

UNIVERSITÀ DEGLI STUDI DI LECCE
DIPARTIMENTO DI MATEMATICA
“ENNIO DE GIORGI”

Carlo Sempi

DIPARTIMENTO DI MATEMATICA
“ENNIO DE GIORGI”
UNIVERSITÀ DEGLI STUDI DI LECCE

Introduzione alla Teoria Ergodica



Quaderno 2/2005: ISBN 88-8305-025-8

Università di Lecce - Coordinamento di SIBA

UNIVERSITÀ DEGLI STUDI DI LECCE
DIPARTIMENTO DI MATEMATICA
“ENNIO DE GIORGI”

Carlo Sempi

DIPARTIMENTO DI MATEMATICA
“ENNIO DE GIORGI”
UNIVERSITÀ DEGLI STUDI DI LECCE

Introduzione alla Teoria Ergodica



Quaderno 2/2005: ISBN 88-8305-025-8

Università di Lecce - Coordinamento di SIBA

QUADERNI DI MATEMATICA

Una pubblicazione a cura del

DIPARTIMENTO DI MATEMATICA

“ENNIO DE GIORGI”

UNIVERSITÀ DEGLI STUDI DI LECCE

Comitato di Redazione

Giuseppe De Cecco (Direttore)

Lorenzo Barone

Wenchang Chu (Segretario)

I QUADERNI del Dipartimento di Matematica dell'Università degli Studi di Lecce documentano gli aspetti di rilievo dell'attività di ricerca e didattica del Dipartimento. Nei Quaderni sono pubblicati articoli di carattere matematico che siano:

- (1) lavori di rassegna e monografie su argomenti di ricerca;
- (2) testi di seminari di interesse generale, tenuti da docenti o ricercatori del Dipartimento o esterni;
- (3) lavori di specifico interesse didattico.

La pubblicazione dei lavori è soggetta all'approvazione del Comitato di Redazione, che decide tenendo conto del parere di un *referee*, nominato di volta in volta sulla base delle competenze specifiche.

Quaderno 2/2005: ISBN 88-8305-025-8
Università di Lecce - Coordinamento di SIBA

Introduzione alla Teoria Ergodica

Carlo Sempì

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI LECCE
LECCE - ITALIA

Prefazione

Questo Quaderno riproduce con qualche ampliamento le lezioni del corso di Introduzione alla Teoria Ergodica che ho tenuto per i dottorandi del Dipartimento. Era mia intenzione non mostrare tutti i possibili aspetti della teoria, ma solo di fornire appunto un'introduzione; in questo spero di essere stato utile a chi seguí le lezioni, almeno come guida alla letteratura.

Carlo Sempì

Lecce, 6 aprile 2005

Indice

| | |
|----------------------------------------------------------|-----|
| Prefazione | 1 |
| Capitolo 1. I concetti fondamentali | 1 |
| 1. Introduzione | 1 |
| 2. Trasformazioni che conservano la misura | 3 |
| 3. Esempî | 5 |
| 4. La ricorrenza | 9 |
| 5. Ergodicit  | 12 |
| 6. Esempî di trasformazioni ergodiche | 16 |
| 7. Il teorema ergodico di Birkhoff | 18 |
| 8. Corollarî del teorema di Birkhoff | 25 |
| 9. Propriet  pi  forti dell'ergodicit  | 28 |
| 10. Espressione funzionale di ergodicit  e mescolanza | 36 |
| 11. Topologie per le trasformazioni | 42 |
| 12. Categoria | 53 |
| 13. Trasformazioni indotte | 55 |
| 14. Esempî di trasformazioni mescolanti | 61 |
| 15. Misure invarianti | 62 |
| Capitolo 2. Isomorfismo, coniugio, isomorfismo spettrale | 67 |
| 1. Introduzione | 67 |
| 2. Richiami sugli spazi di Hilbert | 72 |
| 3. Isomorfismo e coniugio | 75 |
| 4. Isomorfismi spettrali | 77 |
| Capitolo 3. L'entropia in teoria ergodica | 81 |
| 1. Partizioni e sottoalgebre | 81 |
| 2. La definizione di entropia | 84 |
| 3. Le propriet  dell'entropia di una trasformazione | 93 |
| 4. Il teorema di Kolmogorov-Sinai | 98 |
| 5. Esempî di calcolo dell'entropia | 106 |
| Nota bibliografica | 109 |
| Bibliografia | 113 |

CAPITOLO 1

I concetti fondamentali

1. Introduzione

Non è possibile introdurre la Teoria Ergodica senza fare cenno alle sue origini fisiche.

La Teoria Ergodica rappresenta uno dei tentativi di spiegare le proprietà termodinamiche dei sistemi macroscopici, vale a dire dei sistemi composti da un numero di particelle almeno pari al numero di AVOGADRO, 6×10^{23} . La disciplina che vuole di fatto dimostrare i principi della termodinamica come teoremi a partire dalle equazioni della dinamica è la *Meccanica Statistica*. In particolare, la Meccanica Statistica vuole spiegare come il comportamento irreversibile dei sistemi termodinamici, si pensi al secondo principio della termodinamica, possa derivare dalle equazioni della dinamica che sono reversibili. La letteratura scientifica dedicata alla meccanica statistica è veramente sterminata. Qui citiamo solo, per la termodinamica due classici [16, 55] e [74], per la teoria cinetica ed il cosiddetto teorema H di BOLTZMANN [12], per la meccanica statistica in generale [31, 62, 69]. Più specificamente per i fondamenti della meccanica statistica si può consultare il libro di GIBBS [22] come un modo di porre i fondamenti in maniera assai distante da quella proposta dalla teoria ergodica. Per l'approccio alla meccanica statistica attraverso la teoria ergodica si vedano [37] e [15]. Anche i fisici che preferiscono l'approccio alla meccanica statistica basato sulla teoria degli *ensembles* introdotto da GIBBS e EINSTEIN, si veda, per esempio [23], ritengono tuttavia che la teoria ergodica abbia fornito importanti contributi ai fondamenti della teoria fisica.

Ricordiamo che, nella meccanica classica, lo stato di un sistema con n gradi di libertà è descritto dalle n coordinate generalizzate $q = (q_1, \dots, q_n)$ e dai loro momenti coniugati $p = (p_1, p_2, \dots, p_n)$. Ogni stato di tale sistema è così rappresentato da un punto P di coordinate (q, p) in un sottoinsieme di \mathbf{R}^{2n} detto *spazio delle fasi*. Il moto di tale punto rappresentativo, e, quindi,

del sistema considerato, è retto dalle equazioni di HAMILTON

$$\frac{dq_j}{dt} = \frac{\partial H}{\partial p_j}, \quad \frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j} \quad (j = 1, 2, \dots, n).$$

Qui H è la (funzione) Hamiltoniana del sistema, che è indipendente dal tempo t e, di solito, si esprime come la somma dell'energia cinetica $K(p)$ e dell'energia potenziale $V(q)$. Detto $P_0 = (q_0, p_0)$ lo stato iniziale del sistema, la soluzione delle equazioni di HAMILTON determina lo stato del sistema per istante $t > 0$

$$(q_t, p_t) = T_t (q_0, p_0).$$

Si ottiene pertanto un semigruppò $\{T_t : t \geq 0\}$ di trasformazioni,

$$T_{t+t'} = T_t T_{t'},$$

sullo spazio delle fasi. Tale semigruppò di trasformazioni si dice *flusso hamiltoniano* e gode della proprietà fondamentale espressa dal seguente teorema.

TEOREMA 1.1 (di Liouville). *Il flusso hamiltoniano $\{T_t\}$ conserva la misura di LEBESGUE λ_{2n} nello spazio delle fasi: se A è un sottoinsieme misurabile dello spazio delle fasi si ha, per ogni $t \geq 0$,*

$$\lambda_{2n}(A) = \lambda_{2n}(T_t A).$$

Nel seguito ci occuperemo esclusivamente di trasformazioni che conservano la misura, adottando però uno schema nei quali il tempo è discreto.

Se $T : \Omega \rightarrow \Omega$ è una trasformazione che conserva la misura, l'orbita $\{T^n \omega : n \in \mathbf{Z}\}$ del punto $\omega \in \Omega$ rappresenta la storia completa di tale punto. Se f è una grandezza fisica, passibile di misurazione, allora

$$f(\omega), f(T\omega), \dots, f(T^n \omega), \dots$$

sono i valori di f in istanti successivi. Nella meccanica statistica di un sistema composto da piú (molte) particelle nessuna misurazione è istantanea (si pensi solo alle collisioni molecolari) sicché ogni misurazione non riguarda i valori istantanei, bensí una media temporale

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j \omega),$$

che è solitamente molto difficile, se non impossibile, da calcolare a causa della mancanza della soluzione esplicita delle equazioni di HAMILTON. L'ipotesi ergodica consite nella possibilità di sostituire la media temporale

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j \omega),$$

supposto che tale limite esista, con la media spaziale

$$\int_{\Omega} f d\mu,$$

per la quale non occorrono le soluzioni delle equazioni del moto, ma basta la conoscenza della “geometria” del sistema.

2. Trasformazioni che conservano la misura

DEFINIZIONE 2.1. Dati due spazi mensurali $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$, si dice che una trasformazione $T : \Omega_1 \rightarrow \Omega_2$

- (i) è *misurabile* se appartiene alla tribú \mathcal{F}_1 l'immagine inversa mediante T di ogni insieme di \mathcal{F}_2 , vale a dire se è $T^{-1} \mathcal{F}_2 \subset \mathcal{F}_1$, oppure, equivalentemente, se $T^{-1} A \in \mathcal{F}_1$ per ogni insieme A di \mathcal{F}_2 ;
- (ii) *conserva la misura* se T è misurabile e $\mu_1 (T^{-1} A) = \mu_2(A)$ per ogni insieme $A \in \mathcal{F}_2$.

Spesso si ha la situazione $\Omega_1 = \Omega_2 = \Omega$, si considerano cioè trasformazioni di uno spazio Ω in sé; inoltre spesso tanto μ_1 quanto μ_2 sono misure di probabilità.

Per indicare una trasformazione che conserva la misura si potrebbe ricorrere alla notazione, assai pesante, $T : (\Omega_1, \mathcal{F}_1, \mu_1) \rightarrow (\Omega_2, \mathcal{F}_2, \mu_2)$; tuttavia, ove non sia possibile confusione tra le tribú e le misure in gioco, si scriverà semplicemente $T : \Omega_1 \rightarrow \Omega_2$.

Se $T : \Omega_1 \rightarrow \Omega_2$ e $T' : \Omega_2 \rightarrow \Omega_3$ conservano la misura (tra i rispettivi spazi), altrettanto fa la trasformazione composta $S := T' \circ T : \Omega_1 \rightarrow \Omega_3$ tra gli spazi $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_3, \mathcal{F}_3, \mu_3)$; infatti, se A è un insieme di \mathcal{F}_3 si ha

$$\begin{aligned} \mu_1 (S^{-1} A) &= \mu_1 [(T' \circ T)^{-1} A] = \mu_1 [(T'^{-1} \circ T^{-1}) A] \\ &= \mu_1 [T^{-1} (T'^{-1} A)] = \mu_2 (T'^{-1} A) = \mu_3(A). \end{aligned}$$

La definizione 2.1 è in pratica di applicazione piuttosto difficile. Sono perciò utili i due teoremi che seguono e ai quali occorrerà premettere un richiamo. Ricordiamo che si chiama *semi-anello* una famiglia \mathcal{S} di sottoinsiemi di un insieme non vuoto Ω tale che

- (i) $\emptyset \in \mathcal{S}$;
- (ii) $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$ (\mathcal{S} è stabile rispetto all'intersezione);

(iii) per ogni coppia A e B di insiemi di \mathcal{S} esistono insiemi disgiunti

$$E_1, E_2, \dots, E_n$$

di \mathcal{S} tali che

$$A \setminus B = \cup_{j=1}^n E_j.$$

Ricordiamo (si veda, per esempio, [38, Theorem 1.4], che l'anello $\mathbf{R}(\mathcal{S})$ generato dal semi-anello \mathcal{S} è composto da tutti, e soli, gli insiemi E che possono essere espressi come l'unione

$$E = \cup_{j=1}^n A_j \tag{2.1}$$

di un numero finito di insiemi disgiunti di \mathcal{S} .

TEOREMA 2.1. *Siano $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ spazi mensurali, sia \mathcal{S}_2 un semi-anello contenuto in \mathcal{F}_2 e sia $T : \Omega_1 \rightarrow \Omega_2$ una trasformazione tale che, per ogni insieme appartenente al semi-anello \mathcal{S}_2 , $T^{-1} E$ appartenga alla tribú \mathcal{F}_1 e tale che*

$$\mu_1 (T^{-1} E) = \mu_2 (E). \tag{2.2}$$

Allora la (2.2) vale per tutti gli insiemi dell'anello $\mathbf{R}_2 = \mathbf{R}_2(\mathcal{S}_2)$ generato da \mathcal{S}_2 .

DIMOSTRAZIONE. In virtù della (2.1), ogni insieme A di \mathbf{R}_2 si può esprimere come unione disgiunta di un numero finito di insiemi di \mathcal{S} ,

$$A = \cup_{j=1}^n E_j.$$

Pertanto

$$\begin{aligned} \mu_1 (T^{-1} A) &= \mu_1 [T^{-1} (\cup_{j=1}^n E_j)] = \mu_1 (\cup_{j=1}^n T^{-1} E_j) = \sum_{j=1}^n \mu_1 (T^{-1} E_j) \\ &= \sum_{j=1}^n \mu_2 (E_j) = \mu_2 (\cup_{j=1}^n E_j) = \mu_2 (A), \end{aligned}$$

che stabilisce l'asserto. □

TEOREMA 2.2. *Siano $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ spazi mensurali, sia \mathcal{A}_2 un'algebra che genera la tribú \mathcal{F}_2 , $\mathcal{F}_2 = \mathcal{F}(\mathcal{A}_2)$ e sia $T : \Omega_1 \rightarrow \Omega_2$ una trasformazione tale che, per ogni insieme B di \mathcal{A}_2 ,*

- (i) $T^{-1} B \in \mathcal{F}_1$
- (ii) $\mu_1 (T^{-1} B) = \mu_2 (B)$.

Allora T conserva la misura.

DIMOSTRAZIONE. Si ponga

$$\mathcal{C} := \{A \in \mathcal{F}_2 : T^{-1} A \in \mathcal{F}_1, \mu_1(T^{-1} A) = \mu_1(A)\}$$

Da questa posizione segue che valgono le inclusioni $\mathcal{A}_2 \subset \mathcal{C} \subset \mathcal{F}_2$. Per dimostrare che $\mathcal{C} = \mathcal{F}_2$, basterà quindi dimostrare che vale l'inclusione $\mathcal{F}_2 \subset \mathcal{C}$. La dimostrazione sarà conclusa quando si sarà dimostrato che \mathcal{C} è una classe monotona; infatti, il teorema della classe monotona (si veda, per esempio, [25, Theorem 6-B]) assicura che sia $\mathcal{F}_2 \subset \mathcal{C}$. \square

3. Esempi

ESEMPIO 3.1. La trasformazione identità $I : \Omega \rightarrow \Omega$, $I\omega = \omega$, per ogni $\omega \in \Omega$, ovviamente conserva la misura sullo spazio $(\Omega, \mathcal{F}, \mu)$.

Prima di dare i prossimi esempi occorre richiamare le misure di HAAR. Per le misure di HAAR si veda [48].

Sia G un gruppo topologico compatto; esiste allora una misura finita μ definita sui boreliani $\mathcal{B}(G)$ di G tale che, per ogni $x \in G$ e per ogni $A \in \mathcal{B}(G)$, sia

$$\mu(xA) = \mu(A).$$

Se ν e μ sono due misure di HAAR definite su $\mathcal{B}(G)$, allora esiste una costante $c > 0$ tale che $\nu = c\mu$. Esiste, pertanto, un'unica misura di probabilità di HAAR definita sullo spazio misurabile $(G, \mathcal{B}(G))$.

ESEMPIO 3.2. Sia \mathcal{K} la circonferenza unitaria nel piano complesso, $\mathcal{K} := \{z \in \mathbf{C} : |z| = 1\}$, la si doti della tribù di BOREL \mathcal{B} e sia μ la misura di HAAR (normalizzata) su \mathcal{B} . Fissato un elemento a di \mathcal{K} , si definisca $T_a : \mathcal{K} \rightarrow \mathcal{K}$ mediante $T_a z := az$. poiché μ è la misura di HAAR si ha

$$\mu(T_a^{-1} A) = \mu(A)$$

per ogni boreliano A .

ESEMPIO 3.3. (del quale l'esempio precedente è un caso particolare). Siano G un gruppo topologico compatto ed a un elemento di G . Allora $T_a : G \rightarrow G$ definita da $T_a z := az$ conserva la misura di HAAR (normalizzata).

ESEMPIO 3.4. (del quale l'esempio precedente è un caso particolare). Siano G un gruppo topologico compatto e $T : G \rightarrow G$ un endomorfismo suriettivo. Allora T preserva la misura di HAAR normalizzata μ . Infatti, si definisca sui boreliani $\mathcal{B}(G)$ di G la misura di probabilità

$$\nu(A) := \mu(T^{-1} A) \quad (A \in \mathcal{B}(G)).$$

Quale che sia $x \in G$ si ha, per $E \in \mathcal{B}(G)$,

$$\nu(Tx \cdot E) = \mu [T^{-1}(Tx \cdot E)] = \mu(x \cdot T^{-1}E) = \mu(T^{-1}E) = \nu(E).$$

Ma T è suriettiva e perciò l'unicità della misura di HAAR dà l'asserto.

ESEMPIO 3.5. Sia G un gruppo topologico compatto. Una trasformazione *affine* A su G è definita da $A(x) := a \cdot T(x)$ ove a è un elemento dato di G e $T : G \rightarrow G$ è un endomorfismo suriettivo di G in sé. La trasformazione A , come composizione di due trasformazioni che conservano la misura di HAAR conserva la misura di HAAR.

ESEMPIO 3.6. (La traslazione, *shift*, bilatera). Si consideri l'insieme finito $X := \{0, 1, \dots, k-1\}$ e sia $p = (p_0, p_1, \dots, p_{k-1})$ una legge di probabilità. Naturalmente, qui e nel seguito supporremo, senza richiamarlo ogni volta, che sia $p_j > 0$ per ogni indice $j \in \{0, 1, \dots, k-1\}$. Sia

$$\Omega := \prod_{n \in \mathbf{Z}} X$$

l'insieme $X^{\mathbf{Z}}$ di tutte le successioni doppiamente infinite di elementi di X ,

$$\Omega = \{(\dots, x_{-n}, \dots, x_{-1}, x_0, x_1, \dots, x_n, \dots) : x_n \in X \ (n \in \mathbf{Z})\},$$

munito della tribú prodotto

$$\mathcal{B} := \dots \otimes \mathcal{P}(X) \otimes \mathcal{P}(X) \otimes \dots \otimes \mathcal{P}(X) \otimes \dots$$

Si consideri ora la misura prodotto μ definita sullo spazio misurabile (Ω, \mathcal{B}) . Si definisca la trasformazione $T : \Omega \rightarrow \Omega$ mediante

$$T(\{x_n\}) = \{y_n : n \in \mathbf{Z}\},$$

ove, per ogni $n \in \mathbf{Z}$, $y_n := x_{n+1}$. La trasformazione T cosí definita conserva la misura di tutti i cilindri misurabili e, quindi, di tutti le unioni finite e disgiunte di cilindri misurabili; l'insieme di tali unioni costituisce un'algebra sicché il Teorema 2.2 assicura che T conservi la misura.

ESEMPIO 3.7. Siano X e p come nell'esempio precedente e sia

$$\Omega := \prod_{n \in \mathbf{Z}_+} X$$

l'insieme $X^{\mathbf{Z}_+}$ di tutte le successioni di elementi di X

$$\Omega = \{(x_0, x_1, \dots, x_n, \dots) : x_n \in X \ (n \in \mathbf{Z}_+)\},$$

munito della tribú prodotto

$$\mathcal{B} := \mathcal{P}(X) \otimes \mathcal{P}(X) \otimes \dots \otimes \mathcal{P}(X) \otimes \dots$$

Si consideri ora la misura prodotto μ definita sullo spazio misurabile (Ω, \mathcal{B}) . Di nuovo, si definisca la trasformazione $T : \Omega \rightarrow \Omega$ mediante

$$T(\{x_n\}) = \{y_n : n \in \mathbf{Z}_+\},$$

ove, per ogni $n \in \mathbf{Z}_+$, $y_n := x_{n+1}$. Lo stesso ragionamento dell'esempio precedente mostra che T conserva la misura; si osservi però, che mentre

la traslazione bilatera è invertibile, tale non è la traslazione unilatera. Si osservi, inoltre, che se l'insieme misurabile A è invariante, $A = T^{-1}A$, rispetto alla traslazione unilatera T , allora si ha $A = TA$.

Infatti, sia $\omega \in TA$; allora, $T^{-1}\omega \in A = T^{-1}A$. Per definizione di T^{-1} si ha, quindi, $TT^{-1}\omega \in A$. Ma $TT^{-1} = I$ (l'operatore identità) sicché ω appartiene ad A . Dunque, $TA \subset A$. Ora sia $\omega \in A$, ovvero, $\omega = TT^{-1}\omega \in A$, sicché $T^{-1}\omega \in T^{-1}A = A$, cioè $T^{-1}\omega$ appartiene ad $T^{-1}A = A$, o, ancora, $\omega \in TA$ che significa $A \subset TA$. Le due inclusioni danno l'asserto.

Il viceversa non è vero, nel senso che esistono insiemi A per i quali $A = TA$ senza che essi siano invarianti. Un tale insieme è dato da

$$A := \{(x, x, \dots, x, \dots) : x \in X\}.$$

In questo caso si ha, per ogni $x \in X$, $T(x, x, \dots, x, \dots) = (x, x, \dots, x, \dots)$, vale a dire $TA = A$. D'altro canto si ha anche

$$T^{-1}(x, x, \dots, x, \dots) = \cup_{y \in X} \{(y, x, x, \dots, x, \dots)\}.$$

Perciò, $T^{-1}A \supset A$ ma $A \neq T^{-1}A$.

ESEMPIO 3.8. (La traslazione, *shift*, markoviana bilatera). Si ritorni all'esempio 3.6 e si consideri come lí l'insieme $X := \{0, 1, \dots, k-1\}$, il prodotto

$$\Omega := \prod_{n \in \mathbf{Z}} X,$$

munito della tribú prodotto $\mathcal{B} := \dots \otimes \mathcal{P}(X) \otimes \mathcal{P}(X) \otimes \dots \otimes \mathcal{P}(X) \otimes \dots$ e la traslazione bilatera

$$T\{x_n\} := \{y_n\},$$

ove, per ogni $n \in \mathbf{Z}$, $y_n := x_{n+1}$.

Si supponga ora che, per ogni $n \in \mathbf{Z}$, sia assegnata una funzione $p_n : X^{n+1} \rightarrow [0, 1]$ tale che

$$\sum_{i=0}^{k-1} p_0(i) = 1 \tag{a}$$

$$p_n(i(0), i(1), \dots, i(n)) = \sum_{i=0}^{k-1} p_{n+1}(i(0), i(1), \dots, i(n), i). \tag{b}$$

Sul semianello \mathcal{S}_e dei cilindri elementari si definisca una funzione d'insieme $\mu : \mathcal{S}_e \rightarrow \mathbf{R}_+$ mediante

$$\begin{aligned} \mu(\{\{x_n\} : x_s = i(0), x_{s+1} = i(1), \dots, x_{s+n} = i(n)\}) \\ := p_n(i(0), i(1), \dots, i(n)). \end{aligned}$$

Allora il teorema di coerenza di KOLMOGOROV assicura che μ possa essere estesa ad una misura di probabilità su (Ω, \mathcal{B}) , che indicheremo ancora con μ , e che, per i Teoremi 2.1 e 2.2, è conservata dalla trasformazione T .

Si riottiene l'Esempio 3.6 prendendo

$$p_n(a_0, a_1, \dots, a_n) = p_{a_0} p_{a_1} \cdots p_{a_n}.$$

Siano ora Π la matrice di transizione $k \times k$ di una catena di MARKOV che ha X come insieme degli stati e $p = (p_0, p_1, \dots, p_{k-1})$ un vettore (riga) di probabilità e si ponga

$$p_n(i(0), i(1), \dots, i(n)) := p_{i(0)} p_{i(0)i(1)} p_{i(1)i(2)} \cdots p_{i(n-1)i(n)}.$$

Con questa scelta della funzione p_n la trasformazione T , che, come si è visto sopra, conserva la misura, si dice *traslazione, shift, bilatera di MARKOV* (p, Π) . La traslazione bilatera dell'Esempio 3.6 è una traslazione di MARKOV se si prende $p = (p_0, p_1, \dots, p_{k-1})$ e $\Pi = (p_{ij})$ con $p_{ij} = p_j$ quale che sia $i \in \{0, 1, \dots, k-1\}$.

La trasformazione di MARKOV unilatera si costruisce come sopra effettuando gli ovvî cambiamenti.

ESEMPIO 3.9. Sia dato il quadrato unitario $[0, 1]^2$, dotato della tribú di BOREL $\mathcal{B}([0, 1]^2)$ e della restrizione λ_2 della misura di LEBESGUE a questa; allora, la proiezione canonica $\pi(x, y) := x$ conserva la misura tra gli spazi $([0, 1]^2, \mathcal{B}([0, 1]^2), \lambda_2)$ e $([0, 1], \mathcal{B}([0, 1]), \lambda)$.

OSSERVAZIONE 3.1. Siano Ω_1 un arbitrario insieme non vuoto, $(\Omega_2, \mathcal{F}_2, \mu_2)$ un qualsiasi spazio di probabilità e $T : \Omega_1 \rightarrow \Omega_2$ un'arbitraria trasformazione suriettiva; si possono, allora, scegliere una tribú \mathcal{F}_1 di sottoinsiemi di Ω_1 ed una misura di probabilità μ_1 su \mathcal{F}_1 in modo che T conservi la misura tra $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$. Basta porre

$$\mathcal{F}_1 := \{T^{-1} B : B \in \mathcal{F}_2\}$$

e definire μ_1 mediante

$$\mu_1(T^{-1} B) := \mu_2(B).$$

Viceversa, se $(\Omega_1, \mathcal{F}_1, \mu_1)$ è uno spazio di probabilità qualsiasi, se Ω_2 è un arbitrario insieme non vuoto e $T : \Omega_1 \rightarrow \Omega_2$ un'arbitraria trasformazione suriettiva, si possono scegliere una tribú \mathcal{F}_2 di sottoinsiemi di Ω_2 ed una misura di probabilità μ_2 su \mathcal{F}_2 in modo che T conservi la misura. Basta prendere

$$\mathcal{F}_2 := \{B \subset \Omega_2 : T^{-1} B \in \mathcal{F}_1\}$$

e definire μ_2 mediante

$$\mu_2(B) := \mu_1(T^{-1} B).$$

ESEMPIO 3.10. Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità e su questo si consideri il processo stocastico a tempo discreto $\{X_n : n \in \mathbf{Z}\}$. Si supponga che tale processo sia stazionario, nel senso che, per ogni scelta di un numero naturale $k \in \mathbf{N}$, di k numeri interi n_1, n_2, \dots, n_k , di k boreliani B_1, B_2, \dots, B_k e di un numero intero $n \in \mathbf{Z}$ si abbia

$$\mu \left(\bigcap_{j=1}^k X_{n_j}^{-1}(B_j) \right) = \mu \left(\bigcap_{j=1}^k X_{n_j+n}^{-1}(B_j) \right).$$

Un processo stazionario corrisponde ad una trasformazione che conserva la misura nel seguente modo. Si considerino lo spazio delle successioni reali doppiamente infinite.

$$\mathbf{R}^{\mathbf{Z}} := \{(\dots, x_{-1}, x_0, x_1, \dots, x_n, \dots) : x_n \in \mathbf{R} \ (n \in \mathbf{Z})\}$$

e la trasformazione $\varphi : \Omega \rightarrow \mathbf{R}^{\mathbf{R}}$ definita, per ogni $n \in \mathbf{Z}$, da

$$(\varphi\omega)_n := X_n(\omega).$$

Si definisca ora una misura di probabilità sui boreliani di $\mathbf{R}^{\mathbf{Z}}$ mediante

$$\nu(B) := \mu(\varphi^{-1}(B)).$$

Infine, sia $T : \mathbf{R}^{\mathbf{Z}} \rightarrow \mathbf{R}^{\mathbf{Z}}$ la trasformazione bilatera

$$(Tx)_n := x_{n+1} \quad (n \in \mathbf{Z}).$$

Come conseguenza della stazionarietà del processo $\{X_n : n \in \mathbf{Z}\}$ la probabilità ν è invariante rispetto a T e, quindi, su tutto $\mathcal{B}(\mathbf{R}^{\mathbf{Z}})$; considerata la proiezione canonica $\pi_n : \mathbf{R}^{\mathbf{Z}} \rightarrow \mathbf{R}$,

$$\pi_n\{x_k : k \in \mathbf{Z}\} := x_n \quad (n \in \mathbf{Z}).$$

Allora $\{\pi_n\}$ ha su $\mathbf{R}^{\mathbf{Z}}$ le stesse leggi congiunte di $\{X_n\}$ su Ω . Pertanto ogni processo stocastico stazionario deriva da una trasformazione che conserva la misura su $\mathbf{R}^{\mathbf{Z}}$.

4. La ricorrenza

La teoria ergodica si interessa delle proprietà asintotiche delle trasformazioni che conservano la misura, vale a dire delle proprietà della successione $\{T^n\}$. Perciò si studiano necessariamente trasformazioni di uno spazio in sé, $T : \Omega \rightarrow \Omega$.

Il primo risultato fu dato da POINCARÉ in termini di probabilità introducendo, implicitamente la nozione di additività numerabile.

DEFINIZIONE 4.1. Siano T una trasformazione su $(\Omega, \mathcal{F}, \mu)$ che conserva la misura, e A un insieme di \mathcal{F} ; un punto ω di A si dice *ricorrente*, rispetto a T e A , se $T^i \omega$ appartiene ad A per almeno un numero naturale i ; si dice, invece, che ω è *fortemente ricorrente* se $T^n \omega$ appartiene ad A per infiniti naturali, cioè $\omega \in \limsup_{n \rightarrow +\infty} T^{-n} A$.

