Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

892

Helmut Cajar

Billingsley Dimension in Probability Spaces

Springer-Verlag
Berlin Heidelberg New York

Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

892

Helmut Cajar

Billingsley Dimension in Probability Spaces

Springer-Verlag
Berlin Heidelberg New York 1981

Author

Helmut Cajar
Mathematisches Institut A, Universität Stuttgart
Pfaffenwaldring 57, 7000 Stuttgart 80
Federal Republic of Germany

AMS Subject Classifications (1980): 10K50, 28D99, 60F15, 60J10

ISBN 3-540-11164-6 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-11164-6 Springer-Verlag New York Heidelberg Berlin

Printed in Germany

Printing and binding: Beltz Offsetdruck, Hemsbach/Bergstr.
2141/3140-543210 D 93

Lecture Notes in Mathematics

For information about Vols. 1–669, please contact your bookseller or Springer-Verlag.

Vol. 670: Fonctions de Plusieurs Variables Complexes III, Proceedings, 1977. Edité par F. Norguet. XII, 394 pages. 1978.

Vol. 671: R. T. Smythe and J. C. Wierman, First-Passage Perculation on the Square Lattice. VIII, 196 pages. 1978.

Vol. 672: R. L. Taylor, Stochastic Convergence of Weighted Sums of Random Elements in Linear Spaces. VII, 216 pages. 1978.

Vol. 673: Algebraic Topology, Proceedings 1977. Edited by P. Hoffman, R. Piccinini and D. Sjerve. VI, 278 pages. 1978.

Vol. 674: Z. Fiedorowicz and S. Priddy, Homology of Classical Groups Over Finite Fields and Their Associated Infinite Loop Spaces. VI, 434 pages. 1978.

Vol. 675: J. Galambos and S. Kotz, Characterizations of Probability Distributions. VIII, 169 pages. 1978.

Vol. 676: Differential Geometrical Methods in Mathematical Physics II, Proceedings, 1977. Edited by K. Bleuler, H. R. Petry and A. Reetz. VI, 626 pages. 1978.

Vol. 677: Séminaire Bourbaki, vol. 1976/77, Exposés 489–506. IV, 264 pages. 1978.

Vol. 678: D. Dacunha-Castelle, H. Heyer et B. Roynette. Ecole d'Eté de Probabilités de Saint-Flour. VII-1977. Edité par P. L. Hennequin. IX, 379 pages. 1978.

Vol. 679: Numerical Treatment of Differential Equations in Applications, Proceedings, 1977. Edited by R. Ansorge and W. Törnig. IX, 163 pages. 1978.

Vol. 680: Mathematical Control Theory, Proceedings, 1977. Edited by W. A. Coppel. IX, 257 pages. 1978.

Vol. 681: Séminaire de Théorie du Potentiel Paris, No. 3, Directeurs: M. Brelot, G. Choquet et J. Deny. Rédacteurs: F. Hirsch et G. Mokobodzki. VII, 294 pages. 1978.

Vol. 682: G. D. James, The Representation Theory of the Symmetric Groups. V, 156 pages. 1978.

Vol. 683: Variétés Analytiques Compactes, Proceedings, 1977. Edité par Y. Hervier et A. Hirschowitz. V, 248 pages. 1978.

Vol. 684: E. E. Rosinger, Distributions and Nonlinear Partial Differential Equations. XI, 146 pages. 1978.

Vol. 685: Knot Theory, Proceedings, 1977. Edited by J. C. Hausmann. VII, 311 pages. 1978.

Vol. 686: Combinatorial Mathematics, Proceedings, 1977. Edited by D. A. Holton and J. Seberry. IX, 353 pages. 1978.

Vol. 687: Algebraic Geometry, Proceedings, 1977. Edited by L. D. Olson. V, 244 pages. 1978.

Vol. 688: J. Dydak and J. Segal, Shape Theory. VI, 150 pages. 1978.

Vol. 689: Cabal Seminar 76–77, Proceedings, 1976–77. Edited by A.S. Kechris and Y. N. Moschovakis. V, 282 pages. 1978.

Vol. 690: W. J. J. Rey, Robust Statistical Methods. VI, 128 pages. 1978.

Vol. 691: G. Viennot, Algèbres de Lie Libres et Monoïdes Libres. III, 124 pages. 1978.

Vol. 692: T. Husain and S. M. Khaleelulla, Barrelledness in Topological and Ordered Vector Spaces. IX, 258 pages. 1978.

Vol. 693: Hilbert Space Operators, Proceedings, 1977. Edited by J. M. Bachar Jr. and D. W. Hadwin. VIII, 184 pages. 1978.

Vol. 694: Séminaire Pierre Lelong – Henri Skoda (Analyse) Année 1976/77. VII, 334 pages. 1978.

Vol. 695: Measure Theory Applications to Stochastic Analysis, Proceedings, 1977. Edited by G. Kallianpur and D. Kölzow. XII, 261 pages. 1978.

Vol. 696: P. J. Feinsilver, Special Functions, Probability Semigroups, and Hamiltonian Flows. VI, 112 pages. 1978.

Vol. 697: Topics in Algebra, Proceedings, 1978. Edited by M. F. Newman. XI, 229 pages. 1978.

Vol. 698: E. Grosswald, Bessel Polynomials. XIV, 182 pages. 1978.

Vol. 699: R. E. Greene and H.-H. Wu, Function Theory on Manifolds Which Possess a Pole. III, 215 pages. 1979.

