of America **18**, 255–263.
Krein, M. G. (1945): *On a generalization of some investigations of G. Szegö, W. M. Smirnov, and A. N. Kolmogorov.* Doklady Akademii Nauk, SSSR **46**, 91–94, [in Russian].
Krein, M. G. (1945): *On a problem of extrapolation of A. N. Kolmogorov.* Doklady Akademii Nauk, SSSR **46**, 306–309, [in Russian].

Kronfli, N. S. (1971): *Atomicity and determinism in Boolean systems.* International Journal of Theoretical Physics **4**, 141–143.

Kyburg, H. E. & Smokler, H. E. (1964): *Studies in Subjective Probability.* New York: Wiley.

Laha, R. G. & Rohatgi, V. K. (1979): *Probability Theory.* New York: Wiley.

Laplace, P. S. (1814): *Essai Philosophique sur les Probabilités.* English translation from the sixth French edition under the title: *A Philosophical Essay on Probabilities*. 1951. New York: Dover Publications.

Laue, M. von (1955): *Ist die klassische Physik wirklich deterministisch?* Physikalische Blätter **11**, 269–270.

Lindblad, G. (1993): *Irreversibility and randomness in linear response theory.* Journal of Statistical Physics **72**, 539–554.

Loomis, L. H. (1947): *On the representation of σ-complete Boolean algebras.* Bulletin of the American Mathematical Society **53**, 757–760.

Lorenzen, P. (1978): *Eine konstruktive Deutung des Dualismus in der Wahrscheinlichkeitstheorie.* Zeitschrift für allgemeine Wissenschaftstheorie **2**, 256–275.

Łoš, J. (1955): *On the axiomatic treatment of probability.* Colloquium Mathematicum (Wroclaw) **3**, 125–137.

Maistrov, L. E. (1974): *Probability Theory. A Historical Sketch.* New York: Wiley.

Martin-Löf, P. (1966): *The definition of random sequences.* Information and Control **9**, 602–619.

Martin-Löf, P. (1969a): *The literature on von Mises' kollektivs revisited.* Theoria. A Swedish Journal of Philosophy **35**, 12–37.

Martin-Löf, P. (1969b): *Algorithms and randomness.* Review of the International Statistical Institute **37**, 265–272.

Masani, P. (1979): *Commentary on the memoire [30a] on generalized harmonic analysis.* In: P. Masani (ed.): *Norbert Wiener. Collected Works with Commentaries. Volume II*. Cambridge, Massachusetts: MIT Press. Pp.333–379.

Masani, R. R. (1990): *Norbert Wiener, 1894–1964.* Basel: Birkhäuser.

Meixner, J. (1961): *Reversibilität und Irreversibilität in linearen passiven Systemen.* Zeitschrift für Naturforschung **16a**, 721–726.

Meixner, J. (1965): *Linear passive systems.* In: J. Meixner (ed.): *Statistical Mechanics of Equilibrium and Non-equilibrium*. Amsterdam: North-Holland.

Meyenn, K. von (1996): *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil I: 1950–1952.* Berlin: Springer-Verlag.

Middleton, D. (1960): *Statistical Communication Theory.* New York: MacGraw-Hill.

Mises, R. von (1919): *Grundlagen der Wahrscheinlichkeitsrechnung.* Mathematische Zeitschrift **5**, 52–99.

Mises, R. von (1928): *Wahrscheinlichkeit, Statistik und Wahrheit.* Wien: Springer.

Mises, R. von (1931): *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.* Leipzig: Deuticke.

Mises, R. von (1964): *Mathematical Theory of Probability and Statitics.* Edited and Complemented by Hilda Geiringer. New York: Academic Press.

Neumann, J. von (1932a): *Zur Operatorenmethode in der klassischen Mechanik.* Annals of Mathematics **33**, 587–642, 789-791.

Neumann, J. von (1932b): *Proof of the quasiergodic hypothesis.* Proceedings of the National Academy of Sciences of the United States of America **18**, 70–82.

Nielsen, O. E. (1997): *An Introduction to Integration and Measure Theory.* New York: Wiley.