TEOREMA 4.1 (di ricorrenza di POINCARÉ). *Siano $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità, $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura e A un insieme di \mathcal{F} . Allora, quasi tutti i punti di A sono ricorrenti.*

DIMOSTRAZIONE. Se $\mu(A) = 0$ non vi è nulla da dimostrare. Si supponga, allora, che sia $\mu(A) > 0$. Si indichi con B il sottoinsieme di A composto dai punti di A che non tornano mai in A . Si vuole dimostrare che $\mu(B) = 0$. Ora

$B = A \setminus \bigcup_{n \in \mathbf{N}} \{\omega \in A : T^n \omega \in A\} = A \setminus \bigcup_{n \in \mathbf{N}} T^{-n} A = A \cap \left(\bigcap_{n \in \mathbf{N}} T^{-n} A^c \right)$,
ciò che prova, intanto che B è misurabile, $B \in \mathcal{F}$. Se $0 \leq i < j$ si ha

$$T^{-j} B \cap T^{-i} B \subset T^{-j} A \setminus \bigcup_{n \in \mathbf{N}} T^{-i-n} A = \emptyset.$$

Ne segue che la successione $\{T^{-n} B : n \in \mathbf{Z}_+\}$ è composta da insieme misurabili a due a due disgiunti. Ma $\mu(T^{-n} B) = \mu(B)$, per ogni $n \in \mathbf{Z}_+$, perché T conserva la misura. Da

$$\Omega \supset \bigcup_{n \in \mathbf{Z}_+} T^{-n} B$$

scende che

$$1 = \mu(\Omega) \geq \mu\left(\bigcup_{n \in \mathbf{Z}_+} T^{-n} B\right) = \sum_{n \in \mathbf{Z}_+} \mu(T^{-n} B) = \sum_{n \in \mathbf{Z}_+} \mu(B);$$

ma ciò è possibile solo se $\mu(B) = 0$. \square

TEOREMA 4.2 (di ricorrenza forte). *Nelle stesse ipotesi del teorema precedente si supponga che A sia un insieme misurabile con $\mu(A) > 0$. Allora quasi tutti i punti di A sono fortemente ricorrenti.*

DIMOSTRAZIONE. L'insieme dei punti che entrano infinite volte in A è dato da

$$B := \limsup_{n \rightarrow +\infty} T^{-n} A,$$

mentre l'insieme dei punti di A che sono fortemente ricorrenti è $A \cap B$. Posto, per $n \in \mathbf{Z}_+$,

$$C_n := \bigcup_{k \geq n} T^{-k} A,$$

si ha $T^{-1} C_n = C_{n+1}$ e

$$C_0 \supset C_1 \supset \cdots \supset C_n \supset C_{n+1} \supset \cdots$$

con $C_0 \supset A$. Poiché T conserva la misura si ha, per ogni $n \in \mathbf{N}$,

$$\mu(C_{n+1}) = \mu(T^{-1} C_n) = \mu(C_n) = \cdots = \mu(C_0).$$

Ora

$$\mu(B) = \mu\left(\bigcap_{n \in \mathbf{Z}_+} C_n\right) = \lim_{n \rightarrow +\infty} \mu(C_n) = \mu(C_0). \quad (4.1)$$

D'altro canto

$$A = A \cap C_0 = (A \cap B) \cup [A \cap (C_0 \setminus B)],$$

che è un'unione disgiunta, sicché

$$\mu(A \cap B) = \mu(A \cap C_0) - \mu[A \cap (C_0 \setminus B)].$$

Poiché B è incluso in C_0 , $B \subset C_0$, in virtù della (4.1) si ha

$$\mu[A \cap (C_0 \setminus B)] \leq \mu(C_0 \setminus B) = \mu(C_0) - \mu(B) = 0;$$

pertanto

$$\mu(A \cap B) = \mu(A \cap C_0) = \mu(A),$$

ciò che conclude la dimostrazione. \square

L'ipotesi che la trasformazione T conservi la misura è stata usata solo per stabilire che la successione $\{T^{-n}B : n \in \mathbf{N}\}$ è costituita da insiemi a due a due disgiunti e per escludere la possibilità che B abbia misura diversa da zero. Pertanto tutti i teoremi dati possono essere rafforzati eliminando le ipotesi che non sono essenziali.

DEFINIZIONE 4.2. Data una trasformazione $T : \Omega \rightarrow \Omega$, si dice che un insieme $A \subset \Omega$ è *vagante* (in inglese *wandering*) se la successione $\{T^{-n}A : n \in \mathbf{Z}_+\}$ è disgiunta.

DEFINIZIONE 4.3. Una trasformazione misurabile T su $(\Omega, \mathcal{F}, \mu)$ si dice *dissipativa* se esiste un insieme misurabile vagante di misura strettamente positiva; altrimenti si dice *conservativa*.

La dimostrazione del teorema di ricorrenza mostra che se T conserva la misura, essa è anche conservativa.

DEFINIZIONE 4.4. Una trasformazione misurabile T su $(\Omega, \mathcal{F}, \mu)$ si dice *comprimibile*, se esiste un insieme $A \in \mathcal{F}$ tale che

$$A \subset T^{-1}A \quad \text{e} \quad \mu(T^{-1}A \setminus A) > 0;$$

se un tale insieme non esiste T si dice *incomprimibile*.

TEOREMA 4.3. Per una trasformazione misurabile T su $(\Omega, \mathcal{F}, \mu)$ sono equivalenti le affermazioni

- (a) T è comprimibile;
- (b) T è dissipativa.

DIMOSTRAZIONE. (a) \implies (b) Si supponga che T sia comprimibile, sia A un insieme come nella definizione 4.4; posto $B := T^{-1}A \setminus A$, risulta $\mu(B) > 0$.

Inoltre, per ogni $n \in \mathbf{N}$, si ha

$$\begin{aligned} B \cap T^{-n} B &= (T^{-1} A \setminus A) \cap (T^{-n-1} A \setminus T^{-n} A) \\ &= T^{-1} A \cap A^c \cap T^{-n-1} A \cap T^{-n} A^c \\ &= T^{-1} (A \cap T^{-n} A) \cap (A^c \cap T^{-n} A^c) = T^{-1} A \cap T^{-n} A^c \\ &= T^{-1} (A \cap T^{-n+1} A^c) = T^{-1} \emptyset = \emptyset, \end{aligned}$$

perché da $A \subset T^{-1} A$ scendono le inclusioni $T^{-1} A^c \subset A^c$, $T^{-n} A \supset A$ e $T^{-n} A^c \subset A^c$. Ma allora la successione $\{T^{-n} B : n \in \mathbf{Z}_+\}$ è disgiunta e quindi T è dissipativa.

(b) \implies (a) Sia T dissipativa; esiste pertanto un insieme $B \in \mathcal{F}$ tale che sia $\mu(B) > 0$ e $B \cap T^{-n} B = \emptyset$ per ogni $n \in \mathbf{N}$. Posto

$$A := \left(\bigcup_{n \in \mathbf{Z}_+} T^{-n} B \right)^c = \bigcap_{n \in \mathbf{Z}_+} T^{-n} B^c,$$

si ha

$$T^{-1} A = \bigcap_{n \in \mathbf{N}} T^{-n} B^c \supset \bigcap_{n \in \mathbf{Z}_+} T^{-n} B^c = A.$$

Inoltre

$$\begin{aligned} T^{-1} A \setminus A &= \bigcap_{n \in \mathbf{N}} T^{-n} B^c \setminus \bigcap_{n \in \mathbf{Z}_+} T^{-n} B^c \\ &= \left(\bigcup_{n \in \mathbf{N}} T^{-n} B \right)^c \setminus \left(\bigcup_{n \in \mathbf{Z}_+} T^{-n} B \right)^c \\ &= \left(\bigcup_{n \in \mathbf{N}} T^{-n} B \right)^c \cap \left(\bigcup_{n \in \mathbf{Z}_+} T^{-n} B \right) = B, \end{aligned}$$

cosicché $\mu(T^{-1} A \setminus A) = \mu(B) > 0$ e T risulta essere comprimibile. \square

5. Ergodicità

Diamo qui la definizione di trasformazione ergodica.

DEFINIZIONE 5.1. Dato uno spazio misurale $(\Omega, \mathcal{F}, \mu)$, una trasformazione T di Ω in sé che conservi la misura si dice *ergodica* se ogni insieme misurabile che sia invariante rispetto a T , cioè tale che $A \in \mathcal{F}$ e $A = T^{-1} A$, ha misura nulla oppure il suo complementare ha misura nulla, $\mu(A) = 0$ oppure $\mu(A^c) = 0$.

OSSERVAZIONE 5.1. Se μ è una misura di probabilità, T è ergodica se si ha $\mu(A) = 0$ oppure $\mu(A) = 1$ per ogni insieme misurabile invariante A .

Il seguente teorema dà formulazioni equivalenti dell'ergodicità.

TEOREMA 5.1. Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità e $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura. Le seguenti affermazioni sono equivalenti:

- (a) T è ergodica;
 (b) se B è un insieme misurabile tale che $\mu(T^{-1}B \Delta B) = 0$, allora si ha $\mu(B) = 0$ oppure $\mu(B) = 1$;
 (c) se gli insiemi misurabili A e B hanno entrambi probabilità strettamente positiva, $\mu(A) > 0$ e $\mu(B) > 0$, allora esiste un numero naturale n tale che $\mu(T^{-n}A \cap B) > 0$;
 (d) se $A \in \mathcal{F}$ è tale che $\mu(A) > 0$ e se $\tilde{A} := \cup_{n \in \mathbf{N}} T^{-n}A$, allora $\mu(\tilde{A}) = 1$.

DIMOSTRAZIONE. (a) \implies (b) Sia $B \in \mathcal{F}$ un insieme tale che $\mu(T^{-1}B \Delta B) = 0$ e si ponga

$$C := \limsup_{n \rightarrow +\infty} T^{-n}B = \bigcap_{n \in \mathbf{Z}_+} \bigcup_{k \geq n} T^{-k}B.$$

Ovviamente l'insieme C così definito è misurabile, $C \in \mathcal{F}$; inoltre esso è invariante, $T^{-1}C = C$, come subito si vede. Ma allora la condizione (a) implica che sia $\mu(C) = 0$ oppure $\mu(C) = 1$. Basta ora mostrare che è $\mu(C) = \mu(B)$.

Posto $B_n := \bigcup_{k \geq n} T^{-k}B$, si ha $C = \lim_{n \rightarrow +\infty} B_n$ e, per ogni $n \in \mathbf{Z}_+$,

$$T^{-1}B_n = B_{n+1}, \quad B_n \subset B_{n-1}, \quad \mu(B_n) = \mu(B_0).$$

Pertanto

$$\bigcap_{k=0}^n B_k = B_n \quad \text{e} \quad \mu\left(\bigcap_{k=0}^n B_k\right) = \mu(B_n) = \mu(B_0),$$

onde anche $\mu(C) = \mu(B_0)$. D'altro canto, scrivendo B_0 come un'unione disgiunta si ottiene

$$B_0 = \bigcup_{n \in \mathbf{Z}_+} T^{-n}B = B \cup (T^{-1}B \setminus B) \cup [T^{-2}B \setminus T^{-1}B \setminus B] \cup \dots$$

e di qui, ricorrendo alla condizione (b),

$$\begin{aligned} \mu(B_0) &= \mu(B) + \sum_{n \in \mathbf{N}} \mu\left(T^{-n}B \setminus T^{-(n-1)}B \setminus \dots \setminus B\right) \\ &\leq \mu(B) + \sum_{n \in \mathbf{N}} \mu\left(T^{-n}B \setminus T^{-(n-1)}B\right) \\ &= \mu(B) + \sum_{n \in \mathbf{N}} \mu\left[T^{-(n-1)}(T^{-1}B \setminus B)\right] \\ &= \mu(B) + \sum_{n \in \mathbf{N}} \mu(T^{-1}B \setminus B) = \mu(B). \end{aligned}$$

quindi è anche $\mu(C) = \mu(B)$.

(b) \implies (c) Si supponga che sia $\mu(A)\mu(B) > 0$ e, al tempo stesso, che la (c) sia falsa, vale a dire che, per ogni $n \in \mathbf{N}$, sia

$$\mu(T^{-n}A \cap B) = 0.$$

Ne segue che $\mu[(\cup_{n \in \mathbf{N}} T^{-n} A) \cap B] = 0$, perché

$$\begin{aligned} 0 \leq \mu[(\cup_{n \in \mathbf{N}} T^{-n} A) \cap B] &= \mu[\cup_{n \in \mathbf{N}} (T^{-n} A \cap B)] \\ &\leq \sum_{n \in \mathbf{N}} \mu(T^{-n} A \cap B) = 0. \end{aligned}$$

Ponendo

$$A_1 := \cup_{n \in \mathbf{N}} T^{-n} A,$$

si ha

$$T^{-1} A_1 \subset A_1 \quad \text{e} \quad \mu(T^{-1} A_1) = \mu(A_1),$$

di modo che

$$\mu(T^{-1} A_1 \Delta A_1) = \mu(A_1 \setminus T^{-1} A_1) = 0.$$

Quindi la condizione (b) implica $\mu(A_1) = 0$ oppure $\mu(A_1) = 1$. Però $T^{-1} A \subset A_1$, onde

$$0 < \mu(A) = \mu(T^{-1} A) \leq \mu(A_1),$$

che dà $\mu(A_1) = 1$. Ora le tre relazioni

$$\mu(B) > 0, \quad \mu(A_1) = 1, \quad \text{e} \quad \mu(A_1 \cap B) = \mu[(\cup_{n \in \mathbf{N}} T^{-n} A) \cap B] = 0,$$

delle quali la prima vale per ipotesi, mentre le rimanenti due sono appena state dimostrate, non sono compatibili. Infatti, dall'ultima scende

$$1 = \mu[(A_1 \cap B)^c] = \mu(A_1^c \cup B^c) \leq \mu(A_1^c) + \mu(B^c) = \mu(B^c) = 1 - \mu(B),$$

cioè $\mu(B) = 0$, una contraddizione.

(c) \implies (a) Si supponga, se possibile, che la (a) sia falsa; esiste, pertanto, un insieme misurabile e invariante A , $A \in \mathcal{F}$ e $T^{-1} A = A$ con $\mu(A) \in]0, 1[$. Si scelga $B = A^c$, di modo che $\mu(B) = 1 - \mu(A) > 0$. Allora per ogni $n \in \mathbf{N}$ si avrebbe

$$\mu(T^{-n} A \cap A^c) = \mu(A \cap A^c) = \mu(\emptyset) = 0,$$

che contraddice alla (c).

(b) \implies (d) L'insieme misurabile \tilde{A} è, ovviamente, misurabile; inoltre si ha

$$T^{-1} \tilde{A} = \cup_{n=2}^{\infty} T^{-n} A \subset \cup_{n \in \mathbf{N}} T^{-n} A = \tilde{A}.$$

D'altra parte, poiché T conserva la misura, si ha

$$\mu(\tilde{A} \Delta T^{-1} \tilde{A}) = \mu(\tilde{A}) - \mu(T^{-1} \tilde{A}) = 0,$$

sicché $\mu(\tilde{A}) = 1$.

(d) \implies (c) Siano A e B insiemi misurabili con $\mu(A) > 0$ e $\mu(B) > 0$; si supponga, se possibile, che sia $\mu(T^{-n} A \cap B) = 0$ per ogni $n \in \mathbf{N}$. Allora

$$\mu(\tilde{A} \cap B) = \sum_{n \in \mathbf{N}} \mu(T^{-n} A \cap B) = 0.$$

Ma $\tilde{A} = (\tilde{A} \cap B) \cup (\tilde{A} \cap B^c)$; per quanto appena provato si ha

$$1 = \mu(\tilde{A}) = \mu(\tilde{A} \cap B^c) \leq \mu(B^c),$$

onde $\mu(B^c) = 1$ e $\mu(B) = 0$, una contraddizione. \square

Se la trasformazione T è anche invertibile il teorema precedente può essere modificato come segue. La dimostrazione è una lieve variazione di quella appena data.

TEOREMA 5.2. *Se T è anche invertibile le seguenti affermazioni sono equivalenti tra loro e sono equivalenti alle condizioni (a) e (b) del teorema precedente:*

- (c') *se gli insiemi misurabili A e B hanno entrambi probabilità strettamente positiva, $\mu(A) > 0$ e $\mu(B) > 0$, allora esiste un numero intero $n \in \mathbf{Z}$ tale che $\mu(T^{-n} A \cap B) > 0$;*
- (d) *se $A \in \mathcal{F}$ è tale che $\mu(A) > 0$ e se $\tilde{A} := \cup_{n \in \mathbf{Z}} T^{-n} A$, allora $\mu(\tilde{A}) = 1$.*

L'ergodicità può essere caratterizzata usando metodi funzionali che sono spesso utili.

TEOREMA 5.3. *Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità e $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura. Le seguenti affermazioni sono equivalenti:*

- (a) *T è ergodica;*
- (b) *se $f : \Omega \rightarrow \mathbf{C}$ è una funzione misurabile e invariante rispetto a T , vale a dire tale che $(f \circ T)(\omega) = f(\omega)$ per ogni $\omega \in \Omega$, allora f è costante q.c.;*
- (c) *se $f : \Omega \rightarrow \mathbf{C}$ è una funzione misurabile e tale che $(f \circ T)(\omega) = f(\omega)$ per quasi ogni $\omega \in \Omega$, allora f è costante q.c.;*
- (d) *se $f \in L^2$ è invariante, $(f \circ T)(\omega) = f(\omega)$ per ogni $\omega \in \Omega$, allora f è costante q.c.;*
- (e) *se $f \in L^2$ è tale che $(f \circ T)(\omega) = f(\omega)$ per quasi ogni $\omega \in \Omega$, allora f è costante q.c..*

DIMOSTRAZIONE. Sono ovviamente vere le implicazioni (c) \implies (b) \implies (d) e (c) \implies (e) \implies (d).

Per completare la dimostrazione del teorema basterà, quindi, stabilire le implicazioni (a) \implies (c) e (d) \implies (a).

(a) \implies (c) Non è restrittivo supporre che la funzione f sia a valori reali, perché, altrimenti, si possono considerare separatamente le parti reali e immaginaria di f . Per ogni $n \in \mathbf{N}$ e per ogni $k \in \mathbf{Z}$, si definiscano gli insiemi

$$A(k, n) := \left\{ \omega \in \Omega : \frac{k}{2^n} \leq f(\omega) < \frac{k+1}{2^n} \right\}.$$

Vale l'inclusione

$$T^{-1} A(k, n) \Delta A(k, n) \subset \{f \circ T \neq f\}$$

onde $\mu [T^{-1} A(k, n) \Delta A(k, n)] = 0$. Poiché T è ergodica, il Teorema 5.1 assicura che, per ogni $n \in \mathbf{N}$ e per ogni $k \in \mathbf{Z}$, sia $\mu(A(k, n)) = 0$ oppure $\mu(A(k, n)) = 1$. Per ogni $n \in \mathbf{N}$ si ha

$$\cup_{k \in \mathbf{Z}} A(k, n) = \Omega,$$

che è un'unione disgiunta. Perciò, per ogni $n \in \mathbf{N}$, esiste un unico indice $k_n \in \mathbf{Z}$ tale che $\mu(A(k_n, n)) = 1$, mentre $\mu(A(k, n)) = 0$, se $k \neq k_n$. Si ponga

$$B := \cap_{n \in \mathbf{N}} A(k_n, n).$$

Evidentemente è $\mu(B) = 1$; inoltre f è costante in B . Infatti, siano ω e ω' due punti di B ; allora, per ogni $n \in \mathbf{N}$, è

$$0 \leq f(\omega) - f(\omega') \leq \frac{1}{2^n};$$

l'arbitrarietà di n dà ora $f(\omega) = f(\omega')$.

(d) \implies (a) Si considerino l'insieme invariante A , $T^{-1} A = A$ e la sua funzione indicatrice 1_A che, ovviamente, appartiene a L^2 . Si ha, per ogni $\omega \in \Omega$,

$$1_A(\omega) = 1_{T^{-1} A}(\omega) = 1_A(T\omega) = (1_A \circ T)(\omega),$$

cosicché 1_A è q.c. eguale a 0 oppure a 1; integrando, si ha così $\mu(A) = 0$ oppure $\mu(A) = 1$. \square

6. Esempi di trasformazioni ergodiche

Ritorniamo in questa sezione agli esempi della Sezione 2 per esaminare quali delle trasformazioni che conservano la misura lí presentate siano ergodiche.

ESEMPIO 6.1. L'identità I sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ è ergodica se, e solo se, ogni insieme di \mathcal{F} ha probabilità eguale a 0 o a 1.

Per quanto riguarda le rotazioni della circonferenza unitaria in \mathbf{C} , Esempio 3.2, vale il seguente

TEOREMA 6.1. Per la rotazione $T_a z = az$ ($a \in \mathcal{K}$) sono equivalenti le affermazioni:

- (a) T_a è ergodica;
- (b) a non è una radice dell'unità.

DIMOSTRAZIONE. (a) \implies (b) Si supponga che a sia una radice dell'unità; esiste, perciò $p \in \mathbf{N}$ tale che $a^p = 1$; si consideri ora la funzione $f : \mathcal{K} \longrightarrow \mathcal{K}$ definita da $f(z) := z^p$; questa è ovviamente invariante rispetto a T_a , $(f \circ T_a)(z) = f(z)$, ma non è costante sicché T_a non può essere ergodica.

(b) \implies (a) Si supponga che a non sia una radice dell'unità e che f sia invariante rispetto a T_a , $f \circ T_a = f$. Si scriva la serie di FOURIER di f

$$f(z) = \sum_{n \in \mathbf{Z}} c_n z^n.$$

Allora

$$(f \circ T_a)(z) = f(az) = \sum_{n \in \mathbf{Z}} c_n a^n z^n$$

sicché l'invarianza di f dà la relazione, valida per ogni $n \in \mathbf{Z}$,

$$c_n (a^n - 1) = 0,$$

che, per $n \neq 0$, dà, a sua volta $c_n = 0$. Dunque, $f(z) = c_0$, vale a dire che f è costante e, quindi, T_a è ergodica. \square

ESEMPIO 6.2. La traslazione bilatera (Esempio 3.6) è ergodica. Sia \mathcal{A} l'algebra di tutte le unioni finite di rettangoli misurabili e sia B un insieme misurabile invariante, $T^{-1}B = B$, $B \in \mathcal{F}$. Sia dato $\varepsilon > 0$ e si scelga $A \in \mathcal{A}$ in modo che sia $\mu(B \Delta A) < \varepsilon$. Si tenga presente che quali che siano gli insiemi misurabili E e F vale la relazione

$$\begin{aligned} |\mu(E) - \mu(F)| &= |\mu(E \cap F) + \mu(E \setminus F) - \mu(F \cap E) - \mu(F \setminus E)| \\ &\leq \mu(E \setminus F) + \mu(F \setminus E) = \mu(E \Delta F). \end{aligned}$$

Ponendo $E = A$ e $F = B$ si ottiene

$$|\mu(B) - \mu(A)| = \mu(B \Delta A) < \varepsilon.$$

Si può scegliere un numero naturale n_0 tale che l'insieme $C = T^{-n_0}A$ dipenda da coordinate distinte da quelle di A . Pertanto

$$\mu(C \cap A) = \mu(C) \mu(A) = \mu^2(A).$$

Inoltre

$$\mu(B \Delta C) = \mu(T^{-n_0} B \Delta T^{-n_0} A) = \mu(B \Delta A) < \varepsilon$$

e, poiché $B \Delta (A \cap C) \subset (B \Delta A) \cup (B \Delta C)$, si ha $\mu(B \Delta (A \cap C)) < 2\varepsilon$. Di qui

$$|\mu(B) - \mu(A \cap C)| < 2\varepsilon$$

e

$$\begin{aligned} |\mu(B) - \mu^2(B)| &\leq |\mu(B) - \mu(A \cap C)| + |\mu(A \cap C) - \mu^2(B)| \\ &< 2\varepsilon + |\mu^2(A) - \mu^2(B)| \\ &\leq 2\varepsilon + (\mu(A) + \mu(B)) |\mu(A) - \mu(B)| < 4\varepsilon. \end{aligned}$$

Siccome $\varepsilon > 0$ è arbitrario, si ha $\mu(B) = \mu^2(B)$, vale a dire $\mu(B) = 0$ oppure $\mu(B) = 1$.

Un ragionamento simile mostra che è ergodica anche la traslazione unilatera (Esempio 3.7)

Nella sezione 9 studieremo quando le traslazioni di MARKOV bilatere (Esempio 3.8) siano ergodiche.

7. Il teorema ergodico di Birkhoff

Si è visto studiando la ricorrenza che quasi ogni punto di un insieme misurabile A torna infinite volte in A stesso. Ha grande significato nella Meccanica Statistica porsi il problema di stabilire quanto “tempo” un punto di A passi nello stesso insieme. Se 1_A è la funzione indicatrice di A , la risposta è data dal limite

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} 1_A(T^j \omega),$$

se tale limite esiste. A questo problema rispose BIRKHOFF nel 1931 con il seguente teorema del quale riporto la dimostrazione di RIESZ.

TEOREMA 7.1 (Teorema ergodico di BIRKHOFF). *Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurabile σ -finito, sia $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura e sia f una funzione di L^1 . Allora*

- (a) $\frac{1}{n} \sum_{j=0}^{n-1} f(T^j \omega)$ tende q.o. ad una funzione f^* di L^1 ;
- (b) la funzione limite f^* è invariante rispetto a T , vale a dire, $f^* \circ T = f^*$.

Inoltre se lo spazio $(\Omega, \mathcal{F}, \mu)$ è finito, $\mu(\Omega) < +\infty$, si ha

$$\int_{\Omega} f^* d\mu = \int_{\Omega} f d\mu. \quad (7.1)$$

La dimostrazione richiede qualche concetto che sarà utile anche per il seguito.

DEFINIZIONE 7.1. Sia $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura; sullo spazio delle funzioni f definite in Ω e a valori complessi è definito un operatore U_T mediante

$$\forall \omega \in \Omega \quad (U_T f)(\omega) := (f \circ T)(\omega) = f(T\omega).$$

L'operatore U_T si dice *indotto* dalla trasformazione T .

OSSERVAZIONE 7.1. Se f appartiene a L^p , con $p \geq 1$, allora vi appartiene anche $U_T f$, cioè $U_T L^p \subset L^p$. Infatti

$$\|U_T f\|_p^p = \int_{\Omega} |f \circ T|^p d\mu = \int_{\Omega} |f|^p d(\mu T^{-1}) = \int_{\Omega} |f|^p d\mu = \|f\|_p^p.$$

Da quest'ultima relazione scende che U_T è un'isometria su L^p . In particolare U_T è un'isometria sullo spazio di HILBERT L^2 . Non è inutile ricordare che questo implica inoltre che sia

$$\langle U_T f, U_T g \rangle = \langle f, g \rangle$$

per ogni coppia di elementi f e g di L^2 .

TEOREMA 7.2 (Teorema ergodico massimale). Sia $U : L^1 \rightarrow L^1$ un operatore lineare positivo, $U f \geq 0$ se $f \geq 0$, e contrattivo, $\|U\| \leq 1$ e sia N un numero naturale. Posto

$$f_0 := 0, \quad f_n := \sum_{j=0}^{n-1} U^j f, \quad e \quad F_N := \max\{f_n : 0 \leq n \leq N\}.$$

Allora

$$\int_{\{F_N > 0\}} f d\mu \geq 0. \tag{7.2}$$

e

$$\int_{\cup_{N \in \mathbf{N}} \{F_N > 0\}} f d\mu \geq 0. \tag{7.3}$$

DIMOSTRAZIONE. Poiché, per ogni n con $0 \leq n \leq N$, si ha $F_N \geq f_n$ anche $U F_N \geq U f_n$. Ora

$$U f_n = U \left(\sum_{j=0}^{n-1} U^j f \right) = \sum_{j=0}^{n-1} U^{j+1} f = f_{n+1} - f \quad (0 \leq n \leq N),$$

cosicché

$$U F_N + f \geq U f_n + f = f_{n+1}.$$

Perciò, se $F_N(\omega) > 0$, è

$$\begin{aligned} U F_N(\omega) + f(\omega) &\geq \max\{f_{n+1}(\omega) : 0 \leq n \leq N-1\} \\ &= \max\{f_n(\omega) : 1 \leq n \leq N\} = F_N(\omega). \end{aligned}$$

Pertanto nell'insieme $A_N := \{F_N > 0\}$ si ha $f \geq F_N - U F_N$ e, quindi, dato che $F_N = 0$ in A_N^c ,

$$\begin{aligned} \int_{A_N} f d\mu &\geq \int_{A_N} F_N d\mu - \int_{A_N} U F_N d\mu = \int_{\Omega} F_N d\mu - \int_{A_N} U F_N d\mu \\ &\geq \int_{\Omega} F_N d\mu - \int_{\Omega} U F_N d\mu \\ &= \|F_N\|_1 - \|U F_N\|_1 \geq \|F_N\|_1 - \|F_N\|_1 = 0. \end{aligned}$$

Per dimostrare la (7.3), si scriva

$$\int_{\cup_{n \in \mathbf{N}} \{F_N > 0\}} f d\mu = \int_{\cup_{n \in \mathbf{N}} \{F_N > 0\}} f^+ d\mu - \int_{\cup_{n \in \mathbf{N}} \{F_N > 0\}} f^- d\mu$$

e si usi il teorema di convergenza monotona a ciascuno degli ultimi due integrali. \square

Si noti che il teorema ergodico massimale può essere applicato, in particolare, all'operatore U_T indotto da una trasformazione T che conserva la misura. Per un tale operatore si ha, infatti, $\|U_T\| \leq 1$.