Vol. 700: Module Theory, Proceedings, 1977. Edited by C. Faith and S. Wiegand. X, 239 pages. 1979.

Vol. 701: Functional Analysis Methods in Numerical Analysis, Proceedings, 1977. Edited by M. Zuhair Nashed. VII, 333 pages. 1979.

Vol. 702: Yuri N. Bibikov, Local Theory of Nonlinear Analytic Ordinary Differential Equations. IX, 147 pages. 1979.

Vol. 703: Equadiff IV, Proceedings, 1977. Edited by J. Fábera. XIX, 441 pages. 1979.

Vol. 704: Computing Methods in Applied Sciences and Engineering, 1977, I. Proceedings, 1977. Edited by R. Glowinski and J. L. Lions. VI, 391 pages. 1979.

Vol. 705: O. Forster und K. Knorr, Konstruktion verseller Familien kompakter komplexer Räume. VII, 141 Seiten. 1979.

Vol. 706: Probability Measures on Groups, Proceedings, 1978. Edited by H. Heyer. XIII, 348 pages. 1979.

Vol. 707: R. Zielke, Discontinuous Čebyšev Systems. VI, 111 pages. 1979.

Vol. 708: J. P. Jouanolou, Equations de Pfaff algébriques. V, 255 pages. 1979.

Vol. 709: Probability in Banach Spaces II. Proceedings, 1978. Edited by A. Beck. V, 205 pages. 1979.

Vol. 710: Séminaire Bourbaki vol. 1977/78, Exposés 507–524. IV, 328 pages. 1979.

Vol. 711: Asymptotic Analysis. Edited by F. Verhulst. V, 240 pages. 1979.

Vol. 712: Equations Différentielles et Systèmes de Pfaff dans le Champ Complexe. Edité par R. Gérard et J.-P. Ramis. V, 364 pages. 1979.

Vol. 713: Séminaire de Théorie du Potentiel, Paris No. 4. Edité par F. Hirsch et G. Mokobodzki. VII, 281 pages. 1979.

Vol. 714: J. Jacod, Calcul Stochastique et Problèmes de Martingales. X, 539 pages. 1979.

Vol. 715: Inder Bir S. Passi, Group Rings and Their Augmentation Ideals. VI, 137 pages. 1979.

Vol. 716: M. A. Scheunert, The Theory of Lie Superalgebras. X, 271 pages. 1979.

Vol. 717: Grosser, Bidualräume und Vervollständigungen von Banachmoduln. III, 209 pages. 1979.

Vol. 718: J. Ferrante and C. W. Rackoff, The Computational Complexity of Logical Theories. X, 243 pages. 1979.

Vol. 719: Categorial Topology, Proceedings, 1978. Edited by H. Herrlich and G. Preuß. XII, 420 pages. 1979.

Vol. 720: E. Dubinsky, The Structure of Nuclear Fréchet Spaces. V, 187 pages. 1979.

Vol. 721: Séminaire de Probabilités XIII. Proceedings, Strasbourg, 1977/78. Edité par C. Dellacherie, P. A. Meyer et M. Weil. VII, 647 pages. 1979.

Vol. 722: Topology of Low-Dimensional Manifolds. Proceedings, 1977. Edited by R. Fenn. VI, 154 pages. 1979.

Vol. 723: W. Brandal, Commutative Rings whose Finitely Generated Modules Decompose. II, 116 pages. 1979.

Vol. 724: D. Griffeath, Additive and Cancellative Interacting Particle Systems. V, 108 pages. 1979.

Vol. 725: Algèbres d'Opérateurs. Proceedings, 1978. Edité par P. de la Harpe. VII, 309 pages. 1979.

Vol. 726: Y.-C. Wong, Schwartz Spaces, Nuclear Spaces and Tensor Products. VI, 418 pages. 1979.

Vol. 727: Y. Saito, Spectral Representations for Schrödinger Operators With Long-Range Potentials. V, 149 pages. 1979.

continuation on page 107

TABLE OF CONTENTS

Introduction

A number of authors (see, e.g., Besicovitch [7], Knichal [32], Eggleston [23], [24], Volkmann [44], [45], [47], [49] and Cigler [19]) have computed the Hausdorff dimensions (h-dim) of sets of real numbers characterized by digit properties of their g-adic representations. A detailed comparison of the results of these papers shows the following phenomenon: If the Hausdorff dimension of some non-denumerable union of sets M_α, $\alpha \in I$, of the type under consideration as well as the Hausdorff dimensions of the individual sets M_α are known, possibly from different sources, then the relation

$$\text{(SUP)} \qquad \text{h-dim}\Big(\bigcup_{\alpha\in I} M_\alpha\Big) = \sup_{\alpha\in I} \text{h-dim}(M_\alpha)$$

holds, while this equality is, in general, only true for denumerable unions.

In the papers cited above each real number $r \in [0,1]$ is expressed by its g-adic expansion

$$r = \sum_{i=1}^{\infty} \frac{e_i}{g^i} = (e_1, e_2, \ldots), \; e_i \in \{0,1, \ldots, g-1\}$$

for some fixed integer $g \geqslant 2$. The study of dimensions is not affected here by the fact that this expansion is ambiguous for denumerably many r. Then the relative frequency $h_n(r, j)$ by which the digit j occurs in the finite sequence $(e_1, e_2, \ldots, e_n)$ is introduced. The sets under investigation are always of the type which can be described in terms of the limit points of the sequence $\{(h_n(r, 0), h_n(r, 1), \ldots, h_n(r, g-1))\}_{n\in\mathbb{N}}$ of g-tuples in the Euclidean space $\mathbb{R}^g$. Volkmann [49] computed the Haudorff dimension of the smallest sets which can be characterized in this fashion.

