Luigi Amerio · B. Segre (Eds.)

CIME Summer Schools

Sistemi dinamici e teoremi ergodici

21

Varenna, Italy 1960

Springer

FONDAZIONE CIME ROBERTO CONTI

Luigi Amerio • B. Segre (Eds.)

Sistemi dinamici e teoremi ergodici

Lectures given at the
Centro Internazionale Matematico Estivo (C.I.M.E.),
held in Varenna (Como), Italy,
June 2-11, 1960

Springer

FONDAZIONE CIME ROBERTO CONTI

C.I.M.E. Foundation
c/o Dipartimento di Matematica “U. Dini”
Viale Morgagni n. 67/a
50134 Firenze
Italy
cime@math.unifi.it

ISBN 978-3-642-10943-0 e-ISBN: 978-3-642-10945-4
DOI:10.1007/978-3-642-10945-4
Springer Heidelberg Dordrecht London New York

Reprint of the 1st ed. C.I.M.E., Florence, 1960
With kind permission of C.I.M.E.

Printed on acid-free paper

Springer.com

CENTRO INTERNATIONALE MATEMATICO ESTIVO (C.I.M.E)

Reprint of the 1st ed.- Varenna, Italy, June 2-11, 1960

SISTEMI DINAMICI E TEOREMI ERGODICI

P.R. Halmos: Entropy in ergodic theory 1

E. Hopf: Some topics of ergodic theory 47

J. Massera: Les equations différentielles linéaires dans les espaces de Banach 113

L. Amerio: Funzioni quasi-periodiche astratte e problemi di propagazione... 133

L. Markus: Sistemi dinamici con stabilità strutturale............................. 149

G. Prodi: Teoremi ergodici per le equazioni della idrodinamica 159

A. N. Feldzamen: The Alexandra Ionescu Tulcea proof of Mcmillan's theorem 179

CENTRO INTERNATIONALE MATEMATICO ESTIVO

(C.I.M.E)

PAUL R. HALMOS

ENTROPY IN ERGODIC THEORY

ROMA - Istituto Matematico dell' Università - 1960

P.R.Halmos

CONTENTS

PREFACE

CHAPTER I. CONDITIONAL EXPECTATION

Section 1. Definition

Section 2. Examples

Section 3: Algebraic properties

Section 4. Dominated convergence

Section 5. Conditional probability

Section 6. Jensen's inequality

Section 7. Transformations

Section 8. Lattice properties

Section 9. Finite fields

Section 10. The martingale theorem

CHAPTER II. INFORMATION

Section 11. Motivation

Section 12. Definition

Section 13. Transformations

Section 14. Information zero.

Section 15. Additivity

Section 16. Finite additivity

Section 17. Convergence

Section 18. McMillan's theorem

CHAPTER III. ENTROPY

Section 19. Definition

Section 20. Transformations

Section 21. Entropy zero

Section 22. Concavity

Section 23. Additivity

Section 24. Finite additivity

Section 25. Convergence

CHAPTER IV. APPLICATION

Section 26. Relative entropy

Section 27. Elementary properties

Section 28. Strong monotony

Section 29. Algebraic properties

Section 30. Entropy

Section 31. Generated fields

Section 32. Examples

P.R.Halmos

PREFACE

Shannon's theory of information appeared on the mathematical scene in 1948; in 1958 Kolmogorov applied the new subject to solve some relatively old problems of ergodic theory. Neither the general theory nor its special application is as well known among mathematicians as they both deserve to be; the reason, probably, is faulty communication. Most extant expositions of information theory are designed to make the subject palatable to non-mathematicians, with the result that they are full of words like "source" and "alphabet". Such words are presumed to be an aid to intuition; for the serious student, however, who is anxious to get at the root of the matter, they are more likely to be confusing than helpful. As for the recent ergodic application of the theory, the communication trouble there is that so far the work of Kolmogorov and his school exists in Doklady abstracts only, in Russian only. The purpose of these notes is to present a stop-gap exposition of some of the general theory and some of its applications. While a few of the proofs may appear slightly different from the corresponding ones in the literature, no claim is made for the novelty of the results. As a prerequisite, some familiarity with the ideas of the general theory of measure is assumed; Halmos's *Measure theory* (1950) is an adequate reference.