COROLLARIO 7.1. *Se $T : \Omega \rightarrow \Omega$ conserva la misura, se φ è integrabile, $\varphi \in L^1$, se si pone*

$$B_a := \left\{ \sup_{n \in \mathbf{N}} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(T^j \omega) > a \right\},$$

e se A è un insieme misurabile invariante e di misura strettamente positiva e finita, $A \in \mathcal{F}$, $T^{-1}A = A$, $0 < \mu(A) < +\infty$, allora

$$\int_{B_a \cap A} \varphi d\mu \geq a \mu(B_a \cap A). \quad (7.4)$$

DIMOSTRAZIONE. Si supponga dapprima che sia $\mu(\Omega) < +\infty$ e che A coincida con tutto lo spazio Ω . Si ponga $f := \varphi - \alpha$, sicché

$$B_a = \cup_{N \in \mathbf{N}} \{F_N > 0\} = \cup_{N \in \mathbf{N}} A_N.$$

La (7.3) dà, per ogni $N \in \mathbf{N}$,

$$\int_{B_a} f d\mu \geq 0,$$

onde

$$\int_{B_a} \varphi d\mu \geq a \mu(B_a).$$

Il caso generale segue considerando la restrizione di T ad A . \square

DIMOSTRAZIONE DEL TEOREMA DI BIRKHOFF. Si definiscano le funzioni

$$f^* := \limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j,$$

$$f_* := \liminf_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j.$$

Mostriamo che f^* e f_* sono invarianti rispetto a T , $f^* \circ T = f^*$, e $f_* \circ T = f_*$. Si ponga

$$\varphi_n := \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j$$

e si noti che

$$\frac{n+1}{n} \varphi_{n+1} - \varphi_n \circ T = \frac{f}{n}. \quad (7.5)$$

Da quest'ultima relazione segue che

$$\begin{aligned} f^* &= \limsup_{n \rightarrow +\infty} \frac{n+1}{n} \varphi_{n+1} = \limsup_{n \rightarrow +\infty} \left(\varphi_n \circ T + \frac{f}{n} \right) \\ &\leq \limsup_{n \rightarrow +\infty} \varphi_n \circ T + \limsup_{n \rightarrow +\infty} \frac{f}{n} = f^* \circ T. \end{aligned}$$

Similmente, sempre utilizzando la (7.5), si ha

$$\begin{aligned} f^* \circ T &= \limsup_{n \rightarrow +\infty} \varphi_n \circ T = \limsup_{n \rightarrow +\infty} \left(\frac{n+1}{n} \varphi_{n+1} - \frac{f}{n} \right) \\ &\leq \limsup_{n \rightarrow +\infty} \frac{n+1}{n} \varphi_{n+1} - \liminf_{n \rightarrow +\infty} \frac{f}{n} = \limsup_{n \rightarrow +\infty} \frac{n+1}{n} \varphi_{n+1} = f^*. \end{aligned}$$

Le ultime due disequaglianze, insieme, danno l'invarianza di f^* rispetto a T . In maniera analoga si mostra che è invariante rispetto a T la funzione f_* .

Siano ora α e β due numeri reali con $\alpha < \beta$ e si ponga

$$A_{\alpha,\beta} := \{f_* < \alpha < \beta < f^*\}.$$

L'insieme $A_{\alpha,\beta}$ è misurabile, $A_{\alpha,\beta} \in \mathcal{F}$, e invariante rispetto a T , $T^{-1} A_{\alpha,\beta} = A_{\alpha,\beta}$. Se si dimostrasse che $A_{a,b}$ ha misura finita, $\mu(A_{a,b}) < +\infty$, allora si potrebbe usare il Corollario 7.1.

Se $\beta > 0$, in virtù dell'ipotesi che la misura μ sia σ -finita, si può scegliere un insieme misurabile B contenuto in $A_{\alpha,\beta}$ e di misura finita, $B \subset A_{\alpha,\beta}$ e $\mu(B) < +\infty$. Così la funzione $h := f - \beta 1_B$ è integrabile e dal Teorema ergodico massimale 7.2, con H_N definita in maniera analoga a F_N , segue che, per ogni $N \in \mathbf{N}$, si ha

$$\int_{\cup_{N \in \mathbf{N}} \{H_N > 0\}} (f - \beta 1_B) d\mu \geq 0.$$

Si osservi che se il punto ω appartiene a $A_{\alpha,\beta}$ si ha

$$\beta < f^*(\omega) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j \omega),$$

cosicché almeno una delle somme

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j \omega)$$

deve essere maggiore di β . Ne segue che almeno una delle somme

$$\frac{1}{n} \sum_{j=0}^{n-1} \{f(T^j \omega) - \beta 1_B(T^j \omega)\}$$

è positiva. Pertanto

$$B \subset A_{\alpha,\beta} \subset \cup_{N \in \mathbf{N}} \{H_N > 0\}$$

e

$$\begin{aligned} \int_{\Omega} |f| d\mu &\geq \int_B |f| d\mu \geq \int_B f d\mu = \int_{B \cap \cup_{N \in \mathbf{N}} \{H_N > 0\}} f d\mu \\ &\geq \beta \mu(B \cap \cup_{N \in \mathbf{N}} \{H_N > 0\}) = \beta \mu(B), \end{aligned}$$

vale a dire $\mu(B) \leq \|f\|_1 / \beta$. Ogni sottoinsieme B di $A_{\alpha,\beta}$ che sia misurabile e che abbia misura finita ha la misura maggiorata da $\|f\|_1 / \beta$. Ma allora la σ -finitezza dello spazio misurale $(\Omega, \mathcal{F}, \mu)$ implica che anche $A_{\alpha,\beta}$ abbia misura finita, $\mu(A_{\alpha,\beta}) < +\infty$. Infatti, si può scrivere

$$A_{\alpha,\beta} = \cup_{n \in \mathbf{N}} E_n,$$

ove gli insiemi E_n ($n \in \mathbf{N}$) sono misurabili, e si può supporre, senza perdita di generalità, che essi siano a due a due disgiunti, $E_j \cap E_k$ ($j \neq k$); inoltre, essi hanno tutti misura finita, $\mu(E_n) < +\infty$, per ogni $n \in \mathbf{N}$. Ora

$$\mu(A_{\alpha,\beta}) = \mu(\cup_{n \in \mathbf{N}} E_n) = \sum_{n \in \mathbf{N}} \mu(E_n);$$

d'altro canto, ogni somma parziale di tale serie,

$$\sum_{j=1}^n \mu(E_j) = \mu(\cup_{j=1}^n E_j)$$

è limitata da $\|f\|_1/\beta$, perché $\cup_{j=1}^n E_j$ è un sottoinsieme misurabile di $A_{\alpha,\beta}$ con misura finita; quindi $\mu(A_{\alpha,\beta}) \leq \|f\|_1/\beta$.

Se, invece, $\beta \leq 0$, allora si ha necessariamente $\alpha < 0$; si può ora applicare lo stesso ragionamento a $-f$ e a $-\alpha$ anziché a f e a β . Così si giunge nuovamente alla conclusione $\mu(A_{\alpha,\beta}) < +\infty$.

Si ponga ora

$$B_\beta := \left\{ \sup_{n \in \mathbf{N}} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j > \beta \right\}.$$

Poiché $A_{\alpha,\beta} \subset B_\beta$, la (7.4) dà

$$\int_{A_{\alpha,\beta}} f d\mu = \int_{A_{\alpha,\beta} \cap B_\beta} f d\mu \geq \beta \mu(A_{\alpha,\beta} \cap B_\beta) = \beta \mu(A_{\alpha,\beta}). \quad (7.6)$$

Sostituendo f , β , e α con $-f$, $-\alpha$ e $-\beta$ rispettivamente si ottiene $(-f)^* = -f_*$, $(-f)_* = -f^*$ e

$$\int_{A_{\alpha,\beta}} f d\mu \leq \alpha \mu(A_{\alpha,\beta}). \quad (7.7)$$

Sottraendo la (7.6) dalla (7.7) si ottiene $(\alpha - \beta) \mu(A_{\alpha,\beta}) \geq 0$, che, poiché è $\alpha < \beta$, implica $\mu(A_{\alpha,\beta}) = 0$. Ora, esprimendo l'insieme $\{f_* < f^*\}$ come un'unione numerabile,

$$\{f_* < f^*\} = \bigcup_{\substack{\alpha, \beta \in \mathbf{Q} \\ \alpha < \beta}} A_{\alpha,\beta},$$

si ottiene $\mu(\{f_* < f^*\}) = 0$, cioè $f^* = f_*$ q.o.. Così si è mostrato che

$$\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j$$

converge q.o. a $f^* = f_*$.

Rimane da mostrare che il limite f^* è integrabile. A tal fine, si usa il lemma di FATOU. Posto

$$g_n := \left| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right|,$$

si ha

$$0 \leq \int_{\Omega} g_n d\mu = \int_{\Omega} \left| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right| d\mu \leq \frac{1}{n} \sum_{j=0}^{n-1} \int_{\Omega} |f \circ T^j| d\mu = \|f\|_1,$$

sicché, per il lemma di FATOU,

$$\int |f_*| d\mu = \int \liminf_{n \rightarrow +\infty} g_n d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} g_n d\mu \leq \|f\|_1 < +\infty.$$

Perciò $|f_*|$ è integrabile.

(c) Per $k \in \mathbf{Z}$ e $n \in \mathbf{N}$ si ponga

$$A(k, n) := \left\{ \frac{k}{n} \leq f^* < \frac{k+1}{n} \right\}$$

Fissato arbitrariamente $\varepsilon > 0$, si ha $A(k, n) \cap B_{k/n-\varepsilon} = A(k, n)$ e, per il Corollario 7.1,

$$\int_{A(k, n)} f d\mu = \int_{A(k, n) \cap B_{k/n-\varepsilon}} f d\mu \geq \left(\frac{k}{n} - \varepsilon \right) \mu(A(k, n)),$$

cioè

$$\int_{A(k, n)} f d\mu \geq \frac{k}{n} \mu(A(k, n)).$$

Di qui segue

$$\int_{A(k, n)} f^* d\mu \geq \frac{k+1}{n} \mu(A(k, n)) \leq \frac{1}{n} \mu(A(k, n)) + \int_{A(k, n)} f d\mu.$$

Da quest'ultima disuguaglianza, sommando su k scende, per ogni $n \in \mathbf{N}$,

$$\int_{\Omega} f^* d\mu \leq \frac{1}{n} \mu(\Omega) + \int_{\Omega} f d\mu,$$

e, perciò, poiché $\mu(\Omega) < +\infty$,

$$\int_{\Omega} f^* d\mu \leq \int_{\Omega} f d\mu.$$

Lo stesso ragionamento applicato a $-f$ dà

$$\int_{\Omega} (-f)^* d\mu \leq \int_{\Omega} (-f) d\mu \quad \text{cioè} \quad - \int_{\Omega} f_* d\mu \leq - \int_{\Omega} f d\mu.$$

Poiché $f^* = f_*$ q.o., si ha

$$\int_{\Omega} f_* d\mu \geq \int_{\Omega} f d\mu,$$

sicché vale la (7.1). □

OSSERVAZIONE 7.2. Si mostra facilmente con un esempio che l'ipotesi $\mu(\Omega) < +\infty$ che lo spazio abbia misura finita è essenziale affinché valga la (7.1). Si consideri lo spazio misurale $(\mathbf{R}, \mathcal{B}, \lambda)$, ove λ è la misura di LEBESGUE definita sui boreliani \mathcal{B} . La trasformazione $T : \mathbf{R} \rightarrow \mathbf{R}$ definita da $Tx := x + 1$ ovviamente conserva la misura. Per f , la funzione indicatrice dell'intervallo $[0, 1]$ si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j = 0 = f^*,$$

sicch 

$$\int_{\Omega} f^* d\lambda = 0 \quad \text{mentre} \quad \int_{\Omega} f d\lambda = 1.$$

8. Corollar  del teorema di Birkhoff

Il teorema di BIRKHOFF assume una formulazione pi  forte quando la trasformazione T   ergodica.

COROLLARIO 8.1. *Nelle stesse ipotesi del Teorema di BIRKHOFF, se la trasformazione $T : \Omega \rightarrow \Omega$   ergodica, allora la funzione limite f^*   costante q.o.. In particolare*

(a) *se la misura μ   finita, $\mu(\Omega) < +\infty$,  *

$$f^* = \frac{1}{\mu(\Omega)} \int_{\Omega} f d\mu; \tag{8.1}$$

(b) *se la misura μ   infinita, $\mu(\Omega) = +\infty$, allora $f^* = 0$ q.o..*

DIMOSTRAZIONE. Per il teorema di BIRKHOFF la funzione f^*   invariante rispetto a T , ma le uniche funzioni invarianti quando T   ergodica sono costanti q.o.. Se $\mu(\Omega) < +\infty$, vale

$$\int_{\Omega} f^* d\mu = \int_{\Omega} f d\mu,$$

onde la (8.1). Se, invece, $\mu(\Omega) = +\infty$, allora $f^* = 0$ q.c., perch  questa   l'unica funzione che sia, allo stesso tempo, costante ed integrabile. \square

OSSERVAZIONE 8.1. Nel caso di uno spazio di probabilit    facile dare una forma propbabilistica alla funzione limite f^* del teorema di BIRKHOFF. Si osserva subito che la famiglia \mathcal{I} degli insiemi invarianti rispetto a T ,

$$\mathcal{I} := \{ A \in \mathcal{F} : T^{-1}A = A \},$$

è una tribú. Inoltre se una funzione $\varphi \in L^1$ è invariante, $\varphi \circ T = \varphi$ e se A è un insieme invariante, $T^{-1}A = A$, allora è invariante anche $\varphi 1_A$; infatti

$$(\varphi 1_A) \circ T = (\varphi \circ T) (1_A \circ T) = \varphi 1_{T^{-1}A} = \varphi 1_A.$$

La (7.1), applicata alla funzione $f 1_A$ dà

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

sicché la funzione limite f^* è la speranza condizionata di f rispetto alla tribú \mathcal{I} degli insiemi invarianti,

$$f^* = E(f | \mathcal{I}).$$

Il corollario che segue è particolarmente importante perché caratterizza le trasformazioni ergodiche in uno spazio di probabilità.

COROLLARIO 8.2. *Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità e sia $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura. Sono allora equivalenti le affermazioni*

- (a) T è ergodica;
- (b) per ogni coppia A e B di insiemi misurabili, $A, B \in \mathcal{F}$, si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) = \mu(A)\mu(B). \quad (8.2)$$

DIMOSTRAZIONE. (a) \implies (b) Sia T ergodica. Se $f = 1_A$ con A misurabile, il teorema di BIRKHOFF e il Corollario 8.1 danno

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} 1_A \circ T^j = \mu(A) \quad \text{q.c.},$$

relazione dalla quale, moltiplicando ambo i membri per 1_B con $B \in \mathcal{F}$, scende

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} (1_A \circ T^j) 1_B = \mu(A) 1_B \quad \text{q.c..}$$

L'asserto segue ora integrando e ricorrendo al teorema di convergenza dominata.

(b) \implies (a) Sia A un insieme misurabile invariante rispetto a T , $A \in \mathcal{F}$ e $T^{-1}A = A$; prendendo $B = A$ nella (8.2), si ottiene

$$\mu(A) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(A) = [\mu(A)]^2,$$

donde $\mu(A) = 0$ oppure $\mu(A) = 1$. \square

COROLLARIO 8.3 (Teorema ergodico medio di VON NEUMANN). *Sia T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$. Se f è una funzione di L^p , con $p \in [0, +\infty[$, allora esiste una funzione $f^* \in L^p$, invariante rispetto a T , $U_T f^* = f^* \circ T = f^*$ tale che*

$$\lim_{n \rightarrow +\infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - f^* \right\|_p = 0. \quad (8.3)$$

DIMOSTRAZIONE. Sia $\varphi : \Omega \rightarrow \mathbf{R}$ una funzione misurabile e limitata, sicché φ appartiene a L^∞ e, quindi, a L^p per ogni $p \in [0, +\infty]$. Scende dal teorema di BIRKHOFF che

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ T^j = \varphi^* \quad \text{q.c.}$$

ove φ^* è una funzione di L^∞ e quindi di L^p per ogni $p \in [0, +\infty]$. Allora

$$\lim_{n \rightarrow +\infty} \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ T^j - \varphi^* \right|^p = 0 \quad \text{q.c.}$$

sicché il teorema di convergenza dominata dà

$$\lim_{n \rightarrow +\infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ T^j - \varphi^* \right\|_p = 0.$$

Fissato arbitrariamente $\varepsilon > 0$, esiste cosí un numero naturale $N(\varepsilon, \varphi)$ tale che per ogni $n \geq N(\varepsilon, \varphi)$ e per ogni $k \in \mathbf{N}$, si abbia

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ T^j - \frac{1}{n+k} \sum_{j=0}^{n+k-1} \varphi \circ T^j \right\|_p < \frac{\varepsilon}{2}.$$

Per $f \in L^p$ con $p \in [0, +\infty[$ si ponga

$$S(n, f) := \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j.$$

La successione $\{S(n, f) : n \in \mathbf{N}\}$ è di CAUCHY in L^p . Intanto, si ha

$$\|S(n, f)\|_p = \left\| \frac{1}{n} \sum_{j=0}^{n-1} U_T^j f \right\|_p \leq \frac{1}{n} \sum_{j=0}^{n-1} \|U_T^j f\|_p = \frac{1}{n} \sum_{j=0}^{n-1} \|f\|_p = \|f\|_p$$

Si scelga $\varphi \in L^\infty$ in modo che sia $\|f - \varphi\|_p < \varepsilon/4$. Allora, per ogni $n \geq N(\varepsilon, \varphi)$ e per ogni $k \in \mathbf{N}$, è

$$\begin{aligned} & \|S(n, f) - S(n+k, f)\|_p \\ & \leq \|S(n, f) - S(n, \varphi)\|_p + \|S(n, \varphi) - S(n+k, \varphi)\|_p \\ & \quad + \|S(n+k, \varphi) - S(n+k, f)\|_p \\ & < \|f - \varphi\|_p + \frac{\varepsilon}{2} + \|f - \varphi\|_p < \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon. \end{aligned}$$

Siccome L^p è completo, esiste $f^* \in L^p$ tale che

$$\lim_{n \rightarrow +\infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} U_T^j f - f^* \right\|_p = 0.$$

Infine, poiché

$$\frac{n+1}{n} S(n+1, f) - S(n, f \circ T) = \frac{f}{n},$$

f^* risulta essere invariante, cioè $U_T f^* = f^*$. □

9. Proprietà più forti dell'ergodicità

Si è visto che in uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura è ergodica se, e solo se, per ogn scelta degli insiemi A e B in \mathcal{F} , vale la (8.2).

DEFINIZIONE 9.1. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$, una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura si dice

- (a) *totalmente ergodica* se, per ogni $k \in \mathbf{N}$, è ergodica la trasformazione T^k ;
- (b) *debolmente mescolante* se, per ogn scelta degli insiemi A e B in \mathcal{F} , è

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} |\mu(T^{-j} A \cap B) - \mu(A)\mu(B)| = 0; \quad (9.1)$$

- (c) *fortemente mescolante* se, per ogn scelta degli insiemi A e B in \mathcal{F} , è

$$\lim_{n \rightarrow +\infty} \mu(T^{-n} A \cap B) = \mu(A)\mu(B). \quad (9.2)$$

OSSERVAZIONE 9.1. Nella definizione di mescolanza forte non è, in genere possibile sostituire la misura di probabilità μ con una misura arbitraria. Infatti, se nella relazione

$$\lim_{n \rightarrow +\infty} \mu(T^{-n} A \cap B) = \mu(A)\mu(B). \quad (9.3)$$

si prende $A = B = \Omega$ si ottiene $\mu(\Omega) = [\mu(\Omega)]^2$, sicchè, se μ è una misura finita e non banale ($\mu(A) = 0$ per ogni $A \in \mathcal{F}$), è necessariamente $\mu(\Omega) = 1$. Se, invece, μ non è una misura finita, $\mu(\Omega) = +\infty$, si supponga che esista un insieme misurabile A di misura finita e non nulla, $0 < \mu(A) < +\infty$. Allora, prendendo nella (9.3), $B = \Omega$ si trova

$$+\infty > \mu(A) = \mu(T^{-n} A) = \mu(T^{-n} A \cap \Omega) \xrightarrow{n \rightarrow +\infty} \mu(A) \mu(\Omega) = +\infty,$$

che è una contraddizione. La (9.3) ha dunque senso, se $\mu(\Omega) = +\infty$, solo per la misura $\mu : \mathcal{F} \rightarrow \overline{\mathbf{R}}_+$ definita da

$$\mu(A) := \begin{cases} 0, & A = \emptyset, \\ +\infty, & A \neq \emptyset. \end{cases}$$

Tale misura riveste però scarso interesse.

TEOREMA 9.1. *Siano $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità e T una trasformazione che conserva la misura. Allora*

- (a) *se T è fortemente mescolante, essa è anche debolmente mescolante;*
- (b) *se T è debolmente mescolante, essa è anche totalmente ergodica;*
- (c) *se T è totalmente ergodica, essa è anche ergodica.*

DIMOSTRAZIONE. (a) La (9.2) si può scrivere anche nella forma

$$\lim_{n \rightarrow +\infty} |\mu(T^{-n} A \cap B) - \mu(A) \mu(B)| = 0;$$

l'asserto segue immediatamente ricordando che una successione convergente converge anche nel senso di CESÀRO.

(b) Si deve mostrare che, per ogni $k \in \mathbf{N}$ e per ogni scelta di due insiemi misurabili A e B , si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-k \cdot j} A \cap B) = \mu(A) \mu(B).$$

Ora

$$\begin{aligned}
0 &\leq \left| \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-kj} A \cap B) - \mu(A) \mu(B) \right| \\
&= \left| \frac{1}{n} \sum_{j=0}^{n-1} [\mu(T^{-kj} A \cap B) - \mu(A) \mu(B)] \right| \\
&\leq \frac{1}{n} \sum_{j=0}^{n-1} |\mu(T^{-kj} A \cap B) - \mu(A) \mu(B)| \\
&\leq \frac{k(n-1)}{n} \frac{1}{k(n-1)} \sum_{i=0}^{k(n-1)} |\mu(T^{-i} A \cap B) - \mu(A) \mu(B)| \xrightarrow{n \rightarrow +\infty} 0.
\end{aligned}$$

La (c) è ovvia. □

Rimandiamo, per il momento, la presentazione di esempi che illustrino i vari tipi di trasformazione, anticipando solo, sin da ora, che non sarà facile dare esempi di trasformazioni debolmente mescolanti.

Il teorema che segue mostra che le condizioni di ergodicità, di forte ergodicità, di mescolanza debole e di mescolanza forte possono essere controllate anziché su tutti i sottoinsiemi misurabili solo su quella di una sottoalgebra \mathcal{A} di \mathcal{F} che la generi, $\mathcal{F}(\mathcal{A}) = \mathcal{F}$.

TEOREMA 9.2. *Siano $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità, T una trasformazione che conserva la misura e \mathcal{A} un'algebra che genera la tribù \mathcal{F} , $\mathcal{F}(\mathcal{A}) = \mathcal{F}$. Allora*

- (a) T è ergodica se, e solo se, la (8.2) vale per ogni di insiemi A e B appartenenti all'algebra \mathcal{A} ;
- (b) T è debolmente mescolante se, e solo se, la (9.1) vale per ogni di insiemi A e B appartenenti all'algebra \mathcal{A} ;
- (c) T è fortemente mescolante se, e solo se, la (9.2) vale per ogni di insiemi A e B appartenenti all'algebra \mathcal{A} .

DIMOSTRAZIONE. In tutti e tre i casi basta dimostrare che basta controllare le relazioni in questione quando gli insiemi considerati appartengano all'algebra \mathcal{A} .

Si osservi che, fissati $\varepsilon > 0$ e gli insiemi A e B in \mathcal{F} , è possibile scegliere A_0 e B_0 in \mathcal{A} in maniera che sia

$$\mu(A \Delta A_0) < \varepsilon \quad \text{e} \quad \mu(B \Delta B_0) < \varepsilon.$$

Pertanto

$$\begin{aligned}
& \mu [(T^{-n} A \cap B) \Delta (T^{-n} A_0 \cap B_0)] \\
&= \mu [\{T^{-n} A \Delta (T^{-n} A_0 \cap B_0)\} \cap \{B \Delta (T^{-n} A_0 \cap B_0)\}] \\
&\leq \mu [\{T^{-n} (A \Delta A_0)\} \cap (T^{-n} A_0 \cap B_0)] + \mu [(B \Delta T^{-n} A_0) \cap (B \Delta B_0)] \\
&\leq \mu [T^{-n} (A \Delta A_0)] + \mu (B \Delta B_0) = \mu (A \Delta A_0) + \mu (B \Delta B_0) < 2\varepsilon.
\end{aligned}$$

Perciò, posto $C := T^{-n} A \cap B$, $C_0 := T^{-n} A_0 \cap B_0$ e $D := C \cap C_0$, si ha

$$\begin{aligned}
& |\mu (T^{-n} A \cap B) - \mu (T^{-n} A_0 \cap B_0)| = |\mu(C) - \mu(C_0)| \\
& |\mu(C \setminus D) + \mu(D) - \mu(C_0 \setminus D) - \mu(D)| = |\mu(C \setminus D) - \mu(C_0 \setminus D)| \\
& \leq \mu(C \setminus D) + \mu(C_0 \setminus D) = \mu [(T^{-n} A \cap B) \Delta (T^{-n} A_0 \cap B_0)] < 2\varepsilon.
\end{aligned}$$

Così, per ogni $n \in \mathbf{N}$, si ha

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{j=1}^{n-1} \mu (T^{-j} A \cap B) - \mu(A) \mu(B) \right| \\
& \leq \left| \frac{1}{n} \sum_{j=1}^{n-1} \mu (T^{-j} A \cap B) - \mu (T^{-j} A_0 \cap B_0) \right| \\
& \quad + \left| \frac{1}{n} \sum_{j=1}^{n-1} \mu (T^{-j} A_0 \cap B_0) - \mu(A_0) \mu(B_0) \right| \\
& \quad + |\mu(A_0) \mu(B_0) - \mu(A_0) \mu(B)| + |\mu(A_0) \mu(B) - \mu(A) \mu(B)| \\
& \leq 2\varepsilon + \left| \frac{1}{n} \sum_{j=1}^{n-1} \mu (T^{-j} A_0 \cap B_0) - \mu(A_0) \mu(B_0) \right| + 2\varepsilon,
\end{aligned}$$

che dimostra la (a).

(b) Analogamente a quanto fatto sopra, si ottiene, per ogni $j \in \mathbf{Z}_+$

$$\begin{aligned}
& |\mu (T^{-j} A \cap B) - \mu(A) \mu(B)| \\
& \leq |\mu (T^{-j} A \cap B) - \mu (T^{-j} A_0 \cap B_0)| \\
& \quad + |\mu (T^{-j} A_0 \cap B_0) - \mu(A_0) \mu(B_0)| \\
& \quad + |\mu(A_0) \mu(B_0) - \mu(A_0) \mu(B)| + |\mu(A_0) \mu(B) - \mu(A) \mu(B)| \\
& \leq 4\varepsilon + |\mu (T^{-j} A_0 \cap B_0) - \mu(A_0) \mu(B_0)|;
\end{aligned}$$

di qui, prendendo la media aritmetica, scende l'asserto. L'ultima sequenza di diseuguaglianze stabilisce anche la (c). \square

Possiamo sfruttare il corollario appena dato per stabilire quando la traslazione di MARKOV bilatera sia ergodica.

TEOREMA 9.3. *Per una traslazione di MARKOV bilatera- (p, Π) T sono equivalenti le affermazioni:*

- (a) T è ergodica;
- (b) la matrice di transizione Π è ergodica (da ogni stato si può raggiungere un qualsiasi altro stato).

DIMOSTRAZIONE. (a) \implies (b) Si indichi con e_j la funzione indicatrice del cilindro

$$\{ \{x_n : n \in \mathbf{Z}\} : x_0 = j \}.$$

Allora, per il corollario 8.1, si ha q.c.

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{s=0}^{n-1} e_j(T^s \omega) = \int e_j d\mu = p_j > 0.$$

Moltiplicando ambo i membri per e_i , integrando ed usando il teorema di convergenza dominata si ottiene

$$\frac{1}{n} \sum_{s=0}^{n-1} p_i p_{ij}^{(s)} = \frac{1}{n} \int e_i(\omega) e_j(T^s \omega) d\mu(\omega) \xrightarrow{n \rightarrow +\infty} p_i p_j.$$

Pertanto la successione $\{\Pi^n\}$ converge nel senso di CESÀRO ad una matrice limite $A = (a_{ij})$, gli elementi della quale sono dati, quali che sia lo stato $i \in \{0, 1, \dots, k-1\}$, da

$$a_{ij} = p_j > 0.$$

Quali che siano gli stati i e j si ha

$$\frac{1}{n} \sum_{s=0}^{n-1} p_{ij}^{(s)} \xrightarrow{n \rightarrow +\infty} a_{ij} = p_j > 0,$$

sicch  esiste un numero naturale s tale che $p_{ij}^s > 0$. Dunque la matrice Π   ergodica.

(b) \implies (a) Se Π   ergodica pu  essere regolare o ciclica ([36]). In entrambi casi la successione $\{\Pi^n : n \in \mathbf{N}\}$ converge nel senso di CES RO ad una matrice limite A che ha tutte le righe eguali e tutti gli elementi strettamente positivi. Per dimostrare che T   ergodica basta, in virt  del Teorema precedente, far vedere che, per ogni scelta di due cilindri A e B con

$$A = \{ \{x_n : n \in \mathbf{Z}\} : x_a = i_0, x_{a+1} = i_1, \dots, x_{a+r} = i_r \}$$

$$B = \{ \{x_n : n \in \mathbf{Z}\} : x_b = j_0, x_{b+1} = j_1, \dots, x_{b+r} = j_s \}$$

 

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{h=0}^{n-1} \mu(T^{-h} A \cap B) = \mu(A) \mu(B).$$

Ora, se $h > b + s - a$ si ha

$$\mu(T^{-h} A \cap B) = p_{j_0} p_{j_0 j_1} \cdots p_{j_{s-1} j_s} p_{j_s i_0}^{a+h-b-s} p_{i_0 i_1} \cdots p_{i_{r-1} i_r}$$

e quindi

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{h=0}^{n-1} \mu(T^{-h} A \cap B) &= p_{j_0} p_{j_0 j_1} \cdots p_{j_{s-1} j_s} p_{i_0} p_{i_0 i_1} \cdots p_{i_{r-1} i_r} \\ &= \mu(A) \mu(B), \end{aligned}$$

sicch  T   ergodica. □

Il seguente risultato appartiene all'analisi reale "elementare", ma consente di porre in forma differente la propriet  di mescolanza debole.