If we consider also the relative frequencies of blocks $(j_1, j_2, \ldots, j_l)$ of digits within the sequence $(e_1, e_2, \ldots)$ simultaneously for all blocks of arbitrary length l then the limit points obtained by letting n tend to infinity may be identified with special probability measures (W-measures), to be called distribution measures. Sets of real numbers which may be characterized solely by means of distribution measures shall be called saturated sets in the sequel.

Colebrook [20] computed the Hausdorff dimension of the smallest saturated sets. His results also yield the relation (SUP) for saturated sets M whenever both sides of the equation (SUP) are known.

In the present paper the relation (SUP) is proved for arbitrary saturated sets M and arbitrary sets I of indices. Furthermore we shall replace Hausdorff dimension by a more general Billingsley dimension with respect to a non-atomic, ergodic Markov measure P over a sequence space with finite state space. The key for proving the relation (SUP) lies in the representation of the Billingsley dimension as infimum of certain μ-P-dimensions which always satisfy the equation (SUP) trivially. This infimum of the μ-P-dimension shall be investigated in Chapter I within a general frame-work. It shall be studied extensively as a dimension of its own right, to be called P-dimension. In Chapter II we shall compute the Billingsley and Hausdorff dimensions of the smallest saturated sets and establish the relation (SUP) for arbitrary saturated sets, with applications in both directions.

At the beginning of Chapter I we shall state general remarks and conventions in § 1.A. A summary of the ergodic theory of a sequence space with denumerable state space, to the extent as it is needed in the present paper, will be given in § 1.B.

Billingsley [9] uses a fixed stochastic process in order to define his dimension. In § 2 this process shall be replaced by a dimension system. A dimension system consists of a basic space X on which a sequence of decompositions is defined in such a way that each of them is a refinement of the preceding one. These decompositions, in turn, generate a σ-algebra on X. In a dimension system a μ-P-dimension (μ-P-dim(M)) is defined for arbitrary subsets M of X for any given propability measure (W-measure) μ and any non-atomic W-measure P. The infimum of all μ-P-dimensions of a set M, extended over all W-measures μ, is then called the P-dimension of M, written P-dim(M), as mentioned above. The basic properties of these concepts are stated. A number of theorems which Billingsley [10,§ 2] has shown to hold for his dimension are also valid analogously for the P-dimension. In correspondence to the elementary nature of the definitions a large part of the proofs is also elementary, and a first simple criterion for the validity of the relation (SUP) for P-dimensions (Theorem 2.7) is obtained: If there exists a W-measure μ such that, for all set M_α, the μ-P-dimension is equal to the P-dimension, then the equation

$$\text{(SUP)} \quad \text{P-dim}\Big(\bigcup_{\alpha\in I} M_\alpha\Big) = \sup_{\alpha\in I} \text{P-dim}(M_\alpha)$$

is true.

A comparison between P-dimension and Billingsley dimension relative to P is given in § 3. The P-dimension of a set is never smaller than its Billingsley dimension. However, if a dimension system satisfies a certain completeness condition which always holds for sequence spaces with denumerable state space, then both dimensions coincide (Theorem 3.3). Now it is possible to express the Hausdorff dimension as a P-dimension by means of the theorem of Wegmann [51, Satz 2], thus reaching the original problem (SUP) again. Even though one might, in the light of this section, disregard the concept of P-dimension in addition to Billingsley's dimension in many cases, it is nevertheless justified to maintain the former and to investigate it within general dimension systems on account of the elementary approach to definitions and implied properties which it provides.

In view of the criterion stated above concerning the validity of the relation (SUP) for P-dimensions it is of interest to know W-measure which are "as small as possible" or to know lower bounds for sufficiently large families of W-measures relative to the partial ordering "less or equal by dimension" on the set of all W-measures (Def. 2.6) introduced in § 2 already. For this purpose a quasimetric q on the set of all W-measures over a dimension system is introduced (Def. 4.1) and investigated in § 4.A. In addtion to results on the continuity of the μ-P-dimension and the P-dimension with respect to μ and P it is in particular shown that the family of invariant Markov measures of arbitrary order on a sequence space with finite state space is bounded from below by dimension (Theorem 4.5). This shall be of particular interest in Chapter II for establishing the relation (SUP) for saturated sets. In § 4.B we shall consider Markov kernels in order to construct lower bounds for families of W-measures in a more general (and partially more elegant) fashion.

§ 5 is not needed for Chapter II. In this section we define the P-dimension of W-measures in analogy to Kinney/Pitcher [31] who introduced a Hausdorff dimension of W-measures on the interval [0,1]. By means of §§ 2 and 4 we obtain quickly some interesting connections such as a representation of a P-dimension of invariant W-measures in terms of the P-dimension of ergodic W-measures (Theorem 5.5), which augments known representation theorems for invariant W-measures (compare, e.g., Lemma 1.2) as far as their P-dimension is concerned.

Then Chapter II deals with the Billingsley dimension (relative to a non-atomic, ergodic Markov measure P) of the saturated sets of a sequence space with finite state space. As mentioned already, Billingsley dimension may be written as P-dimension in this context such that all the tools of Chapter I are applicable.