Paley, R. E. A. C., Wiener, N. & Zygmund, A. (1933): *Notes on random functions.* Mathematische Zeitschrift **37**, 647–668.

Parthasarathy, K. R. (1967): *Probability Measures on Metric Spaces.* New York: Academic Press.

Pauli, W. (1954): *Wahrscheinlichkeit und Physik.* Dialectica **8**, 112–124.

Perrin, J. (1906): *La discontinuité de la matière.* Revue du mois **1**, 323–343.

Picci, G. (1986): *Application of stochastic realization theory to a fundamental problem of statistical physics.* In: C. I. Byrnes & A. Lindquist (eds.): *Modelling, Identification and Robust Control*. Amsterdam: North-Holland. Pp.211–258.

Picci, G. (1988): *Hamiltonian representation of stationary processes.* In: I. Gohberg, J. W. Helton & L. Rodman (eds.): *Operator Theory: Advances and Applications*. Basel: Birkhäuser. Pp.193–215.

Pinsker, M.S. (1964): *Information and Information Stability of Random Variables and Processes.* San Francisco: Holden–Day.

Plato, J. von (1994): *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective.* Cambridge: Cambridge University Press.

Post, E. L. (1936): *Finite combinatory processes – formultation.* Journal of Symbolic Logic **1**, 103–105.

Prohorov, Yu. V. & Rozanov, Yu. A. (1969): *Probability Theory.* Berlin: Springer.

Reichenbach, H. (1949): *Philosophische Probleme der Quantenmechanik.* Basel: Birkhäuser.

Reichenbach, H. (1994): *Wahrscheinlichkeitslehre.* Braunschweig: Vieweg. 2. Auflage, auf Grundlage der erweiterten amerikanischen Ausgabe bearbeitet und herausgegeben von Godehard Link. Band 7 der Gesammelten Werke von Hans Reichenbach.

Rényi, A. (1955): *A new axiomatic theory of probability.* Acta Mathematica Academia Scientiarum Hungaricae **6**, 285–335.

Rényi, A. (1970a): *Probability Theory.* Amsterdam: North-Holland.

Rényi, A. (1970b): *Foundations of Probability.* San Francisco: Holden-Day.

Rosenblatt, M. (1971): *Markov Processes: Structure and Asymptotic Behavior.* Berlin: Springer.

Ross, W. D. (1924): *Aristotle's Metaphysics. Text and Commentary.* Oxford: Clarendon Press.

Rozanov, Yu. A. (1967): *Stationary Random Processes.* San Francisco: Holden-Day.

Russell, B. (1948): *Human Knowledge. Its Scope and Limits.* London: Georg Allen and Unwin.

SAVAGE, L. J. (1954): *The Foundations of Statistics.* New York: Wiley.

SAVAGE, L. J. (1962): *The Foundations of Statistical Infereence. A Discussion.* London: Methuen.

SCARPELLINI, B. (1979): *Predicting the future of functions on flows.* Mathematical Systems Theory **12**, 281–296.

SCHADACH, D. J. (1973): *Nicht-Boolesche Wahrscheinlichkeitsmasse für Teilraummethoden in der Zeichenerkennung.* In: T. EINSELE, W. GILOI & H.-H. NAGEL (eds.): *Lecture Notes in Economics and Mathematical Systems.* Vol. **83**. Berlin: Springer. Pp.29–35.

SCHEIBE, E. (1964): *Die kontingenten Aussagen in der Physik.* Frankfurt: Athenäum Verlag.

SCHEIBE, E. (1973): *The Logical Analysis of Quantum Mechanics.* Oxford: Pergamon Press.

SCHNORR, C. P (1969): *Eine Bemerkung zum Begriff der zufälligen Folge.* Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **14**, 27–35.

SCHNORR, C. P (1970a): *Über die Definition von effektiven Zufallstests.* Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **15**, 297–312, 313–328.

SCHNORR, C. P (1970b): *Klassifikation der Zufallsgesetze nach Komplexität und Ordnung.* Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **16**, 1–21.

SCHNORR, C. P (1971a): *A unified approach to the definition of random sequencies.* Mathematical System Theory **5**, 246–258.