TEOREMA 9.4. *Per una successione limitata $\{x_n : n \in \mathbf{N}\}$ di numeri reali sono equivalenti le condizioni:*

(a) *la successione $\{|x_n| : n \in \mathbf{Z}_+\}$ converge a 0 nel senso di CES RO,*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} |x_j| = 0;$$

(b) *esiste un sottoinsieme J di \mathbf{Z}_+ di densit  nulla, tale, cio , che*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \text{card}[J \cap \{0, 1, \dots, n-1\}] = 0,$$

per il quale risulta

$$\lim_{\substack{n \rightarrow +\infty \\ n \notin J}} x_n = 0;$$

(c) *vale*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} |x_j|^2 = 0.$$

DIMOSTRAZIONE. Per ogni sottoinsieme F di \mathbf{Z}_+ , si ponga, per semplicit ,

$$\alpha_F(n) := \text{card}[F \cap \{0, 1, \dots, n-1\}].$$

(a) \implies (b) Per ogni $k \in \mathbf{N}$ sia

$$J_k := \left\{ n \in \mathbf{Z}_+ : |x_n| \geq \frac{1}{k} \right\}.$$

Pertanto

$$J_1 \subset J_2 \subset \cdots \subset J_n \subset \cdots \tag{9.4}$$

Inoltre, ogni J_k ha densità nulla, perché

$$\frac{1}{n} \sum_{j=0}^{n-1} |x_j| \geq \frac{1}{n} \frac{1}{k} \alpha_{J_k}(n).$$

Esistono perciò numeri naturali $0 \leq s_0 < s_1 < \dots < s_n < \dots$ tali che per $n \geq s_k$ si abbia

$$\frac{1}{n} \alpha_{J_{k+1}}(n) < \frac{1}{k+1}.$$

Si ponga

$$J := \cup_{k \in \mathbf{Z}_+} \{J_{k+1} \cap [s_k, s_{k+1}[\};$$

Ora dalla (9.4) a da $s_k \leq n < s_{k+1}$ segue che

$$\begin{aligned} J \cap \{0, 1, \dots, n-1\} &= [J_{k-1} \cap \{0, 1, \dots, s_k - 1\}] \cup [J_k \cap \{s_k, s_k + 1, \dots, n-1\}] \\ &\subset [J_{k-1} \cap \{0, 1, \dots, s_k - 1\}] \cup [J_k \cap \{0, 1, \dots, n-1\}], \end{aligned}$$

e, perciò,

$$\begin{aligned} \frac{1}{n} \alpha_J(n) &\leq \frac{1}{n} \{\alpha_{J_k}(s_k) + \alpha_{J_{k+1}}(n)\} \leq \frac{1}{n} \{\alpha_{J_k}(n) + \alpha_{J_{k+1}}(n)\} \\ &< \frac{1}{k-1} + \frac{1}{k}. \end{aligned}$$

Quando n tende a $+\infty$, anche k tende a $+\infty$; di qui segue che J ha densità nulla. Se $n \geq s_k$ e $n \notin J$ scende che n non appartiene a J_{k+1} e perciò,

$$|x_n| < \frac{1}{k+1}$$

che dà l'asserto (b).

(b) \implies (a) Si supponga che sia $|x_n| \leq H$ per ogni $n \in \mathbf{Z}_+$. Fissato $\varepsilon > 0$, esistono due numeri naturali $n_1 = n_1(\varepsilon) \in \mathbf{N}$ e $n_2 = n_2(\varepsilon) \in \mathbf{N}$ tali che sia $|x_n| < \varepsilon$ per ogni naturale $n \geq n_1$ con $n \notin J$ e

$$\frac{1}{n} \alpha_J(n) < \varepsilon,$$

per ogni $n \geq n_2$. Posto $n_0 := \max\{n_1, n_2\}$, per ogni $n \geq n_0$, è

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-1} |x_j| &= \frac{1}{n} \left[\sum_{j \in J \cap \{0, 1, \dots, n-1\}} |x_j| + \sum_{j \notin J \cap \{0, 1, \dots, n-1\}} |x_j| \right] \\ &< \frac{H}{n} \alpha_J(n) + \varepsilon < (H+1)\varepsilon, \end{aligned}$$

onde l'asserto.

(a) \iff (c) Basta osservare che, alla luce di quanto è stato dimostrato sinora si ha equivalenza tra le due condizioni

$$\lim_{\substack{n \rightarrow +\infty \\ n \notin J}} |x_n| = 0 \quad \text{e} \quad \lim_{\substack{n \rightarrow +\infty \\ n \notin J}} |x_n|^2 = 0.$$

□

Sono ora evidenti i seguenti corollari.

COROLLARIO 9.1. *Per una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura su uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ sono equivalenti le affermazioni:*

- (a) T è debolmente mescolante;
- (b) per ogni coppia A e B di insiemi misurabili esiste $J(A, B)$ sottoinsieme di \mathbf{N} di densità nulla tale che

$$\lim_{\substack{n \rightarrow +\infty \\ n \notin J(A, B)}} \mu(T^{-n} A \cap B) = \mu(A) \mu(B).$$

- (c) per ogni coppia A e B di insiemi misurabili esiste $J(A, B)$ sottoinsieme di \mathbf{N} di densità nulla tale che

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} |\mu(T^{-j} A \cap B) - \mu(A) \mu(B)|^2 = 0.$$

Avremo modo di usare il seguente lemma.

LEMMA 9.1. *L'unione di due insiemi di densità nulla è ancora di densità nulla.*

DIMOSTRAZIONE. Siano J_1 e J_2 due sottoinsiemi di \mathbf{N} di densità nulla,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \text{card} [J_j \cap \{0, 1, \dots, n-1\}] = 0 \quad (j = 1, 2).$$

Allora

$$\begin{aligned} 0 &\leq \frac{1}{n} \text{card} [(J_1 \cup J_2) \cap \{0, 1, \dots, n-1\}] \\ &= \frac{1}{n} \text{card} [(J_1 \cap \{0, 1, \dots, n-1\}) \cup (J_2 \cap \{0, 1, \dots, n-1\})] \\ &\leq \frac{1}{n} \text{card} (J_1 \cap \{0, 1, \dots, n-1\}) \\ &\quad + \frac{1}{n} \text{card} (J_2 \cap \{0, 1, \dots, n-1\}) \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

□

10. Espressione funzionale di ergodicità e mescolanza

In questa sezione ricorreremo sistematicamente all'uso dell'operatore U_T indotto dalla trasformazione T , introdotto nella sezione 6.

TEOREMA 10.1. *Per una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura su uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ sono equivalenti le affermazioni:*

- (a) T è ergodica;
- (b) per coppia f e g di funzioni di L^2 è

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \langle U_T^j f, g \rangle = \langle f, 1 \rangle \langle 1, g \rangle; \quad (10.1)$$

- (c) per ogni funzione f di L^2 è

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \langle U_T^j f, f \rangle = \langle f, 1 \rangle \langle 1, f \rangle; \quad (10.2)$$

TEOREMA 10.2. *Per una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura su uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ sono equivalenti le affermazioni:*

- (a) T è debolmente mescolante;
- (b) per coppia f e g di funzioni di L^2 è

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \langle U_T^j f, g \rangle - \langle f, 1 \rangle \langle 1, g \rangle \right| = 0; \quad (10.3)$$

- (c) per ogni funzione f di L^2 è

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \langle U_T^j f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle \right| = 0; \quad (10.4)$$

- (d) per ogni funzione f di L^2 è

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \langle U_T^j f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle \right|^2 = 0; \quad (10.5)$$

TEOREMA 10.3. *Per una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura su uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ sono equivalenti le affermazioni:*

- (a) T è fortemente mescolante;
- (b) per coppia f e g di funzioni di L^2 è

$$\lim_{n \rightarrow +\infty} \langle U_T^n f, g \rangle = \langle f, 1 \rangle \langle 1, g \rangle; \quad (10.6)$$

(c) per ogni funzione f di L^2 è

$$\lim_{n \rightarrow +\infty} \langle U_T^n f, f \rangle = \langle f, 1 \rangle \langle 1, f \rangle; \quad (10.7)$$

I tre teoremi appena enunciati si dimostrano in maniera molto simile. Pertanto basterà dare la dimostrazione di uno solo di essi, per esempio dell'ultimo.

DIMOSTRAZIONE DEL TEOREMA 10.3. (b) \implies (a) Nella (10.6) basta prendere $f = 1_A$ e $g = 1_B$ con $A, B \in \mathcal{F}$.

(a) \implies (c) Per ogni coppia A e B di insiemi misurabili si ha

$$\begin{aligned} \lim_{n \rightarrow +\infty} \langle U_T^n 1_A, 1_B \rangle &= \lim_{n \rightarrow +\infty} \int 1_A \circ T^n 1_B d\mu = \lim_{n \rightarrow +\infty} \int_{T^{-n} A \cap B} d\mu \\ &= \lim_{n \rightarrow +\infty} \mu(T^{-n} A \cap B) = \mu(A) \mu(B) = \langle 1_A, 1 \rangle \langle 1, 1_B \rangle. \end{aligned}$$

Se in quest'ultima relazione si tiene fisso B , si ha, per la linearità del prodotto interno rispetto al primo fattore,

$$\lim_{n \rightarrow +\infty} \langle U_T^n h, 1_B \rangle = \langle h, 1 \rangle \langle 1, 1_B \rangle,$$

per ogni funzione semplice h . Si tenga ora fissa la funzione h , sicché dall'ultima relazione scende

$$\lim_{n \rightarrow +\infty} \langle U_T^n h, h \rangle = \langle h, 1 \rangle \langle 1, h \rangle;$$

pertanto la (10.7) vale per le funzioni semplici.

Sia, ora f una funzione di L^2 e si fissi, arbitrariamente, $\varepsilon > 0$. Si può scegliere una funzione semplice h in modo che sia $\|f - h\|_2 < \varepsilon$. Esiste anche un numero naturale $n_0 = n_0(\varepsilon)$ tale che, per ogni $n \geq n_0$, sia

$$|\langle U_T^n h, h \rangle - \langle h, 1 \rangle \langle 1, h \rangle| < \varepsilon.$$

Per ogni $n \geq n_0$, si ha

$$\begin{aligned} &|\langle U_T^n f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle| \\ &\leq |\langle U_T^n f, f \rangle - \langle U_T^n h, f \rangle| + |\langle U_T^n h, f \rangle - \langle U_T^n h, h \rangle| \\ &\quad + |\langle U_T^n h, h \rangle - \langle h, 1 \rangle \langle 1, h \rangle| + |\langle h, 1 \rangle \langle 1, h \rangle - \langle f, 1 \rangle \langle 1, h \rangle| \\ &\quad + |\langle f, 1 \rangle \langle 1, h \rangle - \langle f, 1 \rangle \langle 1, f \rangle| \\ &= |\langle U_T^n (f - h), f \rangle| + |\langle U_T^n h, f - h \rangle| + \varepsilon + |\langle 1, h \rangle| |\langle h - f, 1 \rangle| \\ &\quad + |\langle f, 1 \rangle| |\langle 1, h - f \rangle| \\ &\leq \|f - h\|_2 \|f\|_2 + \|h\|_2 \|h - f\|_2 + \varepsilon + \|h\|_2 \|h - f\|_2 + \|f\|_2 \|h - f\|_2 \\ &\leq \varepsilon \{1 + 2\|h\|_2 + 2\|f\|_2\}, \end{aligned}$$

onde l'asserto.

(c) \implies (b) Sia f una funzione di L^2 e sia \mathcal{H}_f il piú piccolo sottospazio chiuso di L^2 che contiene f e le funzioni costanti e che sia invariante rispetto a U_T , $U_T \mathcal{H}_f \subset \mathcal{H}_f$. Posto

$$\mathcal{G}_f := \left\{ \varphi \in L^2 : \lim_{n \rightarrow +\infty} \langle U_T^n f, \varphi \rangle = \langle f, 1 \rangle \langle 1, \varphi \rangle \right\},$$

si ha

$$\mathcal{H}_f \subset \mathcal{G}_f \subset L^2.$$

Infatti, per definizione, \mathcal{G}_f è un sottospazio di L^2 e contiene f per ipotesi. Inoltre contiene le funzioni costanti

$$\begin{aligned} \langle U_T^n f, c \rangle &= \bar{c} \int U_T^n f d\mu = \bar{c} \int f \circ T^n d\mu = \bar{c} \int f d(\mu \circ T^{-n}) \\ &= \bar{c} \int f d\mu = \langle f, 1 \rangle \langle 1, c \rangle. \end{aligned}$$

Infine \mathcal{G}_f è invariante rispetto a U_T ; infatti se $\varphi \in \mathcal{G}_f$ si ha

$$\begin{aligned} \lim_{n \rightarrow +\infty} \langle U_T^n f, U_T \varphi \rangle &= \lim_{n \rightarrow +\infty} \int (f \circ T^n) \overline{(\varphi \circ T)} d\mu \\ &= \lim_{n \rightarrow +\infty} \int (f \circ T^{n-1}) \overline{\varphi} d\mu = \lim_{n \rightarrow +\infty} \langle U_T^{n-1} f, \varphi \rangle \\ &= \langle f, 1 \rangle \langle 1, \varphi \rangle = \langle f, 1 \rangle \int \overline{\varphi} d\mu \\ &= \langle f, 1 \rangle \int \overline{\varphi \circ T} d\mu = \langle f, 1 \rangle \langle 1, U_T \varphi \rangle, \end{aligned}$$

sicché anche $U_T \varphi$ appartiene a \mathcal{G}_f .

Se φ è una funzione ortogonale al sottospazio \mathcal{H}_f , $\varphi \perp \mathcal{H}_f$, si ha

$$\langle U_T^n f, \varphi \rangle = 0$$

per ogni $n \in \mathbf{N}$. Inoltre $\langle 1, \varphi \rangle = 0$. Ma allora si ha $\mathcal{H}_f^\perp \subset \mathcal{G}_f$ che insieme all'inclusione $\mathcal{H}_f \subset \mathcal{G}_f$ dà $\mathcal{G}_f = L^2$. \square

Sia $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$. Si può definire la trasformazione $T \times T : \Omega \times \Omega \rightarrow \Omega \times \Omega$ mediante

$$(T \times T)(\omega, \omega') := (T(\omega), T(\omega')).$$

La trasformazione cosí definita conserva la misura sullo spazio di probabilità $(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F}, \mu \otimes \mu)$. Infatti ciò è vero per i rettangoli misurabili (gli insiemi di $\mathcal{F} \times \mathcal{F}$) e, quindi, per le unioni disgiunte e finite di tali rettangoli. In virtù del Teorema 2.2, ciò basta a stabilire che $T \times T$ conserva la misura.

TEOREMA 10.4. *Per una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura su uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ sono equivalenti le affermazioni*

- (a) T è debolmente mescolante;
- (b) $T \times T$ è ergodica;
- (c) $T \times T$ è debolmente mescolante.

DIMOSTRAZIONE. (a) \implies (c) Siano A, B, C e D insiemi di \mathcal{F} . Per il Corollario 9.1, esistono due sottoinsiemi J_1 e J_2 di \mathbf{N} , entrambi di densità nulla, tali che

$$\begin{aligned} \lim_{\substack{n \rightarrow +\infty \\ n \notin J_1}} \mu(T^{-n} A \cap B) &= \mu(A) \mu(B) \\ \lim_{\substack{n \rightarrow +\infty \\ n \notin J_2}} \mu(T^{-n} C \cap D) &= \mu(C) \mu(D). \end{aligned}$$

Di conseguenza, ricorrendo al Lemma 9.1, si ha

$$\begin{aligned} &\lim_{\substack{n \rightarrow +\infty \\ n \notin J_1 \cup J_2}} (\mu \otimes \mu) [(T \times T)^{-n} (A \times C) \cap (B \times D)] \\ &= \lim_{\substack{n \rightarrow +\infty \\ n \notin J_1 \cup J_2}} \mu(T^{-n} A \cap B) \mu(T^{-n} C \cap D) \\ &= \mu(A) \mu(B) \mu(C) \mu(D) = [(\mu \otimes \mu)(A \times C)] [(\mu \otimes \mu)(C \times D)]. \end{aligned}$$

L'asserto vale dunque per i rettangoli misurabili di $\mathcal{F} \otimes \mathcal{F}$ e quindi anche per le unioni disgiunte e finite di tali rettangoli. Queste unioni formano un'algebra \mathcal{A} che genera la tribú prodotto $\mathcal{F} \otimes \mathcal{F}$. Alla luce del Teorema 9.2 ciò implica l'asserto.

L'implicazione (c) \implies (b) è ovvia.

(b) \implies (a) Se A e B sono insiemi misurabili si ha

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j} A \cap B) &= \frac{1}{n} \sum_{j=0}^{n-1} (\mu \otimes \mu) [(T \times T)^{-j} (A \times \Omega) \cap (B \times \Omega)] \\ &\xrightarrow{n \rightarrow +\infty} [(\mu \otimes \mu)(A \times \Omega)] [(\mu \otimes \mu)(B \times \Omega)] \\ &= \mu(A) \mu(B). \end{aligned}$$

Ma si ha anche, dall'ipotesi,

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-1} \mu^2(T^{-j} A \cap B) &= \frac{1}{n} \sum_{j=0}^{n-1} (\mu \otimes \mu) ((T \times T)^{-j} (A \times A) \cap (B \times B)) \\ &\xrightarrow{n \rightarrow +\infty} (\mu \otimes \mu)(A \times A) (\mu \otimes \mu)(B \times B) \\ &= \mu^2(A) \mu^2(B). \end{aligned}$$

Perciò

$$\begin{aligned} & \frac{1}{n} \sum_{j=0}^{n-1} \{\mu(T^{-j} A \cap B) - \mu(A) \mu(B)\}^2 \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \{\mu^2(T^{-j} A \cap B) - 2\mu(T^{-j} A \cap B) \mu(A) \mu(B) + \mu^2(A) \mu^2(B)\} \\ & \xrightarrow{n \rightarrow +\infty} 2\mu^2(A) \mu^2(B) - 2\mu^2(A) \mu^2(B) = 0 \end{aligned}$$

che conclude la dimostrazione. \square

DEFINIZIONE 10.1. Sia $T : \Omega \rightarrow \Omega$ una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$. Si dice che il numero $k \in \mathbf{C}$ è un *autovalore* di T , se esiste una funzione $f \in L^2$ con $f \neq 0$ tale che $U_T f = k f$, o, equivalentemente, tale che $f \circ T = k f$ q.c.. Si dice allora che f è un'*autofunzione* corrispondente all'autovalore k .

Si noti che tutti gli autovalori di T hanno modulo unitario poiché, se vale $U_T f = k f$, è

$$\|f\|_2^2 = \|U_T f\|_2^2 = \|k f\|_2^2 = |k|^2 \|f\|_2^2.$$

Inoltre $k = 1$ è sempre un autovalore (che corrisponde alla funzioni costanti).

DEFINIZIONE 10.2. Si dice che la trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ ha *spettro continuo* se 1 è l'unico autovalore e le funzioni q.c. costanti sono le sole autofunzioni dell'operatore U_T indotto da T .

Al prossimo risultato che lega i concetti di mescolanza debole e di spettro continuo, è opportuno premettere il seguente teorema, per la dimostrazione del quale rimando a [27].

TEOREMA 10.5. Sia $U : \mathcal{H} \rightarrow \mathcal{H}$ un operatore unitario sopra uno spazio di HILBERT (complesso) \mathcal{H} . Per ogni $f \in \mathcal{H}$ esiste un'unica misura μ_f definita sui boreliani della circonferenza unitaria $K := \{z \in \mathbf{C} : |z| = 1\}$ tale che, per ogni $n \in \mathbf{N}$, sia

$$\langle U^n f, f \rangle = \int_K z^n d\mu_f(z).$$

Si noti che se $T : \Omega \rightarrow \Omega$ è una trasformazione invertibile che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$, allora l'operatore U_T indotto da T è unitario; se, inoltre, T ha spettro continuo, allora, per ogni funzione $f \in L^2$ con $\langle f, 1 \rangle = 0$, μ_f non ha atomi.

TEOREMA 10.6. Per una trasformazione $T : \Omega \rightarrow \Omega$ che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ sono equivalenti le affermazioni

- (a) T è debolmente mescolante;
- (b) T ha spettro continuo.

DIMOSTRAZIONE. (a) \implies (b). Si supponga che T sia debolmente mescolante e, se possibile, che k sia un autovalore differente da 1, $U_t f = k f$ con $k \neq 1$. Integrando si ottiene

$$\begin{aligned} \langle f, 1 \rangle &= \int f d\mu = \int f d(\mu \circ T^{-1}) = \int (f \circ T) d\mu = \int U_T f d\mu \\ &= \langle U_T f, 1 \rangle = k \langle f, 1 \rangle. \end{aligned}$$

Di qui scende $\langle f, 1 \rangle = 0$, sicché, ricordando che $|k| = 1$, il Teorema 10.2 dà

$$\begin{aligned} 0 &= \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \langle U_T^j f, f \rangle \right| = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \langle k^j f, f \rangle \right| \\ &= \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} |k|^j |\langle f, f \rangle| = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} |\langle f, f \rangle| = |\langle f, f \rangle|. \end{aligned}$$

Dunque $f = 0$ q.c., sicché k non può essere un autovalore. Se $k = 1$ si ha necessariamente che f è q.c. costante, perché T è ergodica.

(b) \implies (a) Supposto che T abbia spettro continuo, si vuole dimostrare che T è debolmente mescolante; basta, a tal fine, dimostrare che vale la

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \langle U_T^j f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle \right|^2 = 0;$$

Se f è q.c. costante, questa è un'immediata conseguenza della disuguaglianza di SCHWARZ, che in questo caso, si riduce ad un'eguaglianza. Si può, quindi, supporre che sia $\langle f, 1 \rangle = 0$. Per il Teorema 10.5, basta far vedere che, per ogni misura μ sui boreliani di K , si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left| \int_K z^j d\mu \right|^2 = 0. \quad (10.8)$$

Ricorrendo al teorema di FUBINI si ottiene

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-1} \left| \int_K z^j d\mu \right|^2 &= \frac{1}{n} \sum_{j=0}^{n-1} \left\{ \int_K z^j d\mu(z) \right\} \left\{ \int_K \bar{w}^j d\mu(w) \right\} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \iint_{K \times K} (z \bar{w})^j d(\mu \otimes \mu)(z, w) = \iint_{K \times K} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} (z \bar{w})^j \right\} d(\mu \otimes \mu)(z, w). \end{aligned}$$

Se il punto $(z\bar{w})$ non appartiene alla diagonale di $K \times K$, ovvero se $z\bar{w} \neq 1$, allora si ha

$$\frac{1}{n} \sum_{j=0}^{n-1} (z\bar{w})^j = \frac{1}{n} \frac{1 - (z\bar{w})^n}{1 - z\bar{w}} \xrightarrow{n \rightarrow +\infty} 0.$$

Poiché la misura μ è continua, è $(\mu \otimes \mu)(z, z) = 0$, sicché l'integrando nell'ultimo integrale scritto tende a 0 q.o.; inoltre, esso è maggiorato in modulo da 1 sicché la (10.8) è ora una conseguenza del teorema di convergenza dominata. \square

11. Topologie per le trasformazioni

Vogliamo qui introdurre alcune topologie sulla famiglia delle trasformazioni che conservano la misura. Ci limiteremo, seguendo HALMOS, a considerare solo un caso particolare; supporremo così che siano verificate le seguenti ipotesi

- si considereranno solo trasformazioni invertibili sicché l'insieme nel quale si introdurrà la topologia è, di fatto, un gruppo;
- si considereranno come coincidenti le trasformazioni che differiscono su un insieme di misura nulla; in altre parole, anziché il gruppo delle trasformazioni invertibili che conservano la misura, si considererà il gruppo degli automorfismi dell'algebra di misura associata allo spazio;
- si supporrà che lo spazio di probabilità in esame sia costituito dall'intervallo unitario $[0, 1]$ munito della tribú di BOREL \mathcal{B} e della misura di LEBESGUE λ su \mathcal{B} .

Ricordiamo qui la definizione delle *topologie forte e debole* in uno spazio di HILBERT H . Data una famiglia di operatori su H , $\{T_n : n \in I\}$, ove I è un insieme diretto, si dice che tale famiglia *converge fortemente* all'operatore T se, per ogni $f \in H$, si ha

$$\|T_n f - T f\| \longrightarrow 0.$$

Si dice invece che la famiglia in questione *converge debolmente* a T se, per ogni coppia f e g di elementi di H , si ha

$$\langle T_n f, g \rangle \longrightarrow \langle T f, g \rangle.$$

È ben noto che la convergenza forte implica quella debole, senza che sia vero il viceversa. Si è visto che esiste una corrispondenza biunivoca tra gli automorfismi di un'algebra di misura e gli operatori unitari su L^2 , sicché il

gruppo degli automorfismi eredita ogni topologia della famiglia degli operatori unitari su L^2 . In quest'ultimo insieme le due topologie, forte e debole, coincidono.

TEOREMA 11.1. *Per una famiglia $\{U_n : n \in I\}$ di operatori unitari su L^2 su uno spazio di HILBERT H le seguenti proprietà sono equivalenti:*

- (a) $\{U_n : n \in I\}$ converge fortemente a un operatore unitario U ;
- (b) $\{U_n : n \in I\}$ converge debolmente ad un operatore unitario U .

DIMOSTRAZIONE. Poiché la convergenza forte implica quella debole, basta dimostrare l'implicazione (b) \implies (a). Per ogni $f \in H$, si ha

$$\begin{aligned} \|U_n f - U f\| &= \langle U_n f - U f, U_n f - U f \rangle \\ &= \langle U_n f, U_n f \rangle - \langle U_n f, U f \rangle - \langle U f, U_n f \rangle + \langle U f, U f \rangle \\ &= \|f\|^2 - \langle U_n f, U f \rangle - \langle U f, U_n f \rangle + \|f\|^2 \\ &\xrightarrow{n \rightarrow +\infty} 2 \|f\|^2 - 2 \langle U f, U f \rangle = 2 \|f\|^2 - 2 \|f\|^2 = 0 \end{aligned}$$

□

Introduciamo ora la definizione di convergenza che servirà per gli scopi della teoria ergodica.

DEFINIZIONE 11.1. Si dirà che una famiglia $\{T_n : n \in I\}$ di automorfismi converge debolmente a T se $T_n A$ converge a $T A$ per ogni elemento A dell'algebra di misura $\tilde{\mathcal{B}}$, vale a dire se, per ogni insieme $A \in \mathcal{B}$, è

$$\lambda(T_n A \Delta T A) \longrightarrow 0.$$

Una sottobase per gli aperti di questa topologia è data dagli insiemi della forma

$$N(T, A, \varepsilon) := \{S : \lambda(T A \Delta S A) < \varepsilon\}.$$

Consideriamo la famiglia degli intervalli

$$\left\{ \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right] : n \in \mathbf{Z}_+, k = 0, 1, \dots, 2^n - 1 \right\}.$$

Per n fissato in \mathbf{Z}_+ , ognuno di tali intervalli si dirà *di rango n* , mentre si dirà *insieme di rango n* ogni unione, necessariamente finita, di intervalli di rango n . Si dirà *permutazione (diadica di rango n)* una trasformazione invertibile che conserva la misura e che porta ogni intervallo diadico di rango n in un altro intervallo diadico di rango n . La permutazione ciclica di rango n permuta ciclicamente gli intervalli diadici di rango n .

Per dimostrare il Teorema di approssimazione debole occorrerà un risultato preliminare.

La famiglia \mathcal{B}_0 di tutti gli insiemi diadici è densa nell'algebra di misura, è stabile rispetto alle operazioni booleane finite; inoltre la misura di un insieme diadico è un numero razionale diadico. Se E è un insieme di \mathcal{B}_0 e se $r \in [0, \lambda(E)]$ è un numero razionale, allora E ha sottoinsiemi che appartengono a \mathcal{B}_0 e che hanno misura eguale a r . Se E è in \mathcal{B}_0 , anche TE gode della proprietà appena illustrata.

LEMMA 11.1. *Sia $\{E_1, E_2, \dots, E_k\}$ una partizione dell'intervallo unitario; se $\delta > 0$ è arbitrario, e se r_1, r_2, \dots, r_k sono k numeri razionali tali che*

$$|\lambda(E_i) - r_i| < \delta \quad (i = 1, 2, \dots, k),$$

allora esiste una partizione $\{F_1, F_2, \dots, F_k\}$ dell'intervallo unitario in insiemi di \mathcal{B}_0 , $F_i \in \mathcal{B}_0$ per $i = 1, 2, \dots, k$, tali che, per ogni $i = 1, 2, \dots, k$, si abbia

$$\lambda(E_i \Delta F_i) < 2\varepsilon, \quad e \quad \lambda(F_i) = r_i.$$

DIMOSTRAZIONE. Si scelga, per ogni $i = 1, 2, \dots, k$, un insieme $B_i \in \mathcal{B}_0$ tale che

$$\lambda(E_i \Delta B_i) < \gamma,$$

o, equivalentemente

$$|\lambda(E_i) - \lambda(B_i)| = |\lambda(E_i \setminus B_i) - \lambda(B_i \setminus E_i)| \leq \lambda(E_i \Delta B_i) < \gamma.$$

Di conseguenza si ha, per ogni $i = 1, 2, \dots, k$,

$$|\lambda(B_i) - r_i| < \gamma + \delta.$$

Gli insiemi B_i così costruiti non costituiscono necessariamente una partizione dell'intervallo unitario e non hanno necessariamente la misura richiesta. Poiché si ha

$$\lambda[(E_i \cap E_j) \Delta (B_i \cap B_j)] \leq \lambda(E_i \Delta B_i) + \lambda(E_j \Delta B_j) < 2\gamma$$

e

$$\lambda(E_i \cap E_j) = 0,$$

segue che

$$\lambda[B_i \cap (\cup_{j \neq i} B_j)] < 2k\gamma.$$

Si prendano ora gli insiemi di \mathcal{B}_0 definiti, per $i = 1, 2, \dots, k$, da

$$B'_i := B_i \setminus \left(\bigcup_{j \neq i} B_j \right).$$

Per questi insiemi è, per $i = 1, 2, \dots, k$,

$$|\lambda(B'_i) - r_i| < \delta + (2k + 1)\gamma \quad e \quad |\lambda(E_i) - \lambda(B'_i)| < (2k + 1)\gamma.$$

Gli insiemi B'_i sono ora a due a due disgiunti, ma la loro unione non è ancora tutto $[0, 1]$ e la misura di ciascuno di essi non è ancora quella desiderata.

Se $\lambda(B'_i) > r_i$, si consideri un sottoinsieme B''_i di B'_i appartenente a \mathcal{B}_0 e tale che

$$\lambda(B'_i) - \lambda(B''_i) = r_i;$$

si unisca quindi B''_i al complemento degli insiemi B'_i ,

$$C := [0, 1] \setminus \left(\bigcup_{i=1}^k B'_i \right) \in \mathcal{B}_0$$

e si ponga $F_i = B'_i \setminus B''_i$.