Chapter I begins with relatively well known facts about conditional expectations; for the benefit of the reader who does not know this technical probabilistic concept, several standard proofs are reproduced. Standard reference: Doob, *Stochastic processes* (1953). A special case of the martingale convergence theorem is proved by what is essentially Lévy's original method (*Théorie de* *l'addition des variables aléatoires* (1937)). The reader who knows the martingale theorem can skip the whole chapter, except possibly Section 9, and, in particular, equation (9.1).

Chapter II motivates and defines information. Standard reference: Khinchin, *Mathematical foundations of information theory* (1957). The more recent book of Feinstein, *Foundations of information theory* (1958), is quite technical, but highly recommended. The chapter ends with a proof McMillan's theorem (mean convergence); the reader who knows that theorem can skip the chapter after looking at it just long enough to absorb the notation. Almost everywhere convergence probably holds. A recent paper by Breiman (Ann. Math. Stat. 28 (1957) 809-811) asserts it, but that paper has an error; at the time these lines were written the correction has not appeared yet. In any case, for the ergodic application not even mean convergence is necessary; all that is needed is the convergence of the integrals, which is easy to prove directly.

Chapter III studies entropy (average amount of information); all the facts here are direct consequences of the definitions, via the machinery built up in the first two chapters.

Chapter IV contains the application to ergodic theory. In general terms, the idea is that information theory suggests a new invariant (entropy) of measure-preserving transformations. The new invariant is sharp enough to distinguish between some hitherto indistinguishable transformations (e.g., the 2-shift and the 3-shift). The original idea of using this invariant is due to Kolmogorov (Doklady 119 (1958) 861-864 and 124 (1959) 754-755). An improved version of the definition is given by Sinai (Doklady 124 (1959) 768-771), who also computes the entropy of ergodic automorphisms of the torus. The new invariant is in some respects not so sharp as older ones. Thus for instance Rokhlin (Doklady 124 (1959) 980-983) asserts that all translations (in compact abelian groups) have the same entropy (namely zero); he also begins the study of the connection between entropy and spectrum. Much remains to be done along all these lines.

CHAPTER I. CONDITIONAL EXPECTATION

SECTION 1. DEFINITION. We shall work, throughout what follows, with a fixed probability space

$$(X, \mathcal{S}, P).$$

Here X is a non-empty set, $\mathcal{S}$ is a field of subsets of X, and P is a probability measure on $\mathcal{S}$. The word "field" in these notes is an abbreviation for "collection of sets closed under the formation of complements and *countable* unions". A probability measure on a field of subsets of X is a measure P such that

$$P(X) = 1 .$$

Suppose that $\mathcal{C}$ is a subfield of $\mathcal{S}$ and f is an integrable real function on X. If

$$Q(C) = \int_C f \, dP$$

for each C in $\mathcal{C}$, then Q is a signed measure on $\mathcal{C}$, absolutely continuous with respect to P (or, rather, with respect to the restriction of P to $\mathcal{C}$). The Radon-Nikodym theorem implies the existence of an integrable function f^*, measurable $\mathcal{C}$, such that

$$Q(C) = \int_C f^* \, dP$$

for each C in $\mathcal{C}$. The function f^* is uniquely determined (to within a set of measure zero); its dependence on f and $\mathcal{C}$ is indicated by writing

$$f^* = E(f/\mathcal{C}) .$$

The function $E(f/\mathcal{C})$ is called "the conditional expectation of f with respect to $\mathcal{C}$". It is worth while to repeat the characteristic properties of conditional expectation; they are that

(1.1) $\quad E(f/\mathcal{C})$ is measurable $\mathcal{C}$

and

(1.2) $$\int_C E(f/\mathcal{C})\,dP = \int_C f\,dP$$

for each C in $\mathcal{C}$.