First we deal in § 6.A with the distribution measures of a single point of the sequence space in order to be able to define exactly the saturated sets and the smallest saturated sets. Various preliminary arguments are given in § 6.B and § 6.C after which we discuss in § 6.D the problem how and to what extent the Markov measure P, assumed to be non-atomic and ergodic, may be replaced by a more general W-measure P'. § 6.E lists those functions and some of their properties which occur in the following sections.

In § 7 we first give an upper bound for the Billingsley dimension of the smallest saturated sets (Theorem 7.1). In view of the aim of establishing the relation (SUP) for saturated sets an essential role is played by a certain μ_o-P-dimension which serves as an upper bound.

In order to provide a lower bound for the Hausdorff dimension of the smallest saturated set, Colebrook [20] constructs a certain subset of it which consists of all numbers whose g-adic expansion is obtained by juxtaposition of certain specified digit blocks of growing lengths. But only those numbers contribute something to the P-dimension of the subset so constructed for which all transitions from one specified block to the next have positive P-probabilities, and thus Colebrook's procedure can not be applied here.

Therefore we construct a suitable W-measure in order to obtain a lower bound for the Billingsley dimension of the smallest saturated sets (Theorem 7.2), a procedure which is typical for the construction of Billingsley dimensions. But the W-measure which we introduce has an additional property which shall permit us in § 9.C and § 9.D to study sets characterized by the absolute non-occurrence of digits or digit blocks in addition to saturated sets. Finally we are able to compute the Billingsley dimensions of the smallest saturated sets. Now an infimum principle appears as in Colebrook [20] , but in Colebrook's paper the Hausdorff dimension of a smallest saturated set equals, up to the factor ln g, the infimum of the entropies of the distribution measures which describe the set, whereas in the study of a more general P-dimension the entropy has to be modified by a factor which depends on the distribution measure and on P. At the end of § 7 we show that even the smallest saturated sets permit non-denumerable decompositions into subsets with the same dimension.

In § 8 we determine the Billingsley dimension of an arbitrary saturated set which turns out to be equal to the supremum of the Billingsley dimensions of those smallest saturated sets which it contains. This implies immediately the relation (SUP) for the saturated sets. The results of § 3 enable us to obtain the corresponding propositions on the Hausdorff dimension of saturated subsets of the interval [0,1] with respect to g-adic expansions (Theorems 8.1 and 8.2). Thus the "maximum entropy principle" as observed by Billingsley [9] turns out to be a general "supremum infimum principle" of the Billingsley dimension of saturated sets. Furthermore this section contains examples illustrating these theorems and results on the continuity of the P-dimension with respect to the saturated set under consideration.

In § 9 we use the results at our disposal in order to describe the Billingsley or Hausdorff dimension of sets of certain types. In § 9.A we group the points of the sequence space together into saturated sets whose distribution measures are contained in, or contain a given set of W-measures or at least one W-measure from it. The sets studied in § 9.B are obtained by considering only relative frequencies of the digit blocks of some given length. By means of the results so obtained we can prove those of Besicovitch [7], Knichal [32], Eggleston [23] and [24], Volkmann [44], [45], [47], [49], Cigler [19] and Billingsley [9] and [10] as far as they are concerned with Hausdorff or Billingsley dimensions of saturated sets.

In § 9.C we consider, in addition to describing sets by distribution measures, the stipulation that finitely many blocks of digits do not occur in the sequence representing the points of the sequence space. This leads to intersections of saturated sets with Cantor-type sets. Here we extend the relation (SUP) to a class of subsets of the sequence space of which the saturated sets form a proper subclass. This enables us also to determine the Hausdorff and Billingsley dimension, respectively, of the sets studied by Volkmann [46] and [48] and by Steinfeld/Wegmann [43] by the approach taken here. Finally we determine in § 9.D the Billingsley dimension of sets which are characterized by the non-occurrence of denumerably many digit blocks.

As a whole, Chapter II and § 9, in particular, furnish results and methods by means of which the computation of the Billingsley or Hausdorff dimension of a set may be reduced in many cases to an extremal value problem with constraints which in turn may be solved, e.g., by standard methods of calculus. In this context it should be mentioned that the concepts of upper and lower noise ("bruit supêrieur", "bruit infêrieur") of a real number, as introduced by Rauzy [38] for the description of deterministic numbers, provide further examples of saturated sets. By our methods we obtain immediately a theorem of Bernay [6] according to which the set of deterministic numbers has Hausdorff dimension zero.

More general expansions of real numbers (see Galambos [26]) do not always lead to saturated sets for which the methods of Chapter II are directly applicable. Thus the results of Schweiger [40] and Schweiger/Stradner [41], investigating digit extensions of more general arithmetic transformations, are only covered in special cases by these methods. This is true, in particular, in the case of ϑ-adic digit expansions for a real number $\vartheta > 1$ with a terminating ϑ-adic representation of the number one inasmuch as, according to Cigler [18], the digits, considered as random variables on [0,1], form an ergodic Markov chain then.