SCHNORR, C. P (1971b): *Zufälligkeit und Wahrscheinlichkeit. Eine Algorithmische Begründung der Wahrscheinlichkeitstheorie.* Lecture Notes in Mathematics, vol. **218.** Berlin: Springer.

SCHNORR, C. P (1973): *Process complexity and effective random tests.* Journal of Computer and System Sciences **7**, 376–388.

SCHUSTER, H. G. (1984): *Deterministic Chaos. An Introduction.* Weinheim: Physik-Verlag.

SCHWARTZ, L. (1973): *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures.* London: Oxford University Press.

SCRIVEN, M. (1965): *On essential unpredictability in human behavior.* In: B. B. WOLMAN & E. NAGEL (eds.): *Scientific Psychology: Principles and Approaches*. New York: Basic Books.

SIKORSKI, R. (1949): *On inducing of homomorphism by mappings.* Fundamenta Mathematicae **36**, 7–22.

SIKORSKI, R. (1969): *Boolean Algebras.* Berlin: Springer.

SOLOMONOFF, R. J. (1964): *A formal theory of inductive inference.* Information and Control **7**, 1–22, 224–254.

STONE, M. H. (1936): *The theory of representations for Boolean algebras.* Transactions of the American Mathematical Society **40**, 37–111.

SZ.-NAGY, B. & FOIAŞ, C. (1970): *Harmonic Analysis of Operators on Hilbert Space.* Amsterdam: North-Holland.

TORNIER, E. (1933): *Grundlagen der Wahrscheinlichkeitsrechnung.* Acta Mathematica **60**, 239–380.

TURING, A. M. (1936): *On computable numbers, with an application to the Entscheidungsproblem.* Proceedings of the London Mathematical Society **42**, 230–256. Corrections: Ibid. **43** (1937) 544–546.

VENN, J. (1866): *The Logic of Chance.* London. An unaltered reprint of the third edition of 1888 appeared by Chelsea, New York, 1962.

WAISMANN, F. (1930): *Logische Analyse des Wahrscheinlichkeitsbegriffs.* Erkenntnis **1**, 228–248.

WATANABE, S. (1961): *A model of mind-body relation in terms of modular logic.* Synthese **13**, 261–302.

WATANABE, S. (1967): *Karhunen–Loève expansion and factor analysis. Theoretical remarks and applications. Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Prague, 1965).* Prague: Academia. Pp.635–660.

WATANABE, S. (1969a): *Knowing and Guessing. A Quantitative Study of Inference and Information.* New York: Wiley.

WATANABE, S. (1969b): *Modified concepts of logic, probability, and information based on generalized continuous characteristic function.* Information and Control **15**, 1–21.

WEIZSÄCKER, C. F. VON (1973): *Probability and quantum mechanics.* British Journal for the Philosophy of Science **24**, 321–337.

WEIZSÄCKER, C. F. VON (1985): *Aufbau der Physik.* München: Hanser Verlag.

WHITTAKER, E. T. (1943): *Chance, freewill and necessity in the scientific conception of the universe.* Proceedings of the Physical Society (London) **55**, 459–471.

WIENER, N. (1930): *Generalized harmonic analysis.* Acta Mathematica **55**, 117–258.

WIENER, N. (1942): *Response of a nonlinear device to noise.* Cambridge, Massachusetts: M.I.T. Radiation Laboratory. Report No. V-186. April 6, 1942.

WIENER, N. (1949): *Extrapolation, Interpolation, and Smoothing of Stationary Times Series. With Engineering Applications.* New York: MIT Technology Press and Wiley.

WIENER, N. (1958): *A further note on differentiability of autocorrelation functions.* Proceedings of the Institute of Radio Engineers **46**, 1760.

WIENER, N. & AKUTOWICZ, E. J. (1957): *The definition and ergodic properties of the stochastic adjoint of a unitary transformation.* Rendiconti del Circolo Matematico di Palermo **6**, 205–217, 349.

WOLD, H. (1938): *A Study in the Analysis of Stationary Times Series.* Stockholm: Almquist and Wiksell.

ZVONKIN, A. K. & LEVIN, L. A. (1970): *The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms.* Russian Mathematical Surveys **25**, 83–124.