SECTION 2. EXAMPLES. If $\mathcal{C}$ is the largest subfield of $\mathcal{S}$, that is $\mathcal{C} = \mathcal{S}$, then f itself satisfies (1.1) and (1.2), so that

$$E(f/\mathcal{S}) = f .$$

This result has a trivial generalization: since f always satisfies (1.2) ($\int_C f\,dP = \int_C f\,dP$), it follows that if the field $\mathcal{C}$ is such that f is measurable $\mathcal{C}$, then

(2.1) $$E(f/\mathcal{C}) = f .$$

To look at the other extreme, let 2 be the smallest subfield of $\mathcal{S}$, that is the field whose only non-empty member is X . Since the only functions measurable 2 are constants, and since the only constant (in the role of $E(f/\mathcal{C})$) that satisfies (1.2) is $\int_C f\,dP$, it follows that

(2.2) $$E(f/2) = \int f\,dP .$$

The constant $\int_C f\,dP$ is sometimes called the absolute (as opposed to conditional) expectation of f , and, in that case, it is denoted by E(f).

Here is an illuminating example. Suppose that X is the unit square, with the collection of Borel sets in the role of $\mathcal{S}$ and Lebesgue measure in the role of P . We say that a set in $\mathcal{S}$ is "vertical" in case its intersection with each vertical line L in the plane is either empty or else equal to $X \cap L$. The collection $\mathcal{C}$ of all vertical sets in $\mathcal{S}$ is a subfield of $\mathcal{S}$. A function f is measurable $\mathcal{C}$ if and only if it does not depend on its second (vertical) argument; it follows easily that if f is integrable, then

$$E(f/\mathcal{C})(x, y) = \int f(x, u)\, du .$$

SECTION 3. ALGEBRAIC PROPERTIES. Conditional expectation is a generalized integral and in one form or another it has all the properties of an integral. Thus, for instance,

$$E(1/\mathcal{C}) = 1 , \tag{3.1}$$

where this equation, as well as all other asserted equations and inequalities involving conditional expectations, holds almost everywhere. (To prove (3.1), apply (2.1).) If f and g are integrable functions and if a and b are constants, then

$$E(af + bg/\mathcal{C}) = aE(f/\mathcal{C}) + bE(g/\mathcal{C}) . \tag{3.2}$$

(Proof: if C is in $\mathcal{C}$, then the integrals over C of the two sides of (3.2) are equal to each other). If $f \geqq 0$, then

$$E(f/\mathcal{C}) \geqq 0 . \tag{3.3}$$

(Proof: if $C = \{x : E(f/\mathcal{C})(x) < 0\}$, then C is in $\mathcal{C}$ and $\int_C f\, dP = 0$; this implies that $P(C) = 0$). It is a consequence of (3.3) that

$$|E(f/\mathcal{C})| \leqq E(|f|/\mathcal{C}) . \tag{3.4}$$

(Proof: both $|f| - f \geqq 0$ and $|f| + f \geqq 0$, and therefore, by (3.2) and (3.3), both $E(-f/\mathcal{C}) \leqq E(|f|/\mathcal{C})$ and $E(f/\mathcal{C}) \leqq E(|f|/\mathcal{C})$).

Conditional expectations also have the following multiplicative property: if f is integrable and if g is bounded and measurable

$\mathcal{C}$, then

(3.5) $$E(fg/\mathcal{C}) = E(f/\mathcal{C})g \ .$$

Since the right side of (3.5) is measurable $\mathcal{C}$, the thing to prove is that

(3.6) $$\int_C E(f/\mathcal{C})g \ dP = \int_C fg \ dP$$

for each C in $\mathcal{C}$. In case g is the characteristic function of a set in $\mathcal{C}$, (3.6) follows from the defining equation (1.2) for conditional expectations. This implies that (3.6) holds whenever g is a finite linear combination of such characteristic functions, and hence, by approximation, that (3.6) holds whenever g is a bounded function measurable $\mathcal{C}$.