With some additional effort Chapter II could be modified in order to cover Cantor series as studied by Peyrière [36]. The continued fraction algorithm, however, is too difficult to be treated by a modification of Chapter II. Nevertheless, one might expect in view of the results of Billingsley/Henningsen [13] that a theory of saturated sets should be true which is largely analogous to Chapter II. Within the scope of the present paper no attempt has been made to cover this subject. Many authors (see, e.g., R.C. Baker [2], Beardon [4], Beyer [8], Boyd [14], Hawkes [30], Nagasaka [34], Pollington [37] and others) have determined Hausdorff dimensions of sets which can not be interpreted as saturated sets. For these problems Chapter II is of little use. Perhaps parts of Chapter I might be useful in order to obtain simplified computations of the dimensions under consideration. But from the point of view taken in Chapter I saturated sets are only an example of a family of sets with the property (SUP). The paper by Baker and Schmidt [1] shows that families of sets with the property (SUP) may be obtained by means of criteria for the approximation of real numbers by algebraic numbers.

It remains to mention that the concept of P-dimension by itself is not sufficient in order to study Hausdorff measures (for the definition see Rogers [39], and for examples, see Hatano [29] and Steinfeld/Wegmann [43]). These measures may be obtained by several approaches such as metrizing the given space (see Wegmann [50]) or using the method of Sion/Willmott [42].

We should mention two directions in which the concept of P-dimension may be generalized. For one, μ-P-dimensions may be defined not only for W-measures μ but also for more general valuations μ of the cylinders which form the elements of the decompositions (compare Remark 3.2.4). In this way, for each family of valuations μ a concept of dimension is obtained as infimum of the corresponding μ-P-dimensions. Furthermore we might, instead of defining the dimension system by means of a sequence of decompositions of a given space, restrict ourselves to a subsequence, thus obtaining a new definition of dimension. This would be analogous to a generalization of the definition of Hausdorff dimension as studied by Buck [15] and [16]. We shall not go into these two kinds of generalizations, neither shall we investigate the problem under which conditions a dimension so modified coincides with the original one.

I wish to express my sincere thanks to my teacher and Ph.D. supervisor, Professor Dr. Bodo Volkmann, who suggested the subject of the present thesis and supported my work with patience. He also initiated the necessary steps for the publication of this paper as part of the Springer Lecture Notes and provided his help in the preparation of the English version of the text.

My thanks also go to my friend, Dr. Konrad Sandau, for frequent discussions from which many ideas of this paper emerged.

I am indebted to the Springer Publishing Company for having accepted this paper for the "Lecture Notes in Mathematics". Last not least, I owe thanks to Mrs. Elisabeth Schlumberger for the time consuming and strenuous effort of typing the manuscript with extreme care.

Stuttgart, July 1981 Helmut Cajar

CHAPTER I

P-dimension

§ 1. Preliminaries, notation, terminology

1.A Generalities. Definitions, theorems, lemmas, remarks and examples are numbered consecutively in each section. The number of a theorem is also designed to its corollaries. Theorems which we quote from literature are not numbered. No originality is claimed for the content of lemmas even when a proof is given. The end of each proof is marked by the symbol //. Furthermore, the following symbols are used:

$\emptyset$ for the empty set,
$\mathbb{N}$ for the set of natural numbers,
$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$,
$\mathbb{R}$ for the set of real numbers,
$\mathbb{R}^+$ for the set of positive real numbers,
$\mathbb{R}_0^+$ for the set of non-negative real numbers,
λ for Lebesgue measure on $\mathbb{R}$,
N for any subset of $\mathbb{N}$,
I for any (not necessarily denumerable) set of indices,
$|A|$ for the cardinality of a denumerable set A,
χ_A for the characteristic function of a set A.

For brevity, a sequence $\{a_n\}_{n\in\mathbb{N}}$ is also denoted by $\{a_n\}_n$ or simply by $\{a_n\}$. A denumerable set is either finite or countably infinite. Expressions like "inf" and "sup" are permitted to assume values $+\infty$ and $-\infty$. Unless stated otherwise, $\sup \emptyset = 0$ and $\inf \emptyset = +\infty$. The symbols $+\infty$ and $-\infty$ are treated by the ordinary rules of the extended real number system. Furthermore, we use the following conventions:

$$\ln 0 = -\infty, \qquad 0 \cdot \ln 0 = 0,$$

$$\frac{\ln c}{\ln c} = 1 \qquad \forall\, c \in [0,1],$$

$$\frac{\ln 0}{\ln c} = +\infty \qquad \forall\, c \in (0,1],$$

$$\frac{\ln c}{\ln 1} = +\infty \qquad \forall\, c \in [0,1),$$

where ln denotes the natural logarithm.

A metric or quasimetric is also allowed to assume the value $+\infty$. In a quasimetric space (M, δ) we denote the open ε-neighbourhood of a point $x \in M$ and of a subset H of M, respectively, by

$$U(x, \varepsilon) = U^{\delta}(x, \varepsilon) = \{y \in M \mid \delta(x, y) < \varepsilon\}$$

and by

$$U(H, \varepsilon) = U^{\delta}(H, \varepsilon) = \{y \in M \mid \exists\, x \in H : \delta(x, y) < \varepsilon\}.$$

The distance between two non-empty subsets A and B of M is denoted by

$$\delta(A, B) = \inf \{\delta(x, y) \mid x \in A, y \in B\}.$$

A metric or quasimetric δ is explicitly mentioned in topological concepts related to the topology induced by δ whenever it is not clear from the context or if it is different from the ordinary metric or quasimetric on the space considered. In this sense we use the terms δ-closed, δ-separable, or $\overline{A}^{\delta}$ for the closure of a subset A of M relative to the δ-topology.