Se invece si ha $\lambda(B'_i) < r_i$ si prenda un sottoinsieme $B''_i \in \mathcal{B}_0$ di C tale che

$$\lambda(B'_i \cup B''_i) = r_i$$

e si ponga $F_i := B'_i \cup B''_i$. Gli insiemi F_i ($i = 1, 2, \dots, k$) così costruiti costituiscono una partizione di $[0, 1]$ in insiemi di \mathcal{B}_0 e $\lambda(F_i) = r_i$. Per questi insiemi si ha, per $i = 1, 2, \dots, k$,

$$\lambda(E_i \Delta F_i) < \delta + 2(2k + 1)\gamma.$$

Basta ora prendere γ sufficientemente piccolo perché sia

$$\lambda(E_i \Delta F_i) < 2\delta;$$

ciò conclude la dimostrazione. □

Un *intorno diadico* di un automorfismo S di G ha la forma

$$\{T \in G : \lambda(SD \Delta TD) < \varepsilon\},$$

ove D è un insieme diadico. Gli intorni diadici costituiscono una sottobase per la topologia debole di G . Il risultato fondamentale asserisce che le permutazioni sono dense in G .

TEOREMA 11.2 (di approssimazione debole). *Ogni intorno diadico contiene permutazioni cicliche di rango arbitrario.*

DIMOSTRAZIONE. Dimostriamo dapprima che, se P è una permutazione e se

$$N := \{T : \lambda(PD_i \Delta TD_i) < \varepsilon \quad i = 1, 2, \dots, n\}$$

è un intorno diadico di P , allora N contiene permutazioni cicliche di rango arbitrariamente grande. Ogni insieme PD_i è un insieme diadico. Ne segue che esiste un naturale j tale che PD_i sia di rango j . Naturalmente, j può essere scelto arbitrariamente grande; si scieg ora $k \in \mathbf{N}$ tale che sia

$$k > j \quad \text{e} \quad \frac{2^{j+1}}{2^k} < \varepsilon.$$

Ogni intervallo di rango j è l'unione di 2^{k-j} intervalli di rango k . Definiamo ora una permutazione ciclica Q di rango k che appartenga all'intorno N .

Sia I un intervallo di rango j . Se P non lascia I invariato, si definisca Q in modo che mandi il primo, nell'ordine naturale, sottointervallo di rango k contenuto in I su tutto il primo sottointervallo di rango k di PI . Se P non manda PI su I , si definisca Q in modo che mandi il primo sottointervallo di rango k di PI nel primo sottointervallo di rango k di P^2I . Si continui in questo modo sino a che non si raggiunga l'ultimo elemento nel ciclo indotto da P su I , sia esso $P^{q-1}I$. Quindi $P(P^{q-1}I) = I$. Allora si definisca Q in modo che mandi il primo sottointervallo di rango k di $P^{q-1}I$ nel secondo sottointervallo di rango k di I e si proceda come sopra sostituendo ovunque "primo sottointervallo" con "secondo sottointervallo". Alla fine, si definisca Q in modo che mandi il secondo sottointervallo di rango k di $P^{q-1}I$ nel terzo sottointervallo di rango k di I . Si ripeta di nuovo la procedura sino a quando Q non debba essere definito sull'ultimo sottointervallo di rango k di $P^{q-1}I$.

Sia ora F un intervallo di rango j disgiunto da $I, PI, \dots, P^{q-1}I$ e si definisca Q in modo che applichi l'ultimo sottointervallo di rango k di $P^{q-1}I$ sul primo sottointervallo di rango k di F . Si ripeta, quindi, su F l'intera procedura fatta su I e si proceda in maniera analoga sino a quando non si sia completata la definizione di Q sull'intera decomposizione di P . La definizione di Q sarà completata quando si applicherà l'ultimo sottointervallo di rango k dell'ultimo termine della decomposizione di P nel primo sottointervallo di rango k dell'intervallo originale I .

Evidentemente, Q è una permutazione ciclica di rango k . Per costruzione, se E è un arbitrario intervallo di rango j , allora si ha

$$\lambda(PE\Delta QE) < \frac{2}{2^k}.$$

Ma un insieme di rango j è l'unione disgiunta di al più 2^j intervalli di rango j , sicché se D è un insieme di rango j , allora è

$$\lambda(PD\Delta QD) < 2^j \frac{2}{2^k} = \frac{2^{j+1}}{2^k} < \varepsilon.$$

Perciò Q appartiene a N .

Rimane da dimostrare che le permutazioni sono dense in G . A tal fine, sia T un'arbitraria trasformazione e sia

$$N := \{S : \lambda(SD_i\Delta TD_i) < \varepsilon \quad i = 1, 2, \dots, n\}$$

un intorno diadico di T ; vogliamo dimostrare che N contiene una permutazione. Gli insiemi $D_i \cap TD_j$ costituiscono una partizione dell'intervallo unitario. Sia $\delta > 0$ tale che

$$n^2 \delta < \frac{\varepsilon}{2}.$$

Il Lemma 11.1 assicura l'esistenza di una partizione $\{E_{ij}\}$ dell'intervallo unitario in sottoinsiemi diadici tale che sia

$$\lambda((D_i \cap TD_j) \Delta E_{ij}) < \delta.$$

Una seconda applicazione dello stesso lemma assicura l'esistenza di una partizione $\{F_{ij}\}$ dell'intervallo unitario in sottoinsiemi diadici tale che sia

$$\lambda((D_i \cap TD_j) \Delta TF_{ij}) < \delta$$

e che risulti, per tutti gli indici $\lambda(E_{ij}) = \lambda(F_{ij})$. poiché T conserva la misura, si ha anche

$$\lambda((D_i \cap TD_j) \Delta F_{ij}) < \delta.$$

Sia k un numero naturale tale che tutti gli insiemi D_i , E_{ij} e F_{ij} abbiano rango k e sia P una permutazione di rango k che applichi F_{ij} su E_{ij} . Poiché

$$D_j = \cup_i (T^{-1} D_i \cap D_j),$$

si ha

$$\lambda(D_j \Delta \cup_i F_{ij}) < n^2 \delta < \frac{\varepsilon}{2}.$$

Ma allora

$$\lambda(PD_j \Delta \cup_i E_{ij}) < \frac{\varepsilon}{2}.$$

Pertanto

$$\lambda(PD_j \Delta TD_j) \leq \lambda(PD_j \Delta \cup_i E_{ij}) + \lambda(\cup_i E_{ij} \Delta TD_j) < \varepsilon,$$

ciò che conclude la dimostrazione. □

Vi è una seconda topologia sul gruppo G degli automorfismi, detta *topologia uniforme*.

Si introducono su G due metriche

$$d_1(S, T) := \sup \{ \lambda(SA \Delta TA) : A \in \mathcal{B}(G) \}$$

$$d_2(S, T) := \lambda\{S\omega \neq T\omega\}.$$

È immediato controllare, ed è lasciato come esercizio, che tanto d_1 quanto d_2 sono metriche su G . Inoltre, esse sono invarianti rispetto alle operazioni di gruppo (moltiplicazione a destra o a sinistra, e formazione dell'inversa); ciò è ovvio per d_1 . Per quanto riguarda d_2 , è immediata l'invarianza per la moltiplicazione a sinistra. Poiché si ha

$$S\{S\omega \neq T\omega\} = \{S^{-1}\omega \neq T^{-1}\omega\}$$

segue che

$$\begin{aligned} d_2(S^{-1}, T^{-1}) &= \lambda(\{S^{-1}\omega \neq T^{-1}\omega\}) \\ &= \lambda(S\{S\omega \neq T\omega\}) = \lambda(\{S\omega \neq T\omega\}) = d_2(S, T), \end{aligned}$$

ciò che mostra l'invarianza di d_2 rispetto alla formazione dell'inversa. L'invarianza di d_2 rispetto alla moltiplicazione a destra è ora conseguenza dell'invarianza rispetto alla moltiplicazione a sinistra e di quella rispetto alla formazione dell'inversa.

Se per $x \in [0, 1]$ e per $n \in \mathbf{N}$, si ha $T^n x = x$, si dice che T è *periodica in* x ; il minimo $n = n(x) \in \mathbf{N}$ tale che $T^n x = x$ si dice il *periodo di* x . Se $T^n = I$ (I è l'identità su $[0, 1]$), allora si dice che T è periodica e il minimo $n \in \mathbf{N}$ tale che $T^n = I$ si chiama *periodo di* T . Evidentemente se T è periodica è periodica per ogni x ; il viceversa non è necessariamente vero. Se la trasformazione T non è periodica in alcun punto di $[0, 1]$, si dirà che è *aperiodica*. Sia A_n l'insieme di punti di $[0, 1]$ nei quali T è periodica di periodo $n \in \mathbf{N}$,

$$A_n := \{x \in [0, 1] : T^n x = x, T^j x \neq x (j < n)\}, \quad (11.1)$$

e sia A_0 l'insieme dei punti nei quali T è antiperiodica,

$$A_0 := (\cup_{n \in \mathbf{N}} A_n)^c = \cap_{n \in \mathbf{N}} A_n^c. \quad (11.2)$$

LEMMA 11.2. *Se T è periodica di periodo n per quasi ogni punto di $[0, 1]$, allora esiste un boreliano E di $[0, 1]$ di misura $1/n$ tale che gli insiemi $E, TE, \dots, T^{n-1}E$ siano a due a due disgiunti.*

DIMOSTRAZIONE. Se $n = 1$ non vi è nulla da dimostrare. Se $n > 1$ esiste necessariamente un boreliano E_1 tale che

$$\lambda(E_1 \Delta TE_1) > 0.$$

Infatti, se per ogni boreliano E_1 di $[0, 1]$ fosse $\lambda(E_1 \Delta TE_1) = 0$, T sarebbe di periodo 1 per quasi ogni punto. Poiché T conserva la misura si ha

$$\begin{aligned} \lambda(E_1 \setminus TE_1) &= \lambda(E_1) - \lambda(E_1 \cap TE_1) \\ &= \lambda(TE_1) - \lambda(E_1 \cap TE_1) = \lambda(TE_1 \setminus E_1), \end{aligned}$$

segue che $\lambda(E_1 \setminus TE_1) > 0$. Posto

$$F_1 := E_1 \setminus TE_1,$$

l'insieme F_1 è un boreliano di misura positiva che, per costruzione è disgiunto da $TF_1 = TE_1 \setminus TE_1$. Se $n = 2$, la dimostrazione è terminata. Se $n > 2$, esiste necessariamente un sottoinsieme E_2 di F_1 che è misurabile e tale che $\lambda(E_2 \Delta T^2 E_2) > 0$; se così non fosse, T sarebbe di periodo 2 quasi ovunque in F_1 . Posto $F_2 := E_2 \setminus T^2 E_2$, scende come sopra che F_2 e $T^2 F_2$ sono disgiunti. Procedendo in questo modo per induzione, dopo $n - 1$ passi si sono costruiti gli insiemi F_1, F_2, \dots, F_{n-1} di misura positiva tali

che $F_1 \supset F_2 \supset \dots \supset F_{n-1}$ e che, per ogni $j = 1, 2, \dots, n-1$ gli insiemi F_j e $T^j F_j$ siano disgiunti. Posto $E := F_{n-1}$, l'insieme E risulta essere disgiunto da $T^j E$ per ogni $j = 1, 2, \dots, n-1$. La misura di E potrebbe non essere eguale a $1/n$; si appliche allora lo stesso procedimento ora usato al sottoinsieme invariante

$$(E \cup TE \cup \dots \cup T^{n-1}E)^c.$$

Si proceda per induzione terminando al piú dopo infinità numerabile di passi. \square

LEMMA 11.3. *Se T è antiperiodica, allora per ogni $n \in \mathbf{N}$ e per ogni $\varepsilon > 0$, esiste un boreliano E tale che gli insiemi $E, TE, \dots, T^{n-1}E$ siano disgiunti e che sia*

$$\lambda(E \cup TE \cup \dots \cup T^{n-1}E) > 1 - \varepsilon.$$

DIMOSTRAZIONE. Sia p un numero naturale tale che $1/p < \varepsilon$. Ricorrendo allo stesso ragionamento del Lemma 11.2 nel quale si sostituisce pn a n , si costruisce un boreliano F di misura positiva e tale che gli insiemi $F, TF, \dots, T^{pn-1}F$ siano a due a due disgiunti. Inoltre F è massimale nel senso che non esiste alcun boreliano che includa F , abbia misura maggiore di quest'ultimo ed abbia le proprietà elencate per F . Perciò, se F_0 è un sottoinsieme misurabile di $T^{pn-1}F$ con $\lambda(F_0) > 0$, allora esiste almeno un numero naturale $j = 1, 2, \dots, pn$ tale che $\lambda(T^j F_0 \cap F) > 0$, altrimenti l'insieme $F \cup TF_0$ contraddirebbe la massimalità di F . Si definisca ora, per $j = 1, 2, \dots, pn$, il boreliano

$$F_j := (T^{pn-1}F \cap T^{-j}F) \cap \left(\bigcup_{i=1}^{j-1} T^{-i}F^c \right).$$

Allora si ha

$$\lambda(T^{pn-1}F \setminus \bigcup_{j=1}^{pn} F_j) = 0.$$

Gli insiemi di una stessa colonna dello schema

$$\begin{array}{cccc} TF_2 & & & \\ TF_3 & T^2 F_3 & & \\ \vdots & \vdots & & \\ TF_{pn} & T^2 TF_{pn} & \dots & T^{pn-1} TF_{pn} \end{array} \quad (11.3)$$

sono a due a due disgiunti perché si ottengono trasformando mediante la stessa potenza di T insiemi disgiunti; inoltre, due insiemi di colonne differenti sono disgiunti perché le loro immagini inverse mediante opportune potenze di T sono incluse in insiemi distinti di $F, TF, \dots, T^{pn-1}F$.

Gli insiemi $T^i F_j$ ($1 \leq i < j \leq pn$) sono disgiunti da ognuno degli insiemi $T^k F$ ($k = 1, 2, \dots, pn-1$) e quindi dalla loro unione. Infatti, se è $k > i$, allora

$$T^i F_j \cap T^k F \subset T^i (T^{pn-1}F \cap T^{k-i}F) = \emptyset;$$

se, invece, $k \leq i$, allora si avrebbe $0 \leq i - k < j$, sicché, per la definizione di F_j , è vuoto l'insieme

$$T^{i-k} F_j \cap F;$$

perciò è vuoto anche l'insieme

$$T^i F_j \cap T^k F = T^k (T^{i-k} F_j \cap F).$$

Pertanto gli insiemi $TF_1, T^2 F_2, \dots, T^{pn} F_{pn}$ sono sottoinsiemi a due a due disgiunti di F ; quindi

$$\lambda \left(\bigcup_{j=1}^{pn} T^j F_j \right) = \sum_{j=1}^{pn} \lambda (T^j F_j) = \sum_{j=1}^{pn} \lambda (F_j) = \lambda (T^{pn-1} F).$$

Si può ora asserire che l'insieme

$$F^* := \left(\bigcup_{k=0}^{pn-1} T^k F \right) \bigcup \left(\bigcup_{1 \leq i < j \leq pn} T^i F_j \right)$$

è quasi invariante, nel senso che $\lambda (F^* \Delta T^{-1} F^*) = 0$. Siccome T è antiperiodica, si ha

$$\lambda (F^*) = 1,$$

perché altrimenti l'insieme $F \cup (F^*)^c$ permetterebbe di ingrandire l'insieme massimale F . Si ponga, infine

$$E := \left(\bigcup_{k=0}^{p-1} T^{kn} F \right) \bigcup \left(\bigcup_{i=1}^{pn-1} \bigcup_{j=i+n}^{pn} T^i F_j \right).$$

Per costruzione gli insiemi $E, TE, \dots, T^{n-1} E$ sono disgiunti. L'unione di questi insiemi non contiene al più $n - 1$ insiemi di ogni riga dello schema (11.3). Ciò significa che la misura totale degli insiemi della j -esima riga non inclusi nell'unione $E \cup TE \cup \dots \cup T^{n-1} E$ ha misura minore di $n \lambda (F_j)$; di conseguenza

$$\lambda (E \cup TE \cup \dots \cup T^{n-1} E) > 1 - \sum_{j=1}^{pn} n \lambda (F_j) = 1 - n \lambda (F).$$

Gli insiemi $F, TF, \dots, T^{pn-1} F$ sono disgiunti ed hanno tutti la stessa misura sicché

$$n \lambda (F) \leq n \frac{1}{pn} = \frac{1}{p} < \varepsilon.$$

Ciò conclude la dimostrazione □

Le topologie indotte su G dalle metriche d_1 e d_2 sono equivalenti; piú precisamente vale il seguente

LEMMA 11.4. *Valgono le diseguaglianze*

$$\frac{2}{3} d_2 \leq d_1 \leq d_2.$$

DIMOSTRAZIONE. Poiché tanto d_1 quanto d_2 sono invarianti rispetto alle operazioni di gruppo, basta dimostrare che

$$\frac{2}{3} d_2(I, T) \leq d_1(I, T) \leq d_2(I, T)$$

per ogni trasformazione T che conservi la misura. Se

$$F := \{x \in [0, 1] : Tx \neq x\},$$

allora si ha $d_2(I, T) = \lambda(F)$. Si osservi che l'insieme F è invariante rispetto a T e che tale è anche ogni sottoinsieme misurabile di F^c . Per ogni boreliano E di $[0, 1]$, si ha

$$\begin{aligned} E\Delta TE &= [(E\Delta TE) \cap F] \cup [(E\Delta TE) \setminus F] \\ &= [(E \cap F) \Delta (TE \cap F)] \cup [(E \setminus F) \Delta (TE \setminus F)] \\ &= [(E \cap F) \Delta (TE \cap TF)] \cup [(E \setminus F) \Delta (TE \setminus TF)] \\ &= [(E \cap F) \Delta T(E \cap F)] \cup [(E \setminus F) \Delta T(E \setminus F)]; \end{aligned}$$

di qui, dato che $T(E \setminus F) = E \setminus F$,

$$\begin{aligned} \lambda(E\Delta TE) &= \lambda[(E \cap F) \Delta T(E \cap F)] + \lambda[(E \setminus F) \Delta T(E \setminus F)] \\ &= \lambda[(E \cap F) \Delta (TE \cap F)] = \lambda[(E\Delta TE) \cap F] \leq \lambda(F). \end{aligned}$$

Ne segue che $d_1(I, T) \leq d_2(I, T)$.

Si consideri ora la partizione di $[0, 1]$ nei sottoinsiemi

$$\{A_0, A_1, \dots, A_n, \dots\}$$

delle equazioni (11.1) e (11.2). Si applichi il lemma 11.3 all'insieme A_0 con $n = 2$ e $\varepsilon \leq 1/3$. Esiste perciò un boreliano E_0 di A_0 tale che $E_0 \cap TE_0 = \emptyset$ e che

$$\lambda(E_0) \geq \frac{1}{3} \lambda(A_0).$$

Applicando il Lemma 11.2 ad ogni A_n si ha un boreliano E_n di A_n tale che gli insiemi $E_n, TE_n, \dots, T^{n-1}E_n$ siano a due a due disgiunti e che sia

$$\lambda(E_n) = \frac{1}{n} \lambda(A_n).$$

Si definisca ora una successione di insiemi misurabili $\{B_n : n = 0 \text{ e } n \geq 2\}$ come segue. Si ponga

- per $n = 0$, $B_0 := E_0$;
- per $n \geq 2$ pari,

$$B_n := E_n \cup \left(\bigcup_{k=1}^{n/2-1} T^{2k} E_n \right);$$

— per $n \geq 2$ dispari,

$$B_n := E_n \cup \left(\bigcup_{k=1}^{(n-3)/2} T^{2k} E_n \right).$$

Si ha, in ogni caso, $B_n \cap TB_n = \emptyset$. Se $n \geq 2$ è pari, si ha

$$\lambda(B_n) = \frac{1}{2} \lambda(A_n),$$

mentre, se $n \geq 2$ è dispari, si ha

$$\lambda(B_n) = \frac{1}{2} \left(1 - \frac{1}{n} \right) \lambda(A_n).$$

Ne segue che, per $n \geq 2$, è

$$\lambda(B_n) \geq \frac{1}{3} \lambda(A_n).$$

Posto

$$B := F_0 \cup B_2 \cup B_3 \cup \dots \cup B_n \cup \dots,$$

si ha $B \cap TB = \emptyset$ e

$$\lambda(B) \geq \frac{1}{3} (1 - \lambda(A_1)).$$

Quindi

$$d_1(I, T) \geq \lambda(B \Delta TB) \geq \frac{2}{3} (1 - \lambda(A_1)) = \frac{2}{3} d_2(I, T),$$

ciò che conclude la dimostrazione. \square

Per ottenere il teorema di approssimazione per la topologia uniforme avremo bisogno del seguente risultato che interessa indipendente.

TEOREMA 11.3. *Per ogni coppia A e B di boreliani dell'intervallo $[0, 1]$ esiste una trasformazione $T : [0, 1] \rightarrow [0, 1]$ tale che $\lambda(TA \Delta B) = 0$.*

DIMOSTRAZIONE. Se i due boreliani hanno misura nulla, $\lambda(A) = \lambda(B) = 0$, non vi è nulla da dimostrare. Supporremo, pertanto, che sia $\lambda(A) \lambda(B) > 0$. Sia S una trasformazione ergodica su $[0, 1]$; esiste allora un naturale n tale che

$$\lambda(S^n A \cap B) > 0.$$

Si definisca $Tx := S^n x$ se x appartiene a $A \cap T^{-n}B$. Si proceda ora per induzione.

Se $\lambda(A \setminus S^{-n}B) > 0$, si applichi il ragionamento appena usato per trovare un sottoinsieme misurabile C di $A \setminus S^{-n}B$ di misura non nulla, $\lambda(C) > 0$ ed un naturale k tale che

$$\lambda[S^k C \cap (B \setminus S^{-n}A)] > 0;$$

e si definisca $Tx := S^k x$ se x appartiene a C . Procedendo in questo modo, eventualmente per induzione transfinita, si giunge a definire la trasformazione T su tutto A . Applicando poi lo stesso metodo ai due insiemi A^c e B^c , si definisce T su tutto $[0, 1]$. \square

Il seguente teorema asserisce di fatto che la trasformazioni periodiche sono dense in G nella topologia uniforme.

TEOREMA 11.4 (d'approssimazione uniforme). *Per ogni trasformazione aperiodica T , per ogni numero naturale n e per ogni $\varepsilon > 0$ esiste una trasformazione S di periodo n tale che*

$$d_2(S, T) \leq \frac{1}{n} + \varepsilon.$$

DIMOSTRAZIONE. In virtù del Lemma 11.3 esiste un boreliano E tale che gli insiemi $E, TE, \dots, T^{n-1}E$ siano a due a due disgiunti e che

$$\lambda(E \cup TE \cup \dots \cup T^{n-1}E) > 1 - \varepsilon.$$

Se x è in $E \cup TE \cup \dots \cup T^{n-2}E$, si definisca $Sx := Tx$; se x appartiene a $T^{n-1}E$, si ponga $Sx := T^{-n+1}x$. La trasformazione S è così stata definita in $E \cup TE \cup \dots \cup T^{n-1}E$ ed è qui di periodo n ; quale che sia il modo di estendere la definizione di S a tutto $[0, 1]$, si ha

$$d_2(S, T) \leq \lambda(T^{n-1}E) + \varepsilon = \lambda(E) + \varepsilon \leq \frac{1}{n} + \varepsilon.$$

Rimane quindi solo da dimostrare che si può definire S nel complementare di $E \cup TE \cup \dots \cup T^{n-1}E$ in modo che anche qui S sia di periodo n e questo segue facilmente dal Teorema 11.3. \square

12. Categoria

L'introduzione della topologia debole nell'insieme delle trasformazioni che conservano la misura consente di trarre conclusioni importanti sulla grandezza topologica di alcuni sottinsiemi di trasformazioni.

TEOREMA 12.1. *Nella topologia debole le trasformazioni fortemente mescolanti costituiscono un insieme di prima categoria.*

DIMOSTRAZIONE. Per $k \in \mathbb{N}$ sia \mathcal{P}_k l'insieme delle trasformazioni che conservano la misura e che sono periodiche di periodo k ,

$$\mathcal{P}_k := \{T \in G : T^k = I\};$$

si ponga inoltre

$$\mathcal{P}_n^* := \cup_{k>n} \mathcal{P}_k.$$

In virtù del Teorema di approssimazione debole, l'insieme \mathcal{P}_n^* è denso per ogni $n \in \mathbf{N}$.

Posto $E := [0, 1/2]$ si consideri la seguente famiglia di trasformazioni che conservano la misura

$$M_k := \left\{ T \in G : \left| \lambda(E \cap T^k E) - \frac{1}{4} \right| \leq \frac{1}{5} \right\}.$$

Tale insieme è chiuso nella topologia debole. Si definiscano

$$M_n^* := \bigcap_{k>n} M_k \quad \text{e} \quad M^* := \bigcup_{n \in \mathbf{N}} \bigcap_{k>n} M_k = \liminf_{n \rightarrow +\infty} M_k.$$

È evidente che M^* contiene tutte le trasformazioni fortemente mescolanti; infatti, per queste si ha

$$\lim_{n \rightarrow +\infty} \lambda(E \cap T^k E) = \frac{1}{4}.$$

Basta, pertanto, dimostrare che M^* è di prima categoria; a tal fine basta dimostrare che, per ogni $n \in \mathbf{N}$, M_n^* è e, siccome M_n^* è chiuso, basta in effetti mostrare che il suo complementare in G , $G \setminus M_n^*$ è denso in G . Ma

$$G \setminus M_n^* = G \cap (\bigcap_{k>n} M_k)^c = \bigcup_{k>n} G \setminus M_k,$$

sicché basta dimostrare che vale l'inclusione

$$\mathcal{P}_k \subset G \setminus M_k.$$

Questo è però di facile dimostrazione, perché se T appartiene a \mathcal{P}_k si ha $T^k E = E$, sicché

$$\lambda(E \cap T^k E) - \frac{1}{4} = \lambda(E) - \frac{1}{4} = \frac{1}{4};$$

dunque T non appartiene a M_k . □

Il lemma che segue sarà necessario per studiare l'insieme delle trasformazioni debolmente mescolanti.

LEMMA 12.1. *Nella topologia debole la classe coniugata di ogni trasformazione antiperiodica T_0 ,*

$$\{S^{-1}T_0S : S \in G\},$$

è densa in G .

DIMOSTRAZIONE. Sia $N = \{T \in G : \lambda(PD_i \Delta T D_i) < \varepsilon, i = 1, 2, \dots, n\}$ un intorno diadico di una permutazione P ; si vuole dimostrare che esiste una trasformazione S di G tale $S^{-1}T_0S$ appartenga a N . Posto

$$M = \left\{ T \in G : \lambda(PD_i \Delta T D_i) < \frac{\varepsilon}{2}, i = 1, 2, \dots, n \right\},$$

segue dal Teorema di approssimazione debole che l'intorno M contiene una permutazione ciclica Q di rango k maggiore dei ranghi di tutti gli insiemi D_i ($i = 1, 2, \dots, n$) e tale che

$$\frac{1}{2^{k-2}} < \varepsilon.$$

Il Teorema di approssimazione uniforme assicura ora l'esistenza di una trasformazione $R \in G$ che ha ovunque periodo 2^k e tale che

$$d_2(R, T_0) < \frac{2}{2^k} < \frac{\varepsilon}{2}. \quad (12.1)$$

Le trasformazioni Q e R sono coniugate in G . Si numerino gli intervalli diadici di rango k , E_0, E_1, \dots, E_q ($q = 2^k$) in modo tale che sia $QE_i = E_{i+1}$, con gli indici presi modulo q . Sia ora F_0 un boreliano tale che $\lambda(F_0) = 1/q$ e che gli insiemi $F_0, RF_0, \dots, R^{q-1}F_0$ siano a due a due disgiunti. Posto $F_i := R^i F_0$ ($i = 1, 2, \dots, q-1$), sia S una qualsiasi trasformazione che conserva la misura e che manda E_0 su F_0 . Si può estendere la definizione di tale S a tutto l'intervallo unitario in maniera ricorsiva, ponendo, per $x \in E_i$ ($i = 1, 2, \dots, q-1$),

$$Sx := R^i S Q^{-i} x.$$

Schematicamente si ha

$$\begin{array}{cccccccc} E_0 & \xrightarrow{Q} & E_1 & \xrightarrow{Q} & E_2 & \xrightarrow{Q} & \dots & \xrightarrow{Q} & E_{q-2} & \xrightarrow{Q} & E_{q-1} \\ s \downarrow & & s \downarrow \\ F_0 & \xrightarrow{R} & F_1 & \xrightarrow{R} & F_2 & \xrightarrow{R} & \dots & \xrightarrow{R} & F_{q-2} & \xrightarrow{R} & F_{q-1} \end{array}$$

Si vede subito che $Q = S^{-1}RS$.

Poiché la distanza d_2 è invariante per le operazioni di gruppo in G , la (12.1) dà

$$d_2(Q, S^{-1}T_0S) = d_2(S^{-1}RS, S^{-1}T_0S) = d_2(R, T_0) < \frac{\varepsilon}{2}.$$

Pertanto

$$\begin{aligned} \lambda(PD_i \Delta S^{-1}T_0SD_i) &\leq \lambda(PD_i \Delta QD_i) + \lambda(QD_i \Delta S^{-1}T_0SD_i) \\ &\leq \lambda(PD_i \Delta QD_i) + d(Q, S^{-1}T_0S) \\ &< \frac{\varepsilon}{2} + d_2(Q, S^{-1}T_0S) < \varepsilon. \end{aligned}$$

Perciò $Q, S^{-1}T_0S$ appartiene a N . □

13. Trasformazioni indotte

DEFINIZIONE 13.1. Siano $T\Omega \rightarrow \Omega$ una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e A un sottoinsieme misurabile di Ω .

Sia $\omega \in A$ un punto ricorrente (quasi tutti i punti di A per il teorema di ricorrenza di POINCARÉ sono ricorrenti, ritornano, cioè, in A) e si indichi con $n_A(\omega)$ il minimo naturale positivo tale che $T^{n_A(\omega)}\omega$ sia in A ,

$$n_A(\omega) := \{n \in \mathbf{N} : T^n\omega \in A\}.$$

Si dirà *trasformazione indotta da T su A* la trasformazione $T_A : A \rightarrow A$ definita per quasi tutti i punti di A da

$$T_A\omega := T^{n_A(\omega)}\omega.$$

Si noti che, a rigore, T_A non è definita su tutti i punti di A , ma solo su quelli ricorrenti; sui rimanenti punti T_A può essere definita in maniera arbitraria.