SECTION 4. DOMINATED CONVERGENCE. The usual limit theorems for integrals also have their analogues for conditional expectations. Thus if f, g, and f_n are integrable functions, if $|f_n| \leqq g$ and $f_n \to f$ almost everywhere, then

(4.1) $$E(f_n/\mathcal{C}) \to E(f/\mathcal{C})$$

almost everywhere and, also, in the mean. For the proof, write

$$g_n = \sup (|f_n - f|, \ |f_{n+1} - f|, \ |f_{n+2} - f|, \ldots);$$

observe that the sequence $\{g_n\}$ tends monotonely to 0 almost everywhere and that $g_n \leqq 2g$. It follows that the sequence $\{E(g_n/\mathcal{C})\}$ is monotone decreasing and, therefore, has a limit h almost everywhere. Since

(4.2) $$\int h \ dP \leqq \int E(g_n/\mathcal{C}) \ dP = \int g_n \ dP \ ,$$

and since $\int g_n \, dP \to 0$, this implies that $E(g_n/\mathcal{C}) \to 0$ almost everywhere. Since, finally,

$$(4.3) \qquad |E(f_n/\mathcal{C}) - E(f/\mathcal{C})| \leqq E(|f_n - f|/\mathcal{C}) \leqq E(g_n/\mathcal{C}) ,$$

the proof of almost everywhere convergence is complete.

Mean convergence is implied by the inequality

$$(4.4) \qquad \int |E(f_n/\mathcal{C}) - E(f/\mathcal{C})| \, dP \leqq \int |f_n - f| \, dP$$

and the Lebesgue dominated convergence theorem.

SECTION 5. CONDITIONAL PROBABILITY. If A is a measurable set (that is A is in $\mathcal{S}$) and if

$$f = c(A)$$

(where c(A) is the characteristic function of A), we write

$$E(f/\mathcal{C}) = P(A/\mathcal{C}) .$$

The function $P(A/\mathcal{C})$ is called "the conditional probability of A with respect to $\mathcal{C}$". The characteristic properties of conditional probability are that

$$P(A/\mathcal{C}) \quad \text{is measurable } \mathcal{C}$$

and

$$\int_C P(A/\mathcal{C}) \, dP = P(A \cap C)$$

for each C in $\mathcal{C}$. If A is in $\mathcal{C}$, then

$$(5.1) \qquad P(A/\mathcal{C}) = c(A) ,$$

and, in any case,

$$(5.2) \qquad P(A/2) = P(A).$$

For this reason the constant P(A) is sometimes called the absolute (as opposed to conditional) probability of A.

The converse of the conclusion (5.1) is true and sometimes useful. The assertion is that if $P(A/\mathfrak{C})$ is the characteristic function of some set, say B, then

(5.3) $$A \text{ is in } \mathfrak{C}$$

(and therefore B = A). To prove this, note that

$$\int_C c(B)\, dP = P(A \cap C) ,$$

and therefore $P(A \cap C) = P(B \cap C)$ for each C in $\mathfrak{C}$. Since $P(A/\mathfrak{C})$ is measurable $\mathfrak{C}$, the set B itself belongs to $\mathfrak{C}$. It is therefore permissible to put C = B and to put C = X - B; it follows that $B \subset A$ and $A \subset B$ (almost), so that B = A (almost).

Just as conditional expectation has the properties of an integral, conditional probability has the properties of a probability measure. Thus if A is a measurable set, then

$$0 \leqq P(A/\mathfrak{C}) \leqq 1 ,$$

and if $\{A_n\}$ is a disjoint sequence of measurable sets with union A, then

$$P(A/\mathfrak{C}) = \Sigma_n P(A_n/\mathfrak{C}) .$$

SECTION 6. JENSEN'S INEQUALITY. A useful analytic property of integration is known as Jensen's inequality, which we now proceed to state and prove in its generalized (conditional) form.