Definitions from measure and probability theory which are not given here may be found, e.g., in Bauer [3]. A probability measure (W-measure for short) on a measurable space $(X, \underline{X})$ (where X denotes the basic set and $\underline{X}$, a σ-algebra on X) is a measure m on $(X, \underline{X})$ with $m(X) = 1$. The expression "$\forall[m]\ x \in B$" means "for m-almost all $x \in B$". The essential supremum of a non-negative measurable function f on X over a measurable set M (written $m\text{-ess}\sup_{x \in M} f(x)$) is understood to equal zero if $m(M) = 0$. The restriction of the measure m to a measurable subset B of X, defined on the σ-algebra $\{A \in \underline{X} | A \subset B\}$, is denoted by $m|_B$. If two measurable sets A and B are given, the conditional probability

$$m(A/B) = \frac{m(A \cap B)}{m(B)}$$

is defined only if $m(B) > 0$. Products involving undefined conditional probabilities are understood to be zero if they have at least one vanishing factor. Otherwise the product remains undefined. The set of all W-measures on a measurable space $(X, \underline{X})$ is denoted by Π. Thus Π is a convex subset of the linear space of all finite signed measures on $(X, \underline{X})$. The convex closure of a subset Θ of Π is denoted by $<\Theta>$ and also by $<\mu_1, \mu_2, \ldots, \mu_n>$ if $\Theta = \{\mu_1, \mu_2, \ldots, \mu_n\} \subset \Pi$ is a finite set. A convex linear combination $\Sigma_{i \in N} \alpha_i \mu_i$ ($\alpha_i \geq 0$, $\Sigma_{i \in N} \alpha_i = 1$) of W-measures $\mu_i \in \Pi$ is also defined for countably infinite index sets N and is itself a W-measure.

<u>Definition 1.1.</u> A convex linear combination $\Sigma_{i \in N} \alpha_i \mu_i$ of W-measures μ_i is called a non-trivial convex combination if $\alpha_i > 0$ for all $i \in N$. By a face of the convex set Π we mean any convex subset Θ of Π satisfying

$$\Theta \cap <\mu, \nu> \setminus \{\mu, \nu\} \neq \emptyset \Rightarrow <\mu, \nu> \subset \Theta \qquad \forall\, \mu \in \Pi\ \forall\, \nu \in \Pi.$$

By the face of a W-measure $\mu \in \Pi$, to be written as $\Sigma(\mu)$, we mean the smallest face of Π which contains μ.

<u>Remark 1.1.</u> Since the intersection of arbitrarily many faces of Π is again a face, the face of a W-measure is well-defined. It has the following properties:

(1) $\Sigma(\mu)$ is convex.

(2) $\Sigma(\mu) = \{\nu \in \Pi \mid \exists\, \nu' \in \Pi \ \ \exists\, \alpha \in (0,1) : \mu = \alpha\nu + (1 - \alpha)\nu'\}$.

(3) $\nu \in \Sigma(\mu) \Rightarrow \Sigma(\nu) \subset \Sigma(\mu)$.

(4) $\Sigma(\mu') = \Sigma(\mu) \Leftrightarrow \exists\, \nu, \nu' \in \Pi \;\; \exists\, \alpha, \beta \in (0,1) : \mu = \alpha\nu + (1-\alpha)\nu'$,
$\mu' = \beta\nu + (1-\beta)\nu'$.

(5) If μ is non-atomic (i.e. $\forall\, B \in \underline{X}\; \exists\, D \in \underline{X} : D \subset B,\ \mu(D) = \frac{1}{2}\mu(B)$) then all $\nu \in \Sigma(\mu)$ are non-atomic.

(6) $\Sigma(\mu)$ is the set of all W-measure $\nu \in \Pi$ which are absolutely continuous with respect to μ and whose density $\frac{d\nu}{d\mu}$ is μ-almost certainly bounded.

1.B The sequence space $A^{\mathbb{N}}$

For an introduction to ergodic theory the reader should consult, e.g., Billingsley [11] or Denker/Grillenberger/Sigmund [21].

Let A be a denumerable set and let $X := A^{\mathbb{N}}$ be the space of all sequences in A. The set A is called the state space and its elements, states. Subsets of X of the form

$$[\underline{b}] := [b_1, b_2, \ldots, b_l] := \{(x_1, x_2, \ldots) \in X \mid (x_1, x_2, \ldots, x_l) = \underline{b}\},$$
$$\underline{b} = (b_1, b_2, \ldots, b_l) \in A^l$$

are called cylinders of order l. For any block $\underline{b} = (b_1, b_2, \ldots, b_l) \in A^l$, the symbol $\underline{b}'$ denotes the block $\underline{b}' := (b_1, b_2, \ldots, b_{l-1}) \in A^l$. For any point $x \in X$, let $Z_o(x) = X$ and let $Z_n(x)$ be the cylinder of order n containing the point x. Similarly, we let

$$[\underline{b}] := X \text{ for all } \underline{b} = (b_1, \ldots, b_l) \in A^o.$$

By $\underline{X}$ we denote the smallest σ-algebra on X containing all cylinders of all orders. The shift T is the measurable mapping of X onto itself defined by

$$T(x_1, x_2, \ldots) := (x_2, x_3, \ldots) \qquad \forall\, (x_1, x_2, \ldots) \in X.$$

A W-measure on the measurable space $(X, \underline{X})$ is uniquely determined by its values on all cylinders. On the other hand, any set function μ defined on the cylinders and satisfying the conditions

(W1) $\Sigma_{b \in A}\mu([b]) = 1$ and

(W2) $\Sigma_{b \in A}\mu([b_1, \ldots, b_l, b]) = \mu([b_1, \ldots, b_l]) \quad \forall\, (b_1, \ldots, b_l) \in A^l \;\; \forall\, l \in \mathbb{N}$,

may be extended to a W-measure μ on $(X, \underline{X})$ in a unique manner. Instead of $\mu([\underline{b}])$ or $\mu([b_1, b_2, \ldots, b_l])$ we also write $\mu(\underline{b})$ or $\mu(b_1, b_2, \ldots, b_l)$, respectively. In this sense, the symbol $\mu(\underline{b}/\underline{b}') = \frac{\mu(b_1,\ldots,b_l)}{\mu(b_1,\ldots,b_{l-1})}$ may be interpreted as the conditional probability for the transition of the system to the state b_l after having passed through the states $b_1, b_2, \ldots, b_{l-1}$.