Si considerino la traccia della tribú \mathcal{F} su A ,

$$\mathcal{F}_A := \mathcal{F} \cap A = \{A \cap B : B \in \mathcal{F}\}$$

e la probabilità μ_A condizionata da A ,

$$\mu_A(B) = \frac{\mu(A \cap B)}{\mu(A)}.$$

TEOREMA 13.1. *La trasformazione indotta T_A è misurabile rispetto alla tribú traccia \mathcal{F}_A e conserva la misura sullo spazio di probabilità $(A, \mathcal{F}_A, \mu_A)$.*

DIMOSTRAZIONE. Per ogni $n \in \mathbf{N}$ si definiscano gli insiemi

$$\begin{aligned} A_n &:= \{\omega \in A : T\omega \notin A, T^2\omega \notin A, \dots, T^{n-1}\omega \notin A, T^n\omega \in A\} \\ &= \{\omega \in A : n_A(\omega) = n\} \end{aligned}$$

e

$$B_n := \{\omega \in A^c : T\omega \notin A, T^2\omega \notin A, \dots, T^{n-1}\omega \notin A, T^n\omega \in A\}.$$

Ora si ha

$$A_n = (A \cap T^{-n}A) \setminus \bigcup_{j=1}^{n-1} (A \cap T^{-j}A) \quad \text{e} \quad B_n = T^n A \setminus \bigcup_{j=0}^{n-1} T^{-j}A,$$

sicché la misurabilità di T assicura che tanto A_n quanto B_n siano misurabili, per ogni $n \in \mathbf{N}$. Inoltre, per definizione, si ha $A_n \subset A$ ($n \in \mathbf{N}$); infine, se B è un sottoinsieme misurabile di A , $B \in \mathcal{F}_A$, è

$$T_A^{-1}B = \bigcup_{n \in \mathbf{N}} (A_n \cap T^{-n}B), \quad (13.1)$$

come subito si verifica. Questo mostra che T_A è misurabile rispetto a \mathcal{F}_A .

Gli insiemi della successione $\{A_n : n \in \mathbf{N}\}$ sono disgiunti; inoltre

$$\mu(A \setminus \bigcup_{n \in \mathbf{N}} A_n) = 0.$$

D'altro canto si ha

$$T^{-1}A = A_1 \cup B_1$$

e, per ogni $n \in \mathbf{N}$,

$$T^{-1} B_n = A_{n+1} \cup B_{n+1}.$$

Pertanto, se B appartiene a \mathcal{F}_A

$$T^{-1} B = T^{-1} (B \cap A) = (A_1 \cap T^{-1} B) \cup (B_1 \cap T^{-1} B) \quad (13.2)$$

$$T^{-1} (B_n \cap T^{-n} B) = (A_{n+1} \cap T^{-(n+1)} B) \cup (B_{n+1} \cap T^{-(n+1)} B). \quad (13.3)$$

Poiché T conserva la misura su $(\Omega, \mathcal{F}, \mu)$, per ogni $n \in \mathbf{N}$, si ha dalla ultime relazioni applicate ripetutamente

$$\begin{aligned} \mu(B) &= \mu(T^{-1} B) = \mu(A_1 \cap T^{-1} B) + \mu(B_1 \cap T^{-1} B) \\ &= \mu(A_1 \cap T^{-1} B) + \mu(A_2 \cap T^{-2} B) + \mu(B_2 \cap T^{-2} B) = \dots \\ &= \sum_{j=1}^n \mu(A_j \cap T^{-j} B) + \mu(B_n \cap T^{-n} B). \end{aligned}$$

La successione $\{B_n : n \in \mathbf{N}\}$ è formata da insiemi disgiunti e, di conseguenza, ha come limite \emptyset ; perciò

$$\mu(B) = \sum_{n \in \mathbf{N}} \mu(A_n \cap T^{-n} B).$$

Ma per la (13.1) quest'ultima somma è eguale a $\mu(T_A^{-1} B)$, onde $\mu(B) = \mu(T_A^{-1} B)$. \square

TEOREMA 13.2. *Se T è invertibile, anche la trasformazione indotta T_A è invertibile.*

DIMOSTRAZIONE. Poiché T^{-1} conserva la misura, si può definire la trasformazione indotta $(T^{-1})_A$ su A da T^{-1} . Vogliamo mostrare che $T_A^{-1} = (T^{-1})_A$. A tal fine basta far vedere che si ha $(T^{-1})_A T_A \omega = \omega$ per quasi tutti i punti ω di A . Se ω è in A_n allora $T_A \omega = T^n \omega =: y \in A$. Chiaramente $T^{-n} y = \omega \in A$. Si supponga che sia $z = T^{-m} y \in A$ per $m \in \mathbf{N}$ con $1 \leq m < n$. Allora $T^m z = y = T^m (T^{n-m} \omega)$. Ma $T^{n-m} \omega$ non può appartenere a A , ciò che contraddice all'ipotesi che T sia iniettiva. Ne segue che $(T^{-1})_A y = T^{-n} y = \omega$. \square

TEOREMA 13.3. *Se T è ergodica, tale è anche T_A .*

DIMOSTRAZIONE. Sia B un insieme invariante di \mathcal{F}_A , $T^{-1} B = B$. Si ponga

$$D := \cup_{n \in \mathbf{N}} \{(A_n \cap T^{-n} B) \cup (B_n \cap T^{-n} B)\}.$$

poiché $T^{-1}B = B$, la (13.1) dà $B \supset A_n \cap T^{-n}B$ ($n \in \mathbf{N}$) e, quindi, per ogni $n \in \mathbf{N}$, si ha

$$A \setminus B = A \cap B^c \subset A \cap (A_n^c \cup T^{-n}B^c) = (A \cap A_n^c) \cup (A \cap T^{-n}B^c)$$

e, dunque,

$$\begin{aligned} A \setminus B &\subset \bigcap_{n \in \mathbf{N}} (A \cap A_n^c) \cup (A \cap T^{-n}B^c) \\ &= \left\{ \bigcap_{n \in \mathbf{N}} (A \cap A_n^c) \right\} \cup \left\{ \bigcap_{n \in \mathbf{N}} (A \cap T^{-n}B^c) \right\} \\ &= (A \cap A^c) \cup \left\{ \bigcap_{n \in \mathbf{N}} (A \cap T^{-n}B^c) \right\} = \bigcap_{n \in \mathbf{N}} (A \cap T^{-n}B^c). \end{aligned}$$

D'altro canto

$$D^c = \bigcap_{n \in \mathbf{N}} \left\{ (A_n \cup B_n) \cap T^{-n}B \right\}^c = \bigcap_{n \in \mathbf{N}} \left\{ (A_n \cup B_n)^c \cup T^{-n}B^c \right\},$$

sicché $A \setminus B$ è contenuto in D^c . Ora, la (13.2) e la (13.3) danno

$$T^{-1}(A_n \cap T^{-n}B) \subset T^{-1}B = (A_1 \cap T^{-1}B) \cup (B_1 \cap T^{-1}B)$$

e

$$T^{-1}(B_n \cap T^{-n}B) = (A_{n+1} \cap T^{-(n+1)}B) \cup (B_{n+1} \cap T^{-(n+1)}B),$$

sicché $T^{-1}D = D$, cioè D è invariante. Poiché T è ergodica, si ha o $\mu(D) = 0$ o $\mu(D^c) = 0$. Nel primo caso è

$$\mu(D) \geq \sum_{n \in \mathbf{N}} p(A_n \cap T^{-n}B) = \mu(T_A^{-1}B),$$

e, quindi, $\mu(B) = \mu(T_A^{-1}B) = 0$. Se, invece, $\mu(D^c) = 0$, allora

$$\mu(D^c) \geq \mu(A \setminus B) = \mu_A(B^c) \mu(A),$$

vale a dire $\mu_A(B^c) = 0$. □

Sia dato uno spazio misurale $(\Omega, \mathcal{F}, \mu)$ finito o σ -finito. e una successione $\{B_n : n \in \mathbf{Z}_+\}$ di insiemi misurabili e disgiunti tali che $\mu(B_{n+1}) \leq \mu(B_n) < +\infty$. Si supponga che, per ogni $n \in \mathbf{Z}_+$, esista una trasformazione $\sigma_n : B_{n+1} \rightarrow B_n$ invertibile e che conserva la misura tra B_{n+1} e $\sigma_n(B_{n+1})$. Sia, infine, $T_0 : B_0 \rightarrow B_0$ una trasformazione suriettiva e invertibile che conserva la misura. Posto

$$\Omega' := \bigcup_{n \in \mathbf{Z}_+} B_n$$

si definisca un'applicazione $T : \Omega' \rightarrow \Omega'$ come segue

$$T\omega := \begin{cases} \sigma_n^{-1}\omega, & \omega \in \sigma_n(B_{n+1}), \\ T_0 \sigma_0 \dots \sigma_{n-1} \omega, & \omega \in B_n \setminus \sigma_n(B_{n+1}). \end{cases} \quad (13.4)$$

TEOREMA 13.4. *La trasformazione $T : \Omega' \rightarrow \Omega'$ definita dalla (13.4) è invertibile e conserva la misura. Se, poi, è $\mu(\Omega') = 1$, allora T_0 è la trasformazione indotta da T su B_0 .*

DIMOSTRAZIONE. Si supponga che sia $T\omega_1 = T\omega_2$. Vari casi sono possibili.

Caso 1: ω_1 e ω_2 appartengono a $\sigma(B_{n+1})$ per lo stesso indice n . Allora, $T\omega_1 = T\omega_2$ significa $\sigma_n^{-1}\omega_1 = \sigma_n^{-1}\omega_2$, e, quindi, in virtù dell'invertibilità di σ_n , $\omega_1 = \omega_2$.

Caso 2: ω_1 e ω_2 appartengono a $B_n \setminus \sigma(B_{n+1})$ per lo stesso indice n . In questo caso si ha

$$T\omega_1 = T_0\sigma_0 \dots \sigma_{n-1}\omega_1 = T_0\sigma_0 \dots \sigma_{n-1}\omega_2 = T\omega_2,$$

onde $\omega_1 = \omega_2$, perché le trasformazioni $T_0, \sigma_0, \dots, \sigma_{n-1}$ sono tutte invertibili.

Caso 3: $\omega_1 \in \sigma_n(B_{n+1})$ e $\omega_2 \in B_j \setminus \sigma_j(B_{j+1})$; ma, allora, $T\omega_1$ appartiene a B_{n+1} mentre $T\omega_2$ è in B_0 , di modo che non si può avere l'eguaglianza $T\omega_1 = T\omega_2$. Dunque questo caso non si può presentare.

Caso 4: è palesemente impossibile che sia $\omega_1 \in \sigma_i(B_{i+1})$ e $\omega_2 \in \sigma_j(B_{j+1})$ con $i \neq j$.

Caso 5: infine, si potrebbe, *a priori*, avere $\omega_1 \in B_i \setminus \sigma_i(B_{i+1})$ e $\omega_2 \in B_j \setminus \sigma_j(B_{j+1})$ con $i \neq j$. Si supponga, senza perdita di generalità, che sia $i < j$. Dall'eguaglianza $T\omega_1 = T\omega_2$ scende, a causa dell'invertibilità di tutte le trasformazioni in gioco, che $\omega_1 = \sigma_i\sigma_{i+1} \dots \sigma_{j-1}\omega_2$ che appartiene a $\sigma_i(B_{i+1})$ contrariamente all'ipotesi; dunque anche questo caso non si può presentare.

Si è così stabilito che T , definita dalla (13.4) è iniettiva. Sia ora y un punto di Ω' e si supponga che esso appartenga a B_n con $n \in \mathbf{N}$. Poiché σ_{n-1} è, per ipotesi, invertibile, esiste sicuramente un punto z di $\sigma_{n-1}(B_n)$ tale che sia $y = \sigma_{n-1}^{-1}z$, o, equivalentemente $y = Tz$. Se, poi, y appartiene a B_0 , esiste, in virtù dell'invertibilità di T_0 , $y' \in B_0$ tale che $y = T_0y'$. Ora, se y' è in $B_0 \setminus \sigma_0(B_1)$ basta prendere $z = y'$ per avere $y = T_0y' = Tz$. Se, invece y' è in $\sigma_0(B_1)$, esiste $y_1 \in B_1$ tale che $\sigma_1y_1 = y'$; se y_1 è in $B_1 \setminus \sigma_1(B_2)$, si prenda $z = y_1$ e si avrà $Tz = T_0\sigma_0y_1 = y$. procedendo per induzione, se y_1 appartiene a $\sigma_1(B_2)$, esiste $y_2 \in B_2$ tale che $\sigma_1y_2 = y_1$. Se per un indice $n \in \mathbf{N}$ si ha $y_n \in B_n \setminus \sigma_n(B_{n+1})$, si prenda $z = y_n$ e si avrà così $Tz = T_0\sigma_0 \dots \sigma_{n-1}y_n = y$. Si supponga, quindi, che y_n appartenga a $\sigma_n(B_{n+1})$ per ogni $n \in \mathbf{N}$. poiché tutte le trasformazioni T_0, σ_0, \dots ,

σ_n, \dots conservano la misura e poiché $\mu(B_n) \xrightarrow{n \rightarrow +\infty} 0$, risulta che i punti y che danno origine a una tale successione costituiscono un insieme di misura nulla.

Per mostrare che T conserva la misura sia A un sottoinsieme misurabile di Ω' e si scriva

$$A = \bigcup_{n \in \mathbf{Z}_+} \{[A \cap \sigma_n(B_{n+1})] \cup [A \cap (B_n \setminus \sigma_n(B_{n+1}))]\}$$

che è un'unione disgiunta. Ora

$$\begin{aligned} T^{-1}A &= \bigcup_{n \in \mathbf{Z}_+} \{\sigma_n[A \cap \sigma_n(B_{n+1})] \cup \sigma_{n-1}^{-1} \dots \sigma_0^{-1} T_0^{-1}[A \cap (B_n \setminus \sigma_n(B_{n+1}))]\} \end{aligned}$$

che mostra, intanto, che $T^{-1}A$ è misurabile. Inoltre

$$\begin{aligned} \mu(T^{-1}A) &= \sum_{n \in \mathbf{Z}_+} \mu(\sigma_n[A \cap \sigma_n(B_{n+1})]) \\ &\quad + \sum_{n \in \mathbf{Z}_+} \mu(\sigma_{n-1}^{-1} \sigma_{n-2}^{-1} \dots \sigma_0^{-1} T_0^{-1}[A \cap (B_n \setminus \sigma_n(B_{n+1}))]) \\ &= \sum_{n \in \mathbf{Z}_+} \mu(A \cap \sigma_n(B_{n+1})) + \sum_{n \in \mathbf{Z}_+} \mu(A \cap (B_n \setminus \sigma_n(B_{n+1}))) \\ &= \mu(A). \end{aligned}$$

Resta da mostrare che $T_0 = T_{B_0}$. Per definizione,

$$T_{B_0}(\omega) = T^{n_{B_0}(\omega)}(\omega).$$

Se $\omega \in B_0 \setminus \sigma_0(B_1)$, è $n_{B_0}(\omega) = 1$ e, dunque, $T_{B_0}(\omega) = T(\omega) = T_0(\omega)$. Se, invece, $\omega \in \sigma_0(B_1)$, si ha $T(\omega) \in B_1$; se, ora, $T(\omega)$ appartiene a $B_1 \setminus \sigma_1(B_2)$, è $n_{B_0}(\omega) = 2$ e

$$T_{B_0}(\omega) = T^2(\omega) = T(T(\omega)) = T(\sigma_0^{-1}(\omega)) = T_0(\sigma_0 \sigma_0^{-1}(\omega)) = T_0(\omega).$$

Si supponga che $T^k(\omega)$ sia in $B_k \setminus \sigma_k(B_{k+1})$, con $T^j(\omega) \in \sigma_j(B_{j+1})$ ($j = 0, 1, \dots, k-1$). Allora è $n_{B_0}(\omega) = k+1$ e

$$\begin{aligned} T_{B_0}(\omega) &= T^{k+1}(\omega) = T(T^k(\omega)) = T_0 \sigma_0 \dots \sigma_{k-1}(T^k(\omega)) \\ &= T_0 \sigma_0 \dots \sigma_{k-2} \sigma_{k-1} [T(T^{k-1}(\omega))] \\ &= T_0 \sigma_0 \dots \sigma_{k-2} (\sigma_{k-1} \sigma_{k-1}^{-1})(T^{k-1}(\omega)) \\ &= \dots = T_0(\sigma_0 \sigma_0^{-1}(\omega)) = T_0(\omega). \end{aligned}$$

I punti $\omega \in B_0$ tali che $T^n(\omega)$ appartenga a $\sigma_n(B_{n+1})$ per ogni $n \in \mathbf{Z}_+$ costituiscono un insieme di misura nulla. \square

14. Esempi di trasformazioni mescolanti

Incominciamo dando l'esempio di una trasformazione ergodica che non è totalmente ergodica.

Sia Ω un insieme costituito da due punti $\Omega = \{\omega_1, \omega_2\}$ e sulla famiglia delle parti di Ω si definisca una misura di probabilità μ mediante $\mu(\{\omega_j\}) = p_j$ con $p_j = 1/2$ ($j = 1, 2$). La trasformazione $T : \Omega \rightarrow \Omega$ definita da $T\omega_1 = \omega_2$ e $T\omega_2 = \omega_1$, conserva la misura ed è invertibile; essa è ergodica perché i soli insiemi invarianti sono \emptyset e Ω , ma non è totalmente ergodica, perché, per esempio, tanto $\{\omega_1\}$ quanto $\{\omega_2\}$ sono invarianti rispetto a T^2 , che, perciò, non è invariante.

Per contro, se a non è una radice dell'unità, la rotazione T_a della circonferenza unitaria è totalmente ergodica, perché $T_a^k z = a^k z$ e neanche a^k è una radice dell'unità, sicché T_a^k è ergodica quale che sia $k \in \mathbf{N}$.

Riesaminiamo qui gli esempi di trasformazioni ergodiche della sezione 1.5.

ESEMPIO 14.1. L'identità I sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ è fortemente mescolante se, e solo e se, è ergodica, vale a dire quando tutti gli insiemi di \mathcal{F} hanno probabilità eguale a 0 o a 1.

ESEMPIO 14.2. Nessuna rotazione T_a della circonferenza unitaria \mathcal{K} con $a \neq 1$ può essere debolmente mescolante. Infatti, si consideri la funzione $\varphi : \mathcal{K} \rightarrow \mathcal{K}$ definita da $\varphi(z) := z$. Per questa si ha

$$(U_{T_a} \varphi)(z) = \varphi(T_a z) = T_a z = az = a \varphi(z),$$

sicché a è un autovalore e φ un'autofunzione di U_{T_a} ; basta, ora, far riferimento al Teorema 10.6.

ESEMPIO 14.3. La traslazione bilatera è fortemente mescolante Siano A e B due cilindri

$$A = \{ \{x_n : n \in \mathbf{Z}\} : x_a = i(0), x_{a+1} = i(1), \dots, x_{a+r} = i(r) \}$$

$$B = \{ \{x_n : n \in \mathbf{Z}\} : x_b = j(0), x_{b+1} = j(1), \dots, x_{b+r} = j(s) \}.$$

Pur di prendere n sufficientemente grande, piú precisamente, se $n > b+s-a$, si ha

$$\mu(T^{-n} A \cap B) = p_{i(0)} p_{i(1)} \dots p_{i(r)} p_{j(0)} p_{j(1)} \dots p_{j(s)} = \mu(A) \mu(B).$$

Il Teorema 10.6 dà quindi l'asserto.

Analogamente si fa vedere che è fortemente mescolante la traslazione unilatera.

TEOREMA 14.1. *Per una traslazione bilatera di MARKOV- (p, Π) T sono equivalenti le affermazioni:*

- (a) T è fortemente mescolante;
- (b) la matrice di transizione Π è regolare, vale a dire che esiste un numero naturale K tale che la matrice Π^K abbia tutti gli elementi positivi, $\Pi^K > 0$.

DIMOSTRAZIONE. (a) \implies (b) Si può procedere come nella dimostrazione del Teorema 9.3 per verificare che si ha, quali che siano gli stati i e j ,

$$\lim_{n \rightarrow +\infty} p_{ij}^{(n)} = p_j > 0,$$

sicché, per n sufficientemente grande, è $p_{ij}^{(n)} > 0$.

L'implicazione inversa (b) \implies (a) si dimostra come nel Teorema appena richiamato. \square

Non daremo l'esempio di una trasformazione debolmente mescolante che non sia fortemente mescolante, benché, come si sia visto, tali trasformazioni debbano essere numerose. Per una tale trasformazione si veda l'esempio di KAKUTANI riportato in [53, pp. 87–89].

15. Misure invarianti

Si supponga che sia dato uno spazio misurale $(\Omega, \mathcal{F}, \mu)$, ove la misura μ è σ -finita, e che $T : \Omega \rightarrow \Omega$ sia una trasformazione misurabile, che per semplicità si supporrà anche invertibile. Il problema è di trovare una misura ν definita sullo spazio misurabile (Ω, \mathcal{F}) tale che la trasformazione T sia invariante rispetto a ν . Il problema, in questa generalità non ha molto senso, perché, ad esempio, la misura identicamente nulla, soddisfa alla proprietà richiesta. Per eliminare queste misure invarianti banali, si richiede che la misura data μ sia assolutamente continua rispetto a ν , $\mu \ll \nu$. Tuttavia, anche così il problema ammette soluzioni non significative; basta, per esempio prendere come misura ν la misura del contare, che è ovviamente una misura invariante. A questo inconveniente si rimedia chiedendo che ν sia σ -finita.

Vi sono due problemi legati all'esistenza delle misure invarianti: nelle condizioni specificate

- (a) esiste una misura *finita* ν tale che la misura data μ sia assolutamente continua rispetto a ν e che ν risulti invariante rispetto alla trasformazione T ?
- (b) esiste una misura σ -finita ν con le stesse proprietà?

Senza perdita di generalità si può supporre che la misura data μ sia finita. Infatti, perché μ è σ -finita esiste una partizione numerabile di Ω $\{E_n : n \in \mathbf{N}\}$ in insiemi disgiunti e misurabili. Sia $\{\alpha_n\}$ una successione di numeri reali positivi, $\alpha_n > 0$ ($n \in \mathbf{N}$), tali che sia convergente la serie $\sum_{n \in \mathbf{N}} \alpha_n \mu(E_n)$. Si definisca una misura μ_0 su (Ω, \mathcal{F}) mediante

$$\mu_0(E) := \sum_{n \in \mathbf{N}} \alpha_n \mu(E \cap E_n).$$

La misura μ_0 così introdotta è finita ed è equivalente a μ (cioè le due misure μ_0 e μ si annullano sugli stessi insiemi, o, equivalentemente, sono l'una assolutamente continua rispetto all'altra, $\mu \ll \mu_0$ e $\mu_0 \ll \mu$). Pertanto, μ è assolutamente continua rispetto a ν se, e solo e se, tale è anche μ_0 . Sostituendo μ con μ_0 si può, quindi, supporre che μ sia una misura di probabilità $\mu(\Omega) = 1$.

Si può anche supporre, di nuovo senza perdita di generalità, che la trasformazione T non sia *singolare*; una trasformazione (invertibile) si dice singolare se esiste un insieme misurabile E tale che $\mu(E) = 0$, mentre almeno uno tra gli insiemi TE e $T^{-1}E$ ha misura diversa da zero, $\mu(T^{-1}E) \vee \mu(TE) > 0$. Sia $\{\alpha_n : n \in \mathbf{Z}\}$ una successione di numeri reali positivi, $\alpha_n > 0$ ($n \in \mathbf{Z}$), tali che $\sum_{n \in \mathbf{Z}} \alpha_n = 1$. Posto

$$\mu_1(E) := \sum_{n \in \mathbf{Z}} \alpha_n \mu(T^n E),$$

si controlla immediatamente che μ_1 è una misura di probabilità e che μ è assolutamente continua rispetto a μ_1 , $\mu \ll \mu_1$. Se μ è assolutamente continua rispetto ad una misura invariante ν , $\mu \ll \nu$, allora anche μ_1 è assolutamente continua rispetto a ν , $\mu_1 \ll \nu$; infatti, se fosse $\nu(E) = 0$, allora per l'invarianza rispetto a T si avrebbe $\nu(T^n E) = 0$ per ogni $n \in \mathbf{Z}$, e, perciò, anche $\mu(T^n E) = 0$ per ogni $n \in \mathbf{Z}$, sicché si avrebbe $\mu_1(E) = 0$, come asserito. Di conseguenza, il problema di trovare la misura ν quando è data μ , equivale a quello di trovare ν quando è data μ_1 . D'altro canto, se fosse $\mu_1(E) = 0$, si avrebbe anche $\mu_1(TE) = \mu_1(T^{-1}E) = 0$; dunque, si può supporre che T non sia singolare.

Infine, si può limitare la ricerca alle misure invarianti ν che siano equivalenti a μ . Infatti, se esiste una misura invariante σ -finita ν , rispetto alla quale μ è assolutamente continua, esiste un'altra misura invariante e σ -finita ν_0 che è equivalente a μ . Inoltre se ν è finita, tale sarà anche ν_0 . Per stabilire

questo punto, si ricorra al teorema di RADON–NIKODÝM; esiste una funzione positiva $f \geq 0$ tale che, per ogni $E \in \mathcal{F}$, sia

$$\mu(E) = \int_E f d\nu.$$

Posto $A := \{f = 0\}$, si ha $\mu(A) = 0$, e, poiché si è supposto che T non sia singolare, $\mu(T^n A) = 0$ per ogni $n \in \mathbf{Z}$. Si introduca l'insieme

$$B := \cup_{n \in \mathbf{Z}} T^n A$$

che risulta essere misurabile, invariante e di misura μ nulla, $\mu(B) = 0$. Mediante

$$\nu_0(E) := \nu(E \setminus B),$$

si definisca una misura $\nu_0 : \mathcal{F} \rightarrow \mathbf{R}_+$ che è σ -finita ed invariante rispetto a T . Se $\nu_0(E) = 0$, allora $\nu(E \setminus B) = 0$ e, quindi, $\mu(E \setminus B) = 0$; poiché

$$\mu(E) = \mu(E \cap B) + \mu(E \setminus B),$$

segue che $\mu(E) = 0$. Viceversa, se $\mu(E) = 0$, necessariamente si ha anche $\mu(E \setminus B) = 0$. Se $\omega \in E \setminus B$, allora $f(\omega) > 0$, sicché da

$$\mu(E \setminus B) = \int_{E \setminus B} f d\nu$$

segue che $\nu(E \setminus B) = 0$ e, di qui che $\nu_0(E) = 0$.

Il problema della ricerca di una misura invariante è ora ricondotto al seguente: *Data una trasformazione T invertibile, misurabile e non singolare sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ trovare una misura finita (o σ -finita) ν che sia equivalente a μ ed invariante rispetto a T .*

Occorre un nuovo concetto. Sia $\{B_n : n \in \mathbf{N}\}$ una partizione misurabile dell'insieme misurabile A

$$a = \cup_{n \in \mathbf{N}} B_n \quad \text{e} \quad E_j \cap E_k = \emptyset \quad (j \neq k).$$

Se $\{C_n : n \in \mathbf{N}\}$ è un'altra successione di insiemi misurabili a due a due disgiunti con unione eguale a C ,

$$C = \cup_{n \in \mathbf{N}} C_n \quad \text{e} \quad C_j \cap C_k = \emptyset \quad (j \neq k),$$

e se esiste una successione $\{k(n) : n \in \mathbf{N}\}$ di numeri interi tali che per ogni $n \in \mathbf{N}$ sia

$$T^{k(n)} E_n = F_n,$$

allora dice che gli insiemi A e C sono *equivalenti per decomposizione numerabile*.

In analogia con la definizione di DEDEKIND di insieme finito, si dirà che un insieme misurabile A è *limitatto* se A non è equivalente per decomposizione numerabile ad un suo sottoinsieme proprio. Si dirà che A è σ -limitato se è

l'unione di una famiglia numerabile di insiemi limitati. una trasformazione T si dirà limitata (rispettivamente σ -limitata) se l'intero spazio Ω è limitato (rispettivamente σ -limitato).

Se il problema dell'esistenza di una misura invariante ha una soluzione ν , ogni insieme misurabile A che abbia misura finita rispetto a ν , $\nu(A) < +\infty$ è limitato.; se consegue che se ν è una misura finita, T è limitata, se ν è σ -finita, allora T è σ -finita. È stato dimostrato che vale il viceversa: HOPF stabilì che se una trasformazione è limitata, allora esiste una misura finita invariante equivalente a μ , mentre HALMOS dimostrò che se T è σ -limitata, allora esiste una misura invariante σ -finita equivalente a μ .

CAPITOLO 3

L'entropia in teoria ergodica

1. Partizioni e sottoalgebre

Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$, si consideri una partizione misurabile $\{A_\iota : \iota \in I\}$, vale a dire una famiglia di insiemi misurabili, A_ι per ogni $\iota \in I$, a due a due disgiunti $A_\iota \cap A_k = \emptyset$ se $k \neq \iota$, e tale che $\Omega = \cup_\iota A_\iota$. La partizione si dice *numerabile* se $\text{card}(I) = \aleph_0$, *finita*, se $\text{card}(I) < \aleph_0$. Si indicherà con \mathcal{P}_k la famiglia delle partizioni di Ω in al più k insiemi misurabili e disgiunti, e con $\mathcal{P} = \cup_{k \in \mathbf{N}} \mathcal{P}_k$ la famiglia delle partizioni finite di Ω . Data una partizione misurabile e finita, $\pi \in \mathcal{P}_n$ sia $\sigma(\pi) \subset \mathcal{F}$ la tribú generata da π

$$\sigma(\pi) := \left\{ A \in \mathcal{F} : A = \cup_{i=1}^r A_{k(i)}, A_{k(i)} \in \pi \ (i = 1, 2, \dots, r), r \leq n \right\},$$

vale a dire la famiglia delle unioni finite di insiemi di π .

Se (Ω, \mathcal{F}) è uno spazio misurabile, si dirà che $A \in \mathcal{F}$ è un *atomo*, se ogni sottoinsieme misurabile B di A , $B \subset A$, $B \in \mathcal{F}$, o è l'insieme vuoto, $B = \emptyset$ oppure coincide con A , $B = A$, sicché \emptyset è il solo sottoinsieme misurabile proprio di A . Due distinti atomi di (Ω, \mathcal{F}) sono necessariamente disgiunti. Se A_1, A_2, \dots, A_n sono tutti gli atomi di (Ω, \mathcal{F}) e se $\Omega = \cup_{j=1}^n A_j$, allora ogni insieme misurabile si esprime come unione finita e disgiunta di atomi. In questo caso si dice che \mathcal{F} è una tribú *atomica* finita o che è generata dalla famiglia di atomi $\{A_1, A_2, \dots, A_n\}$.