A real-valued function F defined on an interval of the real line is called *convex* if

$$F(ps + qt) \leqq pF(s) + qF(t)$$

whenever s and t are in the domain of F and p and q are non-negative numbers with sum 1. It follows immediately, by induction, that if $t_1, \ldots, t_n$ are in the domain of F and $p_1, \ldots, p_n$ are non-negative numbers with sum 1, then

$$F(\Sigma_{i=1}^{n} p_i t_i) \leqq \Sigma_{i=1}^{n} p_i F(t_i) \quad . \tag{6.1}$$

Suppose now that F is a continuous convex function whose domain is a finite subinterval of $[0, \infty)$, suppose that g is a measurable function on X whose range is (almost) included in the domain of F, and suppose that $\mathcal{C}$ is an arbitrary subfield of $\mathcal{S}$. Jensen s inequality asserts that under these conditions

$$F(E(g/\mathcal{C})) \leqq E(F(g)/\mathcal{C}) \tag{6.2}$$

almost everywhere. Since g is the limit of an increasing sequence of simple functions, and since F is continuous, it is sufficient to prove (6.2) in case

$$g = \Sigma_A \, c(A) t_A$$

where the summation extends over the atoms of some finite subfield of $\mathcal{S}$. If g has this form, then

$$F(g) = \Sigma_A \, c(A) F(t_A)$$

and

$$E(g/\mathcal{C}) = \Sigma_A \, P(A/\mathcal{C}) t_A \; .$$

Since $E(F(g)/\mathcal{C}) = \Sigma_A P(A/\mathcal{C}) F(t_A)$, the inequality (6.2) is in this case a special case of (6.1).

In the extreme case, $\mathcal{C} = \mathcal{S}$, the conditional form of Jensen's inequality reduces to a triviality ($F(g) \leqq F(g)$); in the other extreme case, $\mathcal{C} = 2$, it becomes the classical absolute Jensen's inequality

$$F(\int g \, dP) \leqq \int F(g) \, dP .$$

SECTION 7. TRANSFORMATIONS. Later we shall need to know the effect of measure-preserving transformations on conditional expectations and probabilities. Suppose therefore that T is a measure-preserving transformation on X; this means that if A is measurable, then $T^{-1}A$ is measurable and

$$P(T^{-1}A) = P(A) .$$

(For present purposes T need not be invertible).

If $\mathcal{C}$ is a subfield of $\mathcal{S}$, then

$$T^{-1}\mathcal{C}$$

is the collection (field) of all sets of the form $T^{-1}C$ with C in $\mathcal{C}$; if f is a function on X, then fT is the composite of f and T. The basic change-of-variables result is that if f is integrable, then

$$\int_C f \, dP = \int_{T^{-1}C} fT \, dP$$

for each measurable set C. If, in particular, C is in $\mathcal{C}$, then

$$\int_{T^{-1}C} E(fT/T^{-1}\mathcal{C}) \, dP = \int_{T^{-1}C} fT \, dP$$

$$= \int_C f \, dP = \int_C E(f/\mathcal{C}) \, dP$$

$$= \int_{T^{-1}C} E(f/\mathcal{C})T \, dP .$$

Since both $E(fT/T^{-1}\mathcal{C})$ and $E(f/\mathcal{C})T$ are measurable $T^{-1}\mathcal{C}$, it follows that

(7.1) $$E(ft/T^{-1}\mathcal{C}) = E(f/\mathcal{C})T .$$

Since $c(A)T = c(T^{-1}A)$, this implies that

(7.2) $$P(T^{-1}A/T^{-1}\mathcal{C}) = P(A/\mathcal{C})T .$$