A W-measure μ on $(X, \underline{X})$ is called invariant if it is invariant under the shift T, i.e. if

$\mu(T^{-1}Y) = \mu(Y) \qquad \forall\, Y \in \underline{X}.$

This definition is equivalent to the following condition:

(W3) $\Sigma_{b \in A}\, \mu(b, b_1, b_2, \ldots, b_l) = \mu(\underline{b}) \quad \forall\, \underline{b} = (b_1, b_2, \ldots, b_l) \in A^l \quad \forall\, l \in \mathbb{N}.$

An invariant W-measure μ on $(X, \underline{X})$ is called ergodic if each invariant set has measure 0 or 1, i.e. if all measurable sets Y have the property:

$$Y = T^{-1}Y \;\Rightarrow\; \mu(Y) \in \{0,1\}.$$

This is equivalent to the condition

$$\mu(Y \,\Delta\, T^{-1}Y) = 0 \;\Rightarrow\; \mu(Y) \in \{0,1\} \qquad \forall\, Y \in \underline{X},$$

where Δ denotes the symmetric difference. In addition to the set Π of all W-measures on $(X, \underline{X})$ we consider the set Π_{inv} of all invariant W-measures and the set Π_{erg} of all ergodic W-measures on $(X, \underline{X})$. Clearly $\Pi_{erg} \subset \Pi_{inv} \subset \Pi$.

The entropy $E(\mu)$ of an invariant W-measure μ is defined as

$$E(\mu) := \lim_{n\to\infty} \frac{-1}{n} \sum_{\underline{b} \in A^n} \mu(\underline{b}) \ln \mu(\underline{b}).$$

The quantity $\sum_{b \in A} \mu(b) \ln \mu(b)$ is finite if and only if $E(\mu)$ is finite, in which case one has has representation

$$E(\mu) = \lim_{n\to\infty} -\Sigma_{\underline{b} \in A^n}\, \mu(\underline{b}) \ln \mu(\underline{b}/\underline{b}').$$

The following Individual Ergodic Theorem of Birkhoff is true:

<u>Theorem.</u> Let μ be an invariant W-measure on $(X, \underline{X})$ and let f be a μ-integrable function on X. Then the sequence

$$f_n(x) := \frac{1}{n} \Sigma_{i=o}^{n-1} f(T^i x)$$

of arithmetical means converges μ-almost everywhere to a μ-integrable function $f^*(x)$, and the relations

$$\int f^* d\mu = \int f\, d\mu,$$

$$\lim_{n\to\infty} \int |f_n - f^*| d\mu = 0,$$

$$f^*(Tx) = f^*(x) \qquad \forall\, [\mu]\; x \in X$$

are true.

For an ergodic measure μ the limit function f^* is constant up to a μ-null set. If $f = \chi_B$ is the characteristic function of a measurable set B of X then $f_n(x)$ is the relative frequency of the points $x, Tx, T^2x, \ldots, T^{n-1}x$ contained in B. For an ergodic measure μ one then has

$$\lim_{n\to\infty} f_n(x) \quad = \quad \mu(B) \qquad \forall\,[\mu]\; x \in X.$$

<u>Shannon-McMillan-Breiman Theorem.</u> Let $\mu \in \Pi_{erg}$ with $E(\mu) < \infty$. Then

$$\lim_{n\to\infty} \frac{-1}{n} \ln \mu(Z_n(x)) \quad = \quad E(\mu) \qquad \forall\,[\mu]\; x \in X.$$

A Bernoulli measure on $(X, \underline{X})$is defined as an invariant W-measure on $(X, \underline{X})$ relative to which T is a Bernoulli shift. In this case one has

(W4) $\mu(\underline{b}) = \Pi_{i=1}^{l} \mu(b_i) \quad \forall\, \underline{b} = (b_1, \ldots, b_l) \in A^l \quad \forall\, l \in \mathbb{N}.$

The entropy of a Bernoulli measure is equal to

$E(\mu) = -\Sigma_{b \in A}\, \mu(b) \ln \mu(b).$

By Markov measures of order l, $l \in \mathbb{N}$, we mean invariant W-measures on $(X, \underline{X})$ relative to which T is a Markov shift. (In the literature, Markov measures are generally not assumed to be invariant). Markov measures of order l are characterized by the following "Markov property":

$$\mu(b_1, \ldots, b_n) = \mu(b_1, \ldots, b_{n-1}) \frac{\mu(b_{n-l}, \ldots, b_n)}{\mu(b_{n-l}, \ldots, b_{n-1})} \quad \forall\, (b_1, \ldots, b_n) \in A^n$$

or

$$\mu([b_1, \ldots, b_n] / [b_1, \ldots, b_{n-1}]) = \mu([b_{n-l}, \ldots, b_n] / [b_{n-l}, \ldots, b_{n-1}]) \quad \forall\, n > l.$$

The conditional probability for the transition from the state b_{n-1} to the state b_n depends only on the l states $b_{n-l}, b_{n-l+1}, \ldots, b_{n-1}$ but not on additional states of the more distant past. According to our conventions we may consider Bernoulli measures as Markov measures of order 0. The entropy of a Markov measure μ of order l is

$E(\mu) = -\Sigma_{\underline{b} \in A^{l+1}}\, \mu(b) \ln \mu(\underline{b}/\underline{b}')$, provided $-\Sigma_{b \in A}\, \mu(b) \ln \mu(b) < \infty$.