D'altra parte, se \mathcal{B} è una sottotribú finita di \mathcal{F} le si può far corrispondere una partizione misurabile, rispetto a \mathcal{B} che sarà indicata con $\pi(\mathcal{B})$. e che si dirà la *partizione generata da \mathcal{B}* .

PROPOSIZIONE 1. *Ogni sotto-tribú finita \mathcal{B} di \mathcal{F} è un'algebra atomica finita di \mathcal{F} .*

DIMOSTRAZIONE. Poiché \mathcal{B} è finita, si può scrivere $\mathcal{B} = \{B_1, B_2, \dots, B_s\}$. Si formino le 2^s intersezioni

$$\cap_{i=1}^s C_i,$$

ove, per ogni indice i , si prende $C_i = B_i$ oppure $C_i = B_i^c$. Quelle tra queste intersezioni che risultano non vuote sono atomi di \mathcal{B} ; infatti, ognuna delle intersezioni considerate appartiene a \mathcal{B} , che non può contenere altri sottoinsiemi non vuoti. \square

Per ogni algebra, o, ciò che è equivalente, sotto-tribú finita, \mathcal{B} di \mathcal{F} si indicherà con $\pi(\mathcal{B})$ la famiglia degli atomi che la genera.

La famiglia \mathcal{P} delle partizioni finite di (Ω, \mathcal{F}) può essere ordinata parzialmente come segue. Date due partizioni finite π e π' di (Ω, \mathcal{F}) si dirà che π' è un *raffinamento* di π e si scriverà $\pi \leq \pi'$ se ogni atomo di π è l'unione di atomi di π' .

Con la relazione d'ordine appena introdotta, la famiglia \mathcal{P} delle partizioni finite di (Ω, \mathcal{F}) è un insieme diretto. Infatti siano $\pi = \{A_1, A_2, \dots, A_n\}$ e $\pi' = \{B_1, B_2, \dots, B_m\}$ due partizioni finite di (Ω, \mathcal{F}) . Posto

$$\pi \vee \pi' := \{A_j \cap B_k : i = 1, 2, \dots, n; j = 1, 2, \dots, m\},$$

si controlla immediatamente che $\pi \vee \pi'$ è una partizione finita di (Ω, \mathcal{F}) , che essa è un raffinamento tanto di π quanto di π' e che ogni raffinamento comune di π e di π' è anche necessariamente un raffinamento di $\pi \vee \pi'$.

TEOREMA 1.1. *Siano π e π' due partizioni di (Ω, \mathcal{F}) e \mathcal{B} e \mathcal{C} due sotto-tribú finite di \mathcal{F} ; allora*

- (a) π' è un raffinamento di π se, e solo se, $\sigma(\pi) \subset \sigma(\pi')$;
- (b) \mathcal{B} è inclusa in \mathcal{C} se, e solo se, $\pi(\mathcal{C})$ è un raffinamento di $\pi(\mathcal{B})$, $\pi(\mathcal{B}) \leq \pi(\mathcal{C})$;
- (c) $\pi(\mathcal{B} \vee \mathcal{C}) = \pi(\mathcal{B}) \vee \pi(\mathcal{C})$;
- (d) $\sigma(\pi \vee \pi') = \sigma(\pi) \vee \sigma(\pi')$.

DIMOSTRAZIONE. (a) Si supponga che π' sia un raffinamento di π , $\pi \leq \pi'$. Se l'insieme A appartiene alla tribú $\sigma(\pi)$, è

$$A = \cup_{i=1}^r A_{k(i)} = \cup_{i=1}^r \cup_{j=1}^{s(i)} B_{k(i),j},$$

ove $B_{k(i),j} \in \pi'$ ($i = 1, 2, \dots, r; j = 1, 2, \dots, s(i)$), onde $A \in \sigma(\pi')$. Viceversa, se $\sigma(\pi) \subset \sigma(\pi')$ e A appartiene a π , si ha $A \in \sigma(\pi) \subset \sigma(\pi')$ cosicché A è unione di elementi di π' .

(b) Se $\mathcal{B} \subset \mathcal{C}$ si ha $\pi(\mathcal{B}) \subset \pi(\mathcal{C})$; perciò, se A appartiene a $\pi(\mathcal{B})$ si ha $A = \cup_{j=1}^r C_j$, ove $C_j \in \pi(\mathcal{C})$ ($j = 1, 2, \dots, r$) onde $\pi(\mathcal{B}) \subset \pi(\mathcal{C})$. Viceversa, se $\pi(\mathcal{C})$ è un raffinamento di $\pi(\mathcal{B})$, $\pi(\mathcal{B}) \leq \pi(\mathcal{C})$, e B è in \mathcal{B} si ha

$$B = \cup_{i=1}^p B_i = \cup_{i=1}^p \cup_{j=1}^{s(i)} C_{i,j}$$

ove $B_i \in \pi(\mathcal{B})$ e $C_{i,j} \in \pi(\mathcal{C})$ ($i = 1, 2, \dots, p; j = 1, 2, \dots, s(i)$). Dunque B è in \mathcal{C} .

La (c) e la (d) sono ovvie. □

Se $\pi = \{A_1, A_2, \dots, A_n\}$ e \mathcal{B} sono, rispettivamente, una partizione di (Ω, \mathcal{F}) e una sotto-tribú di \mathcal{F} , si pone, per $r \in \mathbf{Z}$,

$$\begin{aligned} T^r \pi &:= \{T^r A_1, T^r A_2, \dots, T^r A_n\}, \\ T^r \mathcal{B} &:= \{T^r A : A \in \mathcal{B}\}. \end{aligned}$$

TEOREMA 1.2. *Siano T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$, π e π' due partizioni di (Ω, \mathcal{F}) , \mathcal{B} e \mathcal{C} due sotto-tribú di \mathcal{F} : allora, per ogni $n \in \mathbf{N}$, si ha*

- (a) $\pi(T^{-n} \mathcal{B}) = T^{-n} \pi(\mathcal{B})$;
- (b) $\sigma(T^{-n} \pi) = T^{-n} \sigma(\pi)$;
- (c) $T^{-n} (\mathcal{B} \vee \mathcal{C}) = T^{-n} \mathcal{B} \vee T^{-n} \mathcal{C}$;
- (d) $T^{-n} (\pi \vee \pi') = T^{-n} \pi \vee T^{-n} \pi'$;
- (e) se π' è un raffinamento di π , $\pi \leq \pi'$, allora $T^{-n} \pi'$ è un raffinamento di $T^{-n} \pi$, $T^{-n} \pi \leq T^{-n} \pi'$;
- (f) $T^{-n} \mathcal{B} \subset T^{-n} \mathcal{C}$ se $\mathcal{B} \subset \mathcal{C}$.

DIMOSTRAZIONE. (a) L'insieme B appartiene alla partizione $\pi(T^{-n} \mathcal{B})$ se, e solo se, si può scrivere

$$B = \bigcap_{i=1}^r T^{-n} B_i = T^{-n} (\bigcap_{i=1}^r B_i),$$

con $B_i \in \mathcal{B}$ ($i = 1, 2, \dots, r$), o, equivalentemente, se, e solo se, B è in $T^{-n} \pi(\mathcal{B})$.

(b) Dire che l'insieme A appartiene alla sotto-tribú $\sigma(T^{-n} \pi)$ equivale a dire che si può scrivere

$$A = \bigcup_{j=1}^r T^{-n} A_j = T^{-n} (\bigcup_{j=1}^r A_j),$$

con $A_j \in \pi$ ($j = 1, 2, \dots, r$), ciò che equivale a dire che A appartiene a $T^{-n} \sigma(\pi)$.

(c) Dire che l'insieme A appartiene a $T^{-n} (\mathcal{B} \vee \mathcal{C})$ equivale a dire che A può essere scritto nella forma

$$A = T^{-n} (B_j \cap C_k) = T^{-n} B_j \cap T^{-n} C_k,$$

con $B_j \in \mathcal{B}$ e $C_k \in \mathcal{C}$, ciò che equivale a dire che A appartiene a $T^{-n} \mathcal{B} \vee T^{-n} \mathcal{C}$.

(d) L'insieme A appartiene a $T^{-n}(\pi \vee \pi')$ se, e solo se, si può scrivere nella forma

$$A = T^{-n}(B \cap C) = (T^{-n}B) \cap (T^{-n}C),$$

con $B \in \pi$ e $C \in \pi'$, o, ancora, se, e solo se, appartiene a $T^{-n}\pi \vee T^{-n}\pi'$.

(e) Se π' è un raffinamento di π , $\pi \leq \pi'$, allora un insieme A di π si scrive nella forma $A = \cup_{j=1}^n B_j$, con $B_j \in \pi'$ ($j = 1, 2, \dots, n$), onde

$$T^{-n}A = \cup_{j=1}^n (T^{-n}B_j),$$

sicché $T^{-n}\pi'$ è un raffinamento di $T^{-n}\pi$.

La (f) è banale. □

2. La definizione di entropia

Introduciamo il concetto di entropia che viene dalla Meccanica Statistica e dalla Teoria dell'Informazione. Esso fu introdotto nella teoria ergodica da KOLMOGOROV.

DEFINIZIONE 2.1. Dati uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ ed una sotto-tribù finita \mathcal{A} di \mathcal{F} , avente $\pi(\mathcal{A}) := \{A_1, A_2, \dots, A_n\}$ come partizione associata, si definisce *entropia* di \mathcal{A} o di $\pi(\mathcal{A})$, la funzione

$$H(\mathcal{A}) = H(\pi(\mathcal{A})) := - \sum_{j=1}^n \mu(A_j) \log \mu(A_j), \quad (2.1)$$

ove i logaritmi, qui e nel seguito, si intendono in base 2, e si adotta la convenzione

$$0 \log 0 =: 0. \quad (2.2)$$

Per le proprietà dell'entropia si consultino i libri [2, 6, 7, 28]. Di seguito sono riportate due proprietà elementari che serviranno nel seguito.

TEOREMA 2.1. *Per l'entropia $H : \mathcal{P} \rightarrow \mathbf{R}$ valgono le seguenti proprietà:*

- (a) *per ogni partizione finita $\pi \in \mathcal{P}$, $H(\pi) \geq 0$;*
- (b) *per ogni partizione $\pi \in \mathcal{P}_n$, $H(\pi) \leq \log n$.*

DIMOSTRAZIONE. Si supponga che sia $\pi = \{A_1, A_2, \dots, A_n\}$. La proprietà (a) è evidente perché le probabilità $\mu(A_j)$, se non sono nulle, e in questo caso si ricorre alla (2.2), appartengono all'intervallo $]0, 1]$ sicché ogni termine della (2.1) è positivo.

(b) La funzione φ definita nell'intervallo $[0, 1]$ da

$$\varphi(t) := \begin{cases} -t \log t, & t \in]0, 1], \\ 0, & t = 0, \end{cases}$$

è concava perché

$$\varphi''(t) = -\frac{1}{x \ln 2} < 0.$$

Perciò, la diseguaglianza di JENSEN dà

$$\begin{aligned} H(\pi) &= -\sum_{j=1}^n \mu(A_j) \log \mu(A_j) = \sum_{j=1}^n \varphi(\mu(A_j)) \\ &= n \sum_{j=1}^n \frac{\varphi(\mu(A_j))}{n} \leq n \varphi\left(\frac{\sum_{j=1}^n \mu(A_j)}{n}\right) \\ &= n \varphi\left(\frac{1}{n}\right) = \log n, \end{aligned}$$

che è l'asserto. □

DEFINIZIONE 2.2. Dato lo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e due sotto-tribú \mathcal{A} e \mathcal{B} di \mathcal{F} , si dirà che esse sono μ -equivalenti, e si scriverà $\mathcal{A} \stackrel{\circ}{=} \mathcal{B}$, se, per ogni $A \in \mathcal{A}$, esiste $B \in \mathcal{B}$ tale che $\mu(A \Delta B) = 0$ e, per ogni $B \in \mathcal{B}$ esiste $A \in \mathcal{A}$ tale che $\mu(A \Delta B) = 0$. Se, invece, per ogni $A \in \mathcal{A}$, esiste $B \in \mathcal{B}$ tale che $\mu(A \Delta B) = 0$, si scriverà $\mathcal{A} \stackrel{\circ}{\subset} \mathcal{B}$.

PROPOSIZIONE 2. Se \mathcal{A} e \mathcal{B} sono due sotto-tribú finite ed equivalenti di \mathcal{F} , $\mathcal{A} \stackrel{\circ}{=} \mathcal{B}$, allora

$$H(\mathcal{A}) = H(\mathcal{B}).$$

DIMOSTRAZIONE. Se \mathcal{A} e \mathcal{B} sono sotto-tribú finite ed equivalenti di \mathcal{F} e la partizione generata da \mathcal{A} è $\pi(\mathcal{A}) = \{A_1, A_2, \dots, A_n\}$ e $\mu(A_j) > 0$ per $j = 1, 2, \dots, k$, mentre $\mu(A_j) = 0$ per $j = k+1, k+2, \dots, n$, allora per la partizione $\pi(\mathcal{B}) = \{B_1, B_2, \dots, B_s\}$ risulta $\mu(A_j \Delta B_j) = 0$ se $j = 1, 2, \dots, k$, mentre $\mu(B_j) = 0$ per $j \geq k+1$. Allora, se A_j è un insieme della partizione $\pi(\mathcal{A})$ generata da \mathcal{A} con $\mu(A_j) > 0$, esiste $B_j \in \pi(\mathcal{B})$ tale che $\mu(A_j \Delta B_j) = 0$. Pertanto

$$\begin{aligned} \mu(A_j) &= \mu(A_j \setminus B_j) + \mu(A_j \cap B_j) = \mu(A_j \cap B_j) \\ &= \mu(A_j \cap B_j) + \mu(B_j \setminus A_j) = \mu(B_j). \end{aligned}$$

Poiché $\mu(B_j) = 0$ se $j \geq k$, si ha

$$\begin{aligned} H(\mathcal{A}) &= -\sum_{j=1}^n \mu(A_j) \log \mu(A_j) = -\sum_{j=1}^k \mu(A_j) \log \mu(A_j) \\ &= -\sum_{j=1}^k \mu(B_j) \log \mu(B_j) = -\sum_{j=1}^s \mu(B_j) \log \mu(B_j) = H(\mathcal{B}), \end{aligned}$$

ciò che conclude la dimostrazione. \square

DEFINIZIONE 2.3. Date due sottotribú finite \mathcal{A} e \mathcal{B} di \mathcal{F} tali che $\mu(A_j) > 0$ per ogni insieme A_j della partizione $\pi(\mathcal{A}) = \{A_1, A_2, \dots, A_n\}$ generata da \mathcal{A} , si chiama *entropia condizionata di \mathcal{B} data \mathcal{A}* la funzione $H : \mathcal{P} \times \mathcal{P} \rightarrow \mathbf{R}_+$ definita da

$$H(\mathcal{B} | \mathcal{A}) := - \sum_{i=1}^n \mu(A_i) \sum_{j=1}^s \mu(B_j | A_i) \log \mu(B_j | A_i), \quad (2.3)$$

se $\pi(\mathcal{A})$ e $\pi(\mathcal{B}) = \{B_1, B_2, \dots, B_s\}$ sono le partizioni generate da \mathcal{A} e da \mathcal{B} .

Si conviene che la definizione (2.3) conservi il suo significato anche quando qualche insieme della partizione $\pi(\mathcal{A})$ abbia probabilità nulla; in tal caso si pone eguale a 0 il termine corrispondente nella somma (2.3).

PROPOSIZIONE 3. *Se \mathcal{A} , \mathcal{B} e \mathcal{C} sono sotto-tribú finite di \mathcal{F} e se gli insiemi delle partizioni che esse generano sono tutti di probabilità strettamente positiva, quando la corrispondente sotto-tribú appare come condizionante, allora valgono le relazioni:*

- (a) se $\mathcal{N} := \{\emptyset, \Omega\}$ denota la tribú banale, $H(\mathcal{A} | \mathcal{N}) = H(\mathcal{A})$;
- (b) se \mathcal{A} e \mathcal{B} sono equivalenti, $\mathcal{A} \stackrel{\circ}{=} \mathcal{B}$, allora, per ogni $\mathcal{C} \in \mathcal{P}$, è $H(\mathcal{A} | \mathcal{C}) = H(\mathcal{B} | \mathcal{C})$;
- (c) se \mathcal{B} e \mathcal{C} sono equivalenti, $\mathcal{B} \stackrel{\circ}{=} \mathcal{C}$, allora, per ogni $\mathcal{A} \in \mathcal{P}$, è $H(\mathcal{A} | \mathcal{B}) = H(\mathcal{A} | \mathcal{C})$.

DIMOSTRAZIONE. La (a) è ovvia.

(b) In questo caso si ha $\pi(\mathcal{A}) = \{A_1, A_2, \dots, A_n\}$, $\pi(\mathcal{B}) = \{B_1, B_2, \dots, B_s\}$, $\mu(A_j \Delta B_j) = 0$ per $j = 1, 2, \dots, n$, mentre $\mu(B_j) = 0$ per $j = n+1, n+2, \dots, s$. Inoltre, per $j = 1, 2, \dots, n$, si ha $\mu(A_j) = \mu(A_j \cap B_j) = \mu(B_j)$. Se $\pi(\mathcal{C}) = \{C_1, C_2, \dots, C_r\}$, si ha

$$A_i \cap C_j = (A_i \cap C_j \cap B_i) \cup \{(A_i \cap C_j) \setminus B_i\}$$

sicché

$$\mu(A_i \cap C_j) = \mu(A_i \cap C_j \cap B_i) + \mu[(A_i \cap C_j) \setminus B_i] = \mu(A_i \cap C_j \cap B_i).$$

Similmente

$$\mu(B_i \cap C_j) = \mu(A_i \cap C_j \cap B_i)$$

onde

$$\begin{aligned}
 H(\mathcal{A} | \mathcal{C}) &= - \sum_{j=1}^r \mu(C_j) \sum_{i=1}^n \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \\
 &= - \sum_{j=1}^r \mu(C_j) \sum_{i=1}^n \frac{\mu(A_i \cap C_j \cap B_i)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j \cap B_i)}{\mu(C_j)} \\
 &= - \sum_{j=1}^r \mu(C_j) \sum_{i=1}^n \frac{\mu(B_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(B_i \cap C_j)}{\mu(C_j)} = H(\mathcal{B} | \mathcal{C}),
 \end{aligned}$$

poiché si ha $0 \leq \mu(B_i \cap C_j) \leq \mu(B_i) = 0$ per $i = n+1, n+2, \dots, s$.

La dimostrazione della (c) è del tutto analoga. \square

TEOREMA 2.2. *Siano \mathcal{A} , \mathcal{B} e \mathcal{C} sotto-tribú finite di \mathcal{F} . Valgono le proprietà*

(a) *(additività condizionata forte)*

$$H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) = H(\mathcal{A} | \mathcal{C}) + H(\mathcal{B} | \mathcal{A} \vee \mathcal{C});$$

(b) *(additività forte)* $H(\mathcal{A} \vee \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B} | \mathcal{A})$;

(c) $H(\mathcal{A} | \mathcal{C}) \leq H(\mathcal{B} | \mathcal{C})$ se $\mathcal{A} \subset \mathcal{B}$;

(d) $H(\mathcal{A}) \leq H(\mathcal{B})$ se $\mathcal{A} \subset \mathcal{B}$;

(e) $H(\mathcal{A} | \mathcal{B}) \geq H(\mathcal{A} | \mathcal{C})$ se $\mathcal{B} \subset \mathcal{C}$;

(f) $H(\mathcal{A} | \mathcal{C}) \leq H(\mathcal{A})$;

(g) *(subadditività condizionata)* $H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) \leq H(\mathcal{A} | \mathcal{C}) + H(\mathcal{B} | \mathcal{C})$;

(h) *(subadditività)* $H(\mathcal{A} \vee \mathcal{B}) \leq H(\mathcal{A}) + H(\mathcal{B})$.

DIMOSTRAZIONE. (a) Si considerino le partizioni generate dalle sotto-tribú \mathcal{A} , \mathcal{B} e \mathcal{C} ,

$$\begin{aligned}
 \pi(\mathcal{A}) &= \{A_1, A_2, \dots, A_n\}, & \pi(\mathcal{B}) &= \{B_1, B_2, \dots, B_s\} \\
 \pi(\mathcal{C}) &= \{C_1, C_2, \dots, C_r\}.
 \end{aligned}$$

Non è restrittivo supporre che tutti gli atomi abbiano probabilità strettamente positiva.

Se è $\mu(A_i \cap C_k) > 0$ si ha

$$\frac{\mu(A_i \cap B_j \cap C_k)}{\mu(C_k)} = \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(A_i \cap C_k)} \frac{\mu(A_i \cap C_k)}{\mu(C_k)}.$$

Se, invece, è $\mu(A_i \cap C_k) = 0$, si ha anche $\mu(A_i \cap B_j \cap C_k) = 0$ sicché i termini corrispondenti non compaiono nella somma che definisce

$$H(\mathcal{A} \vee \mathcal{B} | \mathcal{C});$$

pertanto

$$\begin{aligned}
H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) &= - \sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(A_i \cap C_k)} \\
&\quad - \sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)} \\
&= H(\mathcal{B} | \mathcal{A} \vee \mathcal{C}) - \sum_{i,k} \mu(A_i \cap C_k) \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)} \\
&= H(\mathcal{A} | \mathcal{C}) + H(\mathcal{B} | \mathcal{A} \vee \mathcal{C}).
\end{aligned}$$

La (b) si ottiene dalla (a) prendendo $\mathcal{C} = \mathcal{N}$.

(c) Poiché $\mathcal{A} \vee \mathcal{B} = \mathcal{B}$ se $\mathcal{A} \subset \mathcal{B}$, la (a) dà

$$H(\mathcal{B} | \mathcal{C}) = H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) = H(\mathcal{A} | \mathcal{C}) + H(\mathcal{B} | \mathcal{A} \vee \mathcal{C}) \geq H(\mathcal{A} | \mathcal{C}).$$

La (d) scende prendendo $\mathcal{C} = \mathcal{N}$ nella (c).

(e) Si ponga

$$\alpha_k := \frac{\mu(B_j \cap C_k)}{\mu(B_j)}, \quad x_k := \frac{\mu(A_i \cap C_k)}{\mu(C_k)};$$

poiché gli α_k definiscono una legge di probabilità e la funzione $\varphi(t) := t \log t$ è convessa, si ha

$$\begin{aligned}
\sum_{k=1}^r \alpha_k x_k \log x_k &\geq \left\{ \sum_{k=1}^r \alpha_k x_k \right\} \log \left\{ \sum_{k=1}^r \alpha_k x_k \right\} \\
&= \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)}.
\end{aligned} \tag{2.4}$$

Infatti, poiché $\mathcal{B} \subset \mathcal{C}$, ogni insieme B_j può essere scritto nella forma

$$B_j = \cup_{h \in I_j} C_{j(h)},$$

ove I_j è un opportuno insieme finito di di indici e gli insiemi che compaiono nell'unione sono disgiunti. Pertanto

$$\begin{aligned}
\sum_{k=1}^s \frac{\mu(A_i \cap C_k) \mu(B_j \cap C_k)}{\mu(C_k)} &= \sum_{k=1}^s \frac{\mu(A_i \cap C_k)}{\mu(C_k)} \delta_{k,j(h)} \mu(C_k) \\
&= \sum_{h \in I_j} \mu(A_i \cap C_{j(h)}) = \mu(A_i \cap B_j).
\end{aligned}$$

Moltiplicando la (2.4) per $\mu(B_j)$ e sommando su i e j si ottiene

$$\begin{aligned} \sum_{i,j} \mu(A_i \cap B_j) \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \\ \leq \sum_{i,j,k} \mu(B_j \cap C_k) \frac{\mu(A_i \cap C_k)}{\mu(C_k)} \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)} \\ = \sum_{i,k} \mu(A_i \cap C_k) \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)}, \end{aligned}$$

vale a dire

$$-H(\mathcal{A} | \mathcal{B}) \leq -H(\mathcal{A} | \mathcal{C}),$$

onde l'asserto.

La (f) segue prendendo $\mathcal{B} = \mathcal{N}$.

(g) Si usino la (a) e la (e), poiché $\mathcal{C} \subset \mathcal{A} \vee \mathcal{C}$,

$$H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) = H(\mathcal{A} | \mathcal{C}) + H(\mathcal{B} | \mathcal{A} \vee \mathcal{C}) \leq H(\mathcal{A} | \mathcal{C}) + H(\mathcal{B} | \mathcal{C}).$$

(h) Si ponga $\mathcal{C} = \mathcal{N}$ nella (g). □

Nel prossimo teorema compaiono le trasformazioni che conservano la misura.

TEOREMA 2.3. *Se T conserva la misura su $(\Omega, \mathcal{F}, \mu)$, allora*

- (a) $H(T^{-1} \mathcal{A} | T^{-1} \mathcal{B}) = H(\mathcal{A} | \mathcal{B});$
- (b) $H(T^{-1} \mathcal{A}) = H(\mathcal{A}).$

DIMOSTRAZIONE. Basta osservare che, poiché T conserva la misura, si ha $\mu(T^{-1} A \cap T^{-1} B) = \mu(A \cap B)$. □

TEOREMA 2.4. *Per due sotto-tribú finite \mathcal{A} e \mathcal{B} di \mathcal{F} le seguenti affermazioni sono equivalenti:*

- (a) $\mathcal{A} \overset{\circ}{\subset} \mathcal{B};$
- (b) $H(\mathcal{A} | \mathcal{B}) = 0.$

DIMOSTRAZIONE. Si considerino le partizioni generate da \mathcal{A} e da \mathcal{B} rispettivamente,

$$\pi(\mathcal{A}) = \{A_1, A_2, \dots, A_n\} \quad \text{e} \quad \pi(\mathcal{B}) = \{B_1, B_2, \dots, B_s\}.$$

Non è restrittivo supporre che tutti gli insiemi di tali partizioni abbiano probabilità non nulla.

(a) \implies (b) L'ipotesi dice che, per ogni coppia di indici, j e k , è

$$\mu(A_j \cap B_k) = 0 \quad \text{oppure} \quad \mu(A_j \cap B_k) = \mu(B_k),$$

sicché $H(\mathcal{A} | \mathcal{B}) = 0$.

(b) \implies (a) Ogni termine della somma che definisce $H(\mathcal{A} | \mathcal{B})$ è nullo, sicché, per ogni coppia di indici, j e k , si ha

$$\mu(A_j \cap B_k) \log \frac{\mu(A_j \cap B_k)}{\mu(B_k)} = 0;$$

pertanto si ha o $\mu(A_j \cap B_k) = 0$ oppure $\mu(A_j \cap B_k) = \mu(B_k)$, ciò che significa $\mathcal{A} \overset{\circ}{\subset} \mathcal{B}$. \square

TEOREMA 2.5. *Per due sotto-tribù finite \mathcal{A} e \mathcal{B} di \mathcal{F} le seguenti affermazioni sono equivalenti:*

- (a) \mathcal{A} e \mathcal{B} sono indipendenti, nel senso che $\mu(A \cap B) = \mu(A) \mu(B)$ per ogni scelta di A in \mathcal{A} e di B in \mathcal{B} ;
- (b) $H(\mathcal{A} | \mathcal{B}) = H(\mathcal{A})$.

DIMOSTRAZIONE. (a) \implies (b) Segue immediatamente dalla definizione di entropia condizionata.

(b) \implies (a) L'eguaglianza supposta vera per ipotesi significa

$$-\sum_{j=1}^n \sum_{k=1}^s \mu(A_j \cap B_k) \log \frac{\mu(A_j \cap B_k)}{\mu(B_k)} = -\sum_{j=1}^n \mu(A_j) \log \mu(A_j).$$

Sfruttando la concavità della funzione $[0, 1] \ni t \mapsto \varphi(t) := -t \log t$, si ha, per ogni indice $j \in \{1, 2, \dots, n\}$,

$$-\sum_{k=1}^s \mu(A_j \cap B_k) \log \frac{\mu(A_j \cap B_k)}{\mu(B_k)} \leq -\mu(A_j) \log \mu(A_j)$$

con eguaglianza solo se il rapporto $\mu(A_j \cap B_k) / \mu(B_k)$ non dipende da k . Dunque tale rapporto è costante rispetto a k ,

$$\frac{\mu(A_j \cap B_k)}{\mu(B_k)} =: \alpha_j.$$

Dall'eguaglianza $\mu(A_j \cap B_k) = \alpha_j \mu(B_k)$ si ottiene sommando su k

$$\alpha_j = \mu(A_j),$$

sicché $\mu(A_j \cap B_k) = \mu(A_j) \mu(B_k)$ e ciò basta a dare l'asserto. \square

TEOREMA 2.6. Si consideri l'insieme \mathcal{P} di tutte le sotto-tribú finite di \mathcal{F} e l'insieme quoziente

$$\tilde{\mathcal{P}} := \mathcal{P} / \overset{\circ}{\equiv}$$

(in \mathcal{P} si identificano due sotto-tribú finite \mathcal{A} e \mathcal{B} di \mathcal{F} se $\mathcal{A} \overset{\circ}{\equiv} \mathcal{B}$). Posto

$$d(\mathcal{A}, \mathcal{B}) := H(\mathcal{A} | \mathcal{B}) + H(\mathcal{B} | \mathcal{A}),$$

la coppia $(\tilde{\mathcal{P}}, d)$ è uno spazio metrico.

DIMOSTRAZIONE. È noto che $d(\mathcal{A}, \mathcal{B}) \geq 0$; il segno d'eguaglianza vale se, e solo se, $\mathcal{A} \overset{\circ}{\equiv} \mathcal{B}$. Inoltre la simmetria vale per definizione. Infine, se \mathcal{C} è un'altra sotto-tribú finita di \mathcal{F} , è

$$H(\mathcal{A} | \mathcal{C}) \leq H(\mathcal{A} \vee \mathcal{B} | \mathcal{C}) = H(\mathcal{B} | \mathcal{C}) + H(\mathcal{A} | \mathcal{B} \vee \mathcal{C}) \leq H(\mathcal{B} | \mathcal{C}) + H(\mathcal{A} | \mathcal{B});$$

similmente si ha $H(\mathcal{C} | \mathcal{A}) \leq H(\mathcal{B} | \mathcal{A}) + H(\mathcal{C} | \mathcal{B})$ che insieme danno

$$d(\mathcal{A}, \mathcal{C}) \leq d(\mathcal{A}, \mathcal{B}) + d(\mathcal{B}, \mathcal{C}),$$

cioè la diseguaglianza triangolare. □

Il seguente lemma di analisi matematica elementare è necessario per poter definire l'entropia di una trasformazione che conserva la misura.