In all other cases, $E(\mu) = \infty$. The Markov property implies that

(W5) $\mu(b_1, \ldots, b_{n+l}) = \mu(b_1, \ldots, b_l) \cdot \Pi_{i=1}^{n} \dfrac{(b_i, \ldots, b_{i+l})}{(b_i, \ldots, b_{i+l-1})}$

for all Markov measures of order l, where the right-hand side is understood to mean zero if one of the conditional probabilities involved is undefined; in this case one of the preceding factors is zero already. A Markov measure of order l is also uniquely determined by its values on the cylinders of the order l + 1.

In order to have a succinct representation for the value assumed by a Markov measure on a given cylinder, we agree to use the following two functions. For two W-measures μ and P and $l \in \mathbb{N}$, let

$E^l(\mu, P) := -\Sigma_{\underline{b} \in A^l}\, \mu(\underline{b}) \ln P(\underline{b}/\underline{b}'),$

where $E^l(\mu, P) = \infty$ if there exists a $\underline{b} \in A^l$ with $\mu(\underline{b}) > 0$ and $P(\underline{b}) = 0$. The function $E^l(\mu, P)$ is affine with respect to μ as long as it remains finite, i.e., for arbitrary $P, \mu, \nu \in \Pi$ with $E^l(\mu, P) < \infty$ and $E^l(\nu, P) < \infty$ the function satisfies the equation

$$E^l(\alpha\mu + (1-\alpha)\nu, P) = \alpha E^l(\mu, P) + (1-\alpha)E^l(\nu, P) \qquad \forall\, \alpha \in [0,1].$$

For any invariant W-measure μ with finite entropy it turns out that

$$E(\mu) = \lim_{l\to\infty} E^l(\mu, \mu).$$

If P is a Markov measure of order l and if μ is an invariant W-measure then

$$E^m(\mu, P) = E^{l+1}(\mu, P) \qquad \forall\, m \geqslant l+1.$$

In this case we shall write $E(\mu, P)$ for $E^{l+1}(\mu, P)$.

For any point $x \in X$ a sequence $\{h_n(x)\}_{n\in\mathbb{N}}$ of W-measures on $(X, \underline{X})$ is defined by the equations

$$h_n(x)(B) := \frac{1}{n}\Sigma_{i=o}^{n-1}\, \chi_B(T^i x) \qquad \forall\, B \in \underline{X} \qquad \forall\, n \in \mathbb{N}.$$

For a block $\underline{b} \in A^l$ we shall also write $h_n(x, \underline{b})$ or $h_n(x; b_1, \ldots, b_l)$ instead of $h_n(x)([\underline{b}])$ or $h_n(x)([b_1, \ldots, b_l])$, respectively. Clearly, $h_n(x, \underline{b})$ is the relative frequency by which the block $(b_1, \ldots, b_l)$ occurs in the sequence $(x_1,x_2,\ldots,x_{n+l-1})$, where x stands for the infinite sequence $(x_1, x_2, x_3, \ldots)$.

With this notation, a Bernoulli measure μ satisfies the relation

$$\text{(W6)} \quad \frac{-1}{n} \ln \mu(Z_n(x)) = -\Sigma_{b\in A}\, h_n(x, b) \ln \mu(b)$$

$$= E^1(h_n(x), \mu) \qquad \forall\, x \in X \quad \forall\, n \in \mathbb{N}.$$

For a Markov measure μ of order l one has

$$\text{(W7)} \quad \frac{-1}{n} \ln \mu(Z_n(x)) = \frac{-1}{n} \ln \mu(Z_l(x)) - \sum_{\underline{b}\in A^{l+1}} h_n(x, b) \ln \mu(\underline{b}/b')$$

$$= \frac{-1}{n} \ln \mu(Z_l(x)) + E^{l+1}(h_n(x), \mu) \;\forall\, x \in X \;\forall\, n \in \mathbb{N}.$$

Every ergodic measure μ satisfies the condition

$$\text{(W8)} \quad \lim_{n\to\infty} h_n(x, b) = \mu(\underline{b}) \qquad \forall\, \underline{b} \in A^l \quad \forall\, l \in \mathbb{N} \qquad \forall\, [\mu]\; x \in X.$$

If the state base A is finite then (W7) and (W8) imply the relation

$$\text{(W9)} \quad \lim_{n\to\infty} \frac{-1}{n} \ln P(Z_n(x)) = E(\mu, P) \qquad \forall\, [\mu]\; x \in X$$

for any ergodic W-measure μ and any Markov measure P.

The weak topology on the set Π of all W-measures on the sequence space $(X, \underline{X})$ is the roughest topology on Π relative to which, for any $l \in \mathbb{N}$ and each block $\underline{b} \in A^l$, the