LEMMA 2.1. Sia $\{a_n : n \in \mathbf{N}\}$ una successione di numeri reali positivi tale che, per ogni scelta di due naturali n e m ,

$$a_{n+m} \leq a_n + a_m.$$

Allora, la successione $\{a_n/n\}$ ammette limite e risulta

$$\lim_{n \rightarrow +\infty} \frac{a_n}{n} = \inf_{n \in \mathbf{N}} \frac{a_n}{n}. \quad (2.5)$$

DIMOSTRAZIONE. Quali che siano i numeri naturali k e n si ha

$$a_{kn} = a_{(k-1)n+n} \leq a_{(k-1)n} + a_n \leq \dots \leq k a_n.$$

Fissato $n \in \mathbf{N}$ si ponga, per ogni $j \geq n$, $j = kn + i$ con $0 \leq i < n$. Allora

$$\frac{a_j}{j} = \frac{a_{kn+i}}{kn+i} \leq \frac{a_i}{kn} + \frac{a_{kn}}{kn} \leq \frac{a_i}{kn} + \frac{k a_n}{kn} = \frac{a_i}{kn} + \frac{a_n}{n}.$$

Si faccia tendere j a $+\infty$; allora anche k tende a $+\infty$, sicché si ha

$$\limsup_{n \rightarrow +\infty} \frac{a_j}{j} \leq \frac{a_n}{n}$$

e, di qui

$$\limsup_{n \rightarrow +\infty} \frac{a_j}{j} \leq \inf_{n \in \mathbf{N}} \frac{a_n}{n}.$$

Poiché, d'altro canto,

$$\liminf_{n \rightarrow +\infty} \frac{a_j}{j} \geq \inf_{n \in \mathbf{N}} \frac{a_n}{n},$$

segue l'asserto. □

La definizione di entropia di una trasformazione che conserva la misura si basa sul seguente

TEOREMA 2.7. *Siano T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e \mathcal{A} una sotto-tribù finita di \mathcal{F} ; allora esiste il limite*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right).$$

DIMOSTRAZIONE. Posto

$$a_n := H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right),$$

si controlla immediatamente, in virtù della (h) del Teorema 2.2 e della (b) del Teorema 2.3, che è verificata la (2.5). L'asserto è ora una conseguenza immediata del Lemma 2.1. □

DEFINIZIONE 2.4. Dati uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e una trasformazione T che conserva la misura si chiama *entropia di T rispetto alla sotto-tribù finita \mathcal{A} di \mathcal{F}* il limite

$$h(T, \mathcal{A}) := \lim_{n \rightarrow +\infty} \frac{1}{n} H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right).$$

Si dice, invece, *entropia di T*

$$h(T) := \sup \{h(T, \mathcal{A}) : \mathcal{A} \in \mathcal{P}\}.$$

La dimostrazione della seguente proposizione è immediata.

LEMMA 2.2. *Valgono le proprietà*

- (a) *per ogni sotto-tribù finita \mathcal{A} di \mathcal{F} , $h(T, \mathcal{A}) \geq 0$;*
- (b) *$h(T) \geq 0$.*

L'introduzione dell'entropia nella Teoria ergodica è motivata dal seguente

TEOREMA 2.8. *L'entropia è un invariante per coniugio (e, quindi, per isomorfismo).*

DIMOSTRAZIONE. Siano $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ due spazi di probabilità e T_1 e T_2 due trasformazioni, ciascuna delle quali conserva la misura sullo spazio con lo stesso indice. Si supponga che T_1 e T_2 siano coniugate. Esiste allora un isomorfismo $\Phi : (\tilde{\mathcal{F}}_2, \tilde{\mu}_2) \longrightarrow (\tilde{\mathcal{F}}_1, \tilde{\mu}_1)$ di algebre di misura tale che $\Phi \tilde{T}_2^{-1} = \tilde{T}_1^{-1} \Phi$. Sia \mathcal{A}_2 una sotto-tribú finita di \mathcal{F}_2 e sia $\pi_2 := \{A_1, A_2, \dots, A_s\}$ la partizione che \mathcal{A}_2 genera. Si scelgano insiemi B_1, B_2, \dots, B_s in modo che $\pi_1 := \{B_1, B_2, \dots, B_s\}$ sia una partizione misurabile di Ω_1 e che risulti $\tilde{B}_j = \Phi(\tilde{A}_j)$ ($j = 1, 2, \dots, s$). Sia \mathcal{A}_1 la sotto-tribú finita generata dalla partizione π_1 . Ora

$$\mu_1 \left(\bigcap_{j=0}^{n-1} T_1^{-j} B_{k(j)} \right) = \mu_2 \left(\bigcap_{j=0}^{n-1} T_2^{-j} A_{k(j)} \right)$$

ove $k(j) \in \{1, 1, \dots, s\}$, poiché

$$\begin{aligned} \Phi \left[\bigcap_{j=0}^{n-1} \left(T_2^{-j} A_{k(j)} \right) \right] &= \Phi \left(\bigcap_{j=0}^{n-1} \tilde{T}_2^{-j} \tilde{A}_{k(j)} \right) \\ &= \bigcap_{j=0}^{n-1} \Phi \tilde{T}_2^{-j} \tilde{A}_{k(j)} = \bigcap_{j=0}^{n-1} \tilde{T}_1^{-j} \Phi \left(\tilde{A}_{k(j)} \right) \\ &= \bigcap_{j=0}^{n-1} \tilde{T}_1^{-j} \tilde{B}_{k(j)} = \bigcap_{j=0}^{n-1} \left(T_1^{-j} B_{k(j)} \right) \end{aligned}$$

Di conseguenza, si ha

$$H \left(\bigvee_{j=0}^{n-1} T_1^{-j} \mathcal{A}_1 \right) = H \left(\bigvee_{j=0}^{n-1} T_2^{-j} \mathcal{A}_2 \right)$$

che implica $h(T_1, \mathcal{A}_1) = h(T_2, \mathcal{A}_2)$, che, a sua volta implica $h(T_1) \geq h(T_2)$. Per simmetria, si ottiene $h(T_1) \leq h(T_2)$ e, quindi $h(T_1) = h(T_2)$. \square

3. Le proprietà dell'entropia di una trasformazione

TEOREMA 3.1. *Siano $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità, T una trasformazione che conserva la misura e \mathcal{A} e \mathcal{B} sotto-tribú finite di \mathcal{F} . Allora*

- (a) $0 \leq h(T, \mathcal{A}) \leq H(\mathcal{A})$;
- (b) $h(T, \mathcal{A} \vee \mathcal{B}) \leq h(T, \mathcal{A}) + h(T, \mathcal{B})$;
- (c) $h(T, \mathcal{A}) \leq h(T, \mathcal{B})$ se $\mathcal{A} \subset \mathcal{B}$;
- (d) $h(T, \mathcal{A}) \leq h(T, \mathcal{B}) + H(\mathcal{A} | \mathcal{B})$.

DIMOSTRAZIONE. (a) Si ha

$$\frac{1}{n} H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \leq \frac{1}{n} \sum_{j=0}^{n-1} H(T^{-j} \mathcal{A}) = \frac{1}{n} \sum_{j=0}^{n-1} H(\mathcal{A}) = H(\mathcal{A}).$$

Per la (b) è

$$\begin{aligned} H \left[\bigvee_{j=0}^{n-1} T^{-j} (\mathcal{A} \vee \mathcal{B}) \right] &= H \left[\bigvee_{j=0}^{n-1} (T^{-j} \mathcal{A} \vee T^{-j} \mathcal{B}) \right] \\ &\leq H \left[\left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \vee \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{B} \right) \right] \\ &\leq H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) + H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{B} \right). \end{aligned}$$

(c) Dall'inclusione $\mathcal{A} \subset \mathcal{B}$ scende $\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \subset \bigvee_{j=0}^{n-1} T^{-j} \mathcal{B}$ di modo che il risultato segue dalla (d) del teorema 2.2.

Infine, per quanto riguarda la (d), si ha

$$\begin{aligned} H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) &\leq H \left[\left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \vee \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{B} \right) \right] \\ &\leq H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{B} \right) + H \left[\left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \middle| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{B} \right) \right]; \end{aligned}$$

ma, per la subaddittività condizionata risulta

$$\begin{aligned} H \left[\left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \middle| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{B} \right) \right] &\leq \sum_{j=1}^{n-1} H \left[T^{-j} \mathcal{A} \middle| \left(\bigvee_{k=0}^{n-1} T^{-k} \mathcal{B} \right) \right] \\ &\leq \sum_{j=1}^{n-1} H [T^{-j} \mathcal{A} | T^{-j} \mathcal{B}] = \sum_{j=1}^{n-1} H(\mathcal{A} | \mathcal{B}) = n H(\mathcal{A} | \mathcal{B}), \end{aligned}$$

onde

$$H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \leq H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{B} \right) + n H(\mathcal{A} | \mathcal{B});$$

di qui scende l'asserto. □

TEOREMA 3.2. *Se, nelle ipotesi del Teorema precedente, la trasformazione T è anche invertibile, si ha, per ogni $m \in \mathbf{N}$,*

$$h(T, \mathcal{A}) = h \left(T, \bigvee_{j=-m}^m T_j A \right)$$

DIMOSTRAZIONE. Dalla definizione di h segue

$$\begin{aligned} h \left(T, \bigvee_{j=-m}^m T^{-j} A \right) &= \lim_{k \rightarrow +\infty} \frac{1}{k} H \left[\bigvee_{j=0}^{k-1} T^{-j} \left(\bigvee_{i=-m}^m T^{-i} A \right) \right] \\ &= \lim_{k \rightarrow +\infty} \frac{1}{k} H \left(\bigvee_{j=-m}^{m+k-1} T^{-j} A \right). \end{aligned}$$

D'altra parte, risulta

$$\begin{aligned} H \left(\bigvee_{j=-m}^{m+k-1} T^{-j} A \right) &= H \left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A \right) \vee \left(\bigvee_{j=m}^{m+k-1} T^{-j} A \right) \right] \\ &= H \left(\bigvee_{j=m}^{m+k-1} T^{-j} A \right) + H \left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A \right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A \right) \right] \\ &= H \left[T^{-m} \left(\bigvee_{j=0}^{k-1} T^{-j} A \right) \right] + H \left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A \right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A \right) \right] \\ &= H \left(\bigvee_{j=0}^{k-1} T^{-j} A \right) + H \left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A \right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A \right) \right]. \end{aligned}$$

Per concludere la dimostrazione basterà far vedere che

$$\lim_{k \rightarrow +\infty} \frac{1}{k} H \left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A \right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A \right) \right] = 0.$$

La (f) del Teorema 2.2 dà

$$0 \leq \frac{1}{k} H \left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A \right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A \right) \right] \leq \frac{1}{k} H \left(\bigvee_{j=-m}^{m-1} T^{-j} A \right),$$

ma quest'ultima espressione tende a zero la tendere di k a $+\infty$. \square

Vi è una maniera alternativa di definire l'entropia.

TEOREMA 3.3. *Siano T una trasformazione che conserva la misura sullo spazio di probabilità (Ω, \mathcal{F}, P) e \mathcal{A} una sotto-tribù finita di \mathcal{F} ; allora*

$$h(T, \mathcal{A}) = \lim_{n \rightarrow +\infty} H \left[\mathcal{A} \left| \bigvee_{j=1}^n T^{-j} \mathcal{A} \right. \right]. \quad (3.1)$$

DIMOSTRAZIONE. Poiché si ha, ovviamente,

$$\bigvee_{j=1}^n T^{-j} \mathcal{A} \subset \bigvee_{j=1}^{n+1} T^{-j} \mathcal{A},$$

scende dalla (e) del Teorema 2.2 che

$$H \left[\mathcal{A} \left| \bigvee_{j=1}^n T^{-j} \mathcal{A} \right. \right] \geq H \left[\mathcal{A} \left| \bigvee_{j=1}^{n+1} T^{-j} \mathcal{A} \right. \right]$$

di modo che il limite che compare nella (3.1) esiste finito.

Si mostrerà per induzione che

$$H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) = H(\mathcal{A}) + \sum_{j=1}^{n-1} H \left[\mathcal{A} \left| \bigvee_{i=1}^j T^{-i} \mathcal{A} \right. \right]. \quad (3.2)$$

Questa è senz'altro vera per $n = 2$, giacché, in questo caso coincide con la (b) del Teorema 2.2. Si supponga, perciò la (3.2) vera per $r \in \mathbf{N}$ con $n \geq 2$; allora, in virtù dell'ipotesi d'induzione, si ha, per $r + 1$,

$$\begin{aligned} H \left(\bigvee_{j=0}^r T^{-j} \mathcal{A} \right) &= H \left[\left(\bigvee_{j=1}^{r-1} T^{-j} \mathcal{A} \right) \vee \mathcal{A} \right] \\ &= H \left(\bigvee_{j=1}^r T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \left| \bigvee_{i=1}^r T^{-i} \mathcal{A} \right. \right] \\ &= H \left(\bigvee_{j=0}^{r-1} T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \left| \bigvee_{i=1}^r T^{-i} \mathcal{A} \right. \right] \\ &= H(\mathcal{A}) + \sum_{j=1}^{r-1} H \left[\mathcal{A} \left| \bigvee_{i=1}^j T^{-i} \mathcal{A} \right. \right] + H \left[\mathcal{A} \left| \bigvee_{i=1}^r T^{-i} \mathcal{A} \right. \right] \\ &= H(\mathcal{A}) + \sum_{j=1}^r H \left[\mathcal{A} \left| \bigvee_{i=1}^j T^{-i} \mathcal{A} \right. \right], \end{aligned}$$

sicché la (3.2) è dimostrata. Si divida ora la (3.2) per n e se ne consideri il limite per $n \rightarrow +\infty$. Poiché il limite nel senso di CESÀRO coincide con il

limite ordinario per una successione convergente, il secondo membro tende al secondo membro della (3.1), mentre il primo membro tende, per definizione, a $h(T, \mathcal{A})$. \square

COROLLARIO 3.1. *La successione*

$$\left\{ \frac{1}{n} H \left(\bigvee_{j=0}^r T^{-j} \mathcal{A} \right) : n \in \mathbf{N} \right\}$$

tende decrescendo a $h(T, \mathcal{A})$.

DIMOSTRAZIONE. Segue dalla (3.2) che

$$\begin{aligned} H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) &= H(\mathcal{A}) + \sum_{j=1}^{n-1} H \left[\mathcal{A} \left| \bigvee_{i=1}^j T^{-i} \mathcal{A} \right. \right] \\ &\geq H \left[\mathcal{A} \left| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \right. \right] + (n-1) H \left[\mathcal{A} \left| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \right. \right] \\ &= n H \left[\mathcal{A} \left| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \right. \right]. \end{aligned}$$

Pertanto

$$n \left\{ H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \left| \bigvee_{j=1}^n T^{-j} \mathcal{A} \right. \right] \right\} \leq (n+1) H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right)$$

e, quindi,

$$\begin{aligned} n H \left(\bigvee_{j=0}^n T^{-j} \mathcal{A} \right) &= n \left\{ H \left(\bigvee_{j=1}^n T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \left| \bigvee_{j=1}^n T^{-j} \mathcal{A} \right. \right] \right\} \\ &= n \left\{ H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \left| \bigvee_{j=1}^n T^{-j} \mathcal{A} \right. \right] \right\} \\ &\leq (n+1) H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right), \end{aligned}$$

onde

$$\frac{1}{n} H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \geq \frac{1}{n+1} H \left(\bigvee_{j=0}^n T^{-j} \mathcal{A} \right)$$

che conclude la dimostrazione. \square

4. Il teorema di Kolmogorov–Sinai

TEOREMA 4.1. *Siano m un numero naturale e T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$, risulta*

$$h(T^m) = m h(T).$$

Se, inoltre, la trasformazione T è anche invertibile, allora, per ogni $m \in \mathbf{Z}$, è

$$h(T^m) = |m| h(T).$$

DIMOSTRAZIONE. Si dimostrerà dapprima che, se \mathcal{A} è una sotto-tribú finita di \mathcal{F} , è

$$h\left(T^m, \bigvee_{j=0}^{m-1} T^{-j} \mathcal{A}\right) = m h(T, \mathcal{A}). \quad (4.1)$$

Infatti

$$\begin{aligned} \lim_{k \rightarrow +\infty} \frac{1}{k} H \left[\bigvee_{i=0}^{k-1} T^{-mi} \left(\bigvee_{j=0}^{m-1} T^{-j} \mathcal{A} \right) \right] &= \lim_{k \rightarrow +\infty} \frac{m}{mk} H \left(\bigvee_{j=0}^{km-1} T^{-j} \mathcal{A} \right) \\ &= m h(T, \mathcal{A}). \end{aligned}$$

Scende dalla (4.1) che

$$\begin{aligned} m h(T) &= m \sup \{ h(T, \mathcal{A}) : \mathcal{A} \in \mathcal{P} \} = \sup \left\{ h \left(T^m, \bigvee_{j=0}^{m-1} T^{-j} \mathcal{A} \right) : \mathcal{A} \in \mathcal{P} \right\} \\ &\leq \sup \{ h(T^m, \mathcal{B}) : \mathcal{B} \in \mathcal{P} \} = h(T^m), \end{aligned}$$

vale a dire

$$m h(T) \leq h(T^m).$$

D'altra parte, per la (c) del Teorema 3.1, è

$$h(T^m, \mathcal{A}) \leq h \left(T^m, \bigvee_{j=0}^{m-1} T^{-j} \mathcal{A} \right) = m h(T, \mathcal{A}),$$

onde

$$h(T^m) \leq m h(T)$$

che dimostra l'asserto.

Si supponga ora che T sia anche invertibile. In virtù di quanto precede, basta dimostrare che $h(T^{-1}) = h(T)$, e, a tal fine basta dimostrare che,

per ogni sotto-tribú finita \mathcal{A} di \mathcal{F} , si ha $h(T, \mathcal{A}) = h(T^{-1}, \mathcal{A})$. Ora, poiché T è invertibile, il Teorema 2.3 (b) assicura che sia

$$H\left(\bigvee_{j=0}^{n-1} T^j \mathcal{A}\right) = H\left[T^{-(n-1)}\left(\bigvee_{j=0}^{n-1} T^j \mathcal{A}\right)\right] = H\left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}\right),$$

che conclude la dimostrazione. \square

Siamo ora in grado di provare il teorema d'approssimazione che servirà per stabilire l'importante teorema di KOLMOGOROV-SINAI

TEOREMA 4.2. *Siano $(\Omega, \mathcal{F}, \mu)$ uno spazio di probabilità, \mathcal{A} un'algebra di \mathcal{F} tale che $\mathcal{F} \stackrel{\circ}{=} \mathcal{F}(\mathcal{A})$, e \mathcal{B} una sotto-tribú finita di \mathcal{F} . Allora, per ogni $\varepsilon > 0$, esiste una sotto-tribú finita \mathcal{C}_ε di \mathcal{A} tale che*

$$H(\mathcal{C}_\varepsilon | \mathcal{B}) + H(\mathcal{B} | \mathcal{C}_\varepsilon) < \varepsilon. \quad (4.2)$$

DIMOSTRAZIONE. Non è restrittivo supporre che tutti gli insiemi della partizione generata da \mathcal{B} ,

$$\pi(\mathcal{B}) = \{B_1, B_2, \dots, B_r\}$$

abbiano probabilità strettamente positiva, $\mu(B_j) > 0$. Infatti, se fosse $\mu(B_j) > 0$ per $j = 1, 2, \dots, s$, mentre $\mu(B_j) = 0$ per $j = s+1, s+2, \dots, r$, per la sotto-tribú \mathcal{B}' generata dalla partizione $\{B_1, B_2, \dots, B_{s-1}, \cup_{j=s}^r B_j\}$ si avrebbe

$$H(\mathcal{B}' | \mathcal{C}_\varepsilon) = H(\mathcal{B} | \mathcal{C}_\varepsilon) \quad \text{e} \quad H(\mathcal{C}_\varepsilon | \mathcal{B}') = H(\mathcal{C}_\varepsilon | \mathcal{B}).$$

In virtù della continuità della funzione $\varphi : [0, 1] \rightarrow \mathbf{R}_+$ definita da $\varphi(x) := -x \log x$ se $x \in]0, 1]$ e da $\varphi(0) := 0$, esiste $\delta_0 \in]0, 1[$ tale che

$$\varphi(x) < \frac{\varepsilon}{2r}$$

se x appartiene a $[0, \delta_0] \cup [1 - \delta_0, 1]$. Il primo passo della dimostrazione consiste nel dimostrare che se una partizione $\pi = \{C_1, C_2, \dots, C_r\}$ soddisfa alle diseuguaglianze

$$\mu(B_j \Delta C_j) \leq \delta := \min \left\{ \delta_0 \frac{\mu(B_i)}{2} : i = 1, 2, \dots, r \right\} \quad (j = 1, 2, \dots, r), \quad (4.3)$$

allora risulta, se $\mathcal{C}_\varepsilon = \mathcal{F}(\pi)$,

$$H(\mathcal{B} | \mathcal{C}_\varepsilon) < \frac{\varepsilon}{2}. \quad (4.4)$$

Infatti, se la partizione π soddisfa alla (4.3), si ha

$$\begin{aligned} \mu(B_j) &= \mu(B_j \setminus C_j) + \mu(B_j \cap C_j) \leq \mu(B_j \Delta C_j) + \mu(C_j) \\ &\leq \delta + \mu(C_j) < \frac{\mu(B_j)}{2} + \mu(C_j), \end{aligned}$$

relazione che dà $\mu(B_j) < 2\mu(C_j)$ e, quindi,

$$\mu(C_j) - \mu(B_j \cap C_j) \leq \mu(B_j \Delta C_j) \leq \delta < \delta_0 \mu(C_j).$$

Perciò

$$\mu(B_j | C_j) = \frac{\mu(B_j \cap C_j)}{\mu(C_j)} \geq 1 - \delta_0 \quad (j = 1, 2, \dots, r),$$

onde

$$\begin{aligned} H(\mathcal{B} | \mathcal{C}_\varepsilon) &= - \sum_{j=1}^r \mu(C_j) \sum_{k=1}^r \mu(B_j | C_j) \log \mu(B_j | C_j) \\ &< \sum_{j=1}^r \mu(C_j) \sum_{k=1}^r \frac{\varepsilon}{2r}, \end{aligned}$$

ciò che dimostra la (4.4). La (4.4) continua a valere se la (4.3) è sostituita dalla disequaglianza più forte

$$\mu(B_j \Delta C_j) \leq \frac{\delta_0}{4} \min \{ \mu(B_i) : i = 1, 2, \dots, r \} \quad (j = 1, 2, \dots, r). \quad (4.5)$$

Ma, se vale la (4.5), risulta

$$\mu(B_j \Delta C_j) < \frac{\mu(B_j)}{4}$$

e, quindi,

$$\mu(B_j) = \mu(B_j \setminus C_j) + \mu(B_j \cap C_j) \leq \mu(B_j \Delta C_j) + \mu(C_j) < \frac{\mu(B_j)}{4} + \mu(C_j),$$

onde

$$\mu(C_j) > \frac{3}{4} \mu(B_j).$$

Infine da quest'ultima disequaglianza e dalla (4.5) si ha

$$\begin{aligned} \mu(B_j \Delta C_j) &< \frac{\delta_0}{4} \frac{4}{3} \min \{ \mu(C_j) : j = 1, 2, \dots, r \} \\ &< \frac{\delta_0}{2} \min \{ \mu(C_j) : j = 1, 2, \dots, r \} \end{aligned}$$

e, quindi, come sopra segue che

$$H(\mathcal{C}_\varepsilon | \mathcal{B}) < \frac{\varepsilon}{2}.$$

Per concludere la dimostrazione basterà far vedere che si può scegliere C_j in \mathcal{A} in modo tale che sia $\mu(C_j) > 0$ e

$$\mu(B_j \Delta C_j) \leq \frac{\delta_0}{4} \min \{ \mu(B_j) : j = 1, 2, \dots, r \} =: \tau \quad (4.6)$$

Si scelga $\lambda > 0$ in modo che sia $2\lambda(r-1)\{1+r(r-1)\} < \tau$. Per ogni indice j si scelga un insieme A_j in \mathcal{A} in modo che sia $\mu(A_j \Delta B_j) < \lambda$. Ora se $k \neq j$, è

$$A_j \cap A_k \subset (A_j \Delta B_j) \cup (A_k \Delta B_k)$$

onde $\mu(A_j \cap A_k) < 2\lambda$. Posto

$$N := \cup_{k \neq j} (A_j \cap A_k),$$

si ha

$$\mu(N) < 2\lambda r(r-1).$$

Sia ponga, ora, $C_j := B_j \setminus N$ per $j = 1, 2, \dots, r-1$ e $C_r := (\cup_{j=1}^{r-1} C_j)^c$, sicché $\pi' := \{C_1, C_2, \dots, C_r\}$ è una partizione finita di Ω in insiemidi \mathcal{A} . Per $j = 1, 2, \dots, r-1$ è

$$B_j \Delta C_j \subset (A_j \Delta B_j) \cup N,$$

onde

$$\mu(B_j \Delta C_j) < 2\lambda[1 + r(r-1)] < \tau,$$

mentre

$$B_r \Delta C_r \subset \cup_{j=1}^{r-1} (B_j \Delta C_j),$$

onde

$$\mu(B_r \Delta C_r) < 2\lambda(r-1)[1 + r(r-1)] < \tau.$$

Poiché la (4.6) è verificata, la dimostrazione è conclusa. \square

Se $\{\mathcal{A}_n : n \in \mathbf{N}\}$ è una successione di algebre contenute in \mathcal{F} si indicherà con $\bigvee_{n \in \mathbf{N}} \mathcal{A}_n$ l'algebra generata dalla successione data.

COROLLARIO 4.1. *Sia $\{\mathcal{A}_n : n \in \mathbf{N}\}$ una successione crescente, $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ di algebre finite contenute in \mathcal{F} , tale che $\mathcal{B} \subset \bigvee_{n \in \mathbf{N}} \mathcal{A}_n$; allora*

$$\lim_{n \rightarrow +\infty} H(\mathcal{B} | \mathcal{A}_n) = 0.$$

DIMOSTRAZIONE. Si ponga $\mathcal{A} := \bigvee_{n \in \mathbf{N}} \mathcal{A}_n$; allora, \mathcal{A} è un'algebra e, per ipotesi, è

$$\mathcal{B} \subset \mathcal{F}(\mathcal{A}).$$

Per il Teorema 4.2, assegnato arbitrariamente $\varepsilon > 0$, esiste una sotto-tribú finita \mathcal{C}_ε di \mathcal{A} tale che $H(\mathcal{B} | \mathcal{C}_\varepsilon) < \varepsilon$. Poiché \mathcal{C}_ε è un'algebra finita, esiste $n_0 \in \mathbf{N}$ tale che $\mathcal{C}_\varepsilon \subset \mathcal{A}_{n_0}$, onde, per $n \geq n_0$, si ha

$$H(\mathcal{B} | \mathcal{A}_n) \leq H(\mathcal{B} | \mathcal{A}_{n_0}) \leq H(\mathcal{B} | \mathcal{C}_\varepsilon) < \varepsilon,$$

cioè l'asserto. \square

Siamo ora pronti alla dimostrazione del fondamentale

TEOREMA 4.3 (Teorema di KOLMOGOROV-SINAI). *Siano T una trasformazione invertibile che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e \mathcal{A} una sotto-tribú finita di \mathcal{F} tale che*

$$\bigvee_{n \in \mathbf{Z}} T^n \mathcal{A} = \mathcal{F}.$$

Allora $h(T) = h(T, \mathcal{A})$.

DIMOSTRAZIONE. Basta dimostrare che, per ogni sotto-tribú finita \mathcal{B} di \mathcal{F} , si ha

$$h(T, \mathcal{B}) \leq h(T, \mathcal{A}). \quad (4.7)$$

Ora

$$\begin{aligned} h(T, \mathcal{B}) &\leq h\left(T, \bigvee_{j=-n}^n T^j \mathcal{A}\right) + H\left(\mathcal{B} \left| \bigvee_{j=-n}^n T^j \mathcal{A}\right.\right) \\ &= h(T, \mathcal{A}) + H\left(\mathcal{B} \left| \bigvee_{j=-n}^n T^j \mathcal{A}\right.\right), \end{aligned}$$

ove abbiamo fatto uso della (d) del Teorema 3.1 e, quindi, del Teorema 3.2. Posto

$$\mathcal{A}_n := \bigvee_{j=-n}^n T^j \mathcal{A},$$

risulta $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ e, per ipotesi, $\mathcal{B} \subset \bigvee_{n \in \mathbf{N}} \mathcal{A}_n$, sicché

$$\lim_{n \rightarrow +\infty} H(\mathcal{B} | \mathcal{A}_n) = 0$$

per il corollario 4.1. La (4.7) è così completamente dimostrata. \square

Se la trasformazione T non è necessariamente invertibile, vale il seguente analogo del Teorema di KOLMOGOROV-SINAI.

TEOREMA 4.4. *Siano T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ ed \mathcal{A} una sotto-tribú finita di \mathcal{F} tale che*

$$\bigvee_{n \in \mathbf{Z}_+} T^{-n} \mathcal{A} \overset{\circ}{=} \mathcal{F}.$$

Allora $h(T) = h(T, \mathcal{A})$.

DIMOSTRAZIONE. La dimostrazione è simile a quella del teorema di KOLMOGOROV-SINAI. Si ha, per ogni $n \in \mathbf{N}$,

$$h(T, \mathcal{A}) = h\left(T, \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}\right). \quad (4.8)$$

Infatti

$$h\left(T, \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}\right) = \lim_{n \rightarrow +\infty} \frac{1}{k} H\left[\bigvee_{s=0}^{k-1} T^{-s} \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}\right)\right],$$

ma

$$\begin{aligned}
 H \left[\bigvee_{s=0}^{k-1} T^{-s} \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \right] &= H \left(\bigvee_{j=0}^{n+k-2} T^{-j} \mathcal{A} \right) \\
 &= H \left[\left(\bigvee_{j=0}^{n-2} T^{-j} \mathcal{A} \right) \vee \left(\bigvee_{j=n-1}^{n+k-2} T^{-j} \mathcal{A} \right) \right] \\
 &= H \left(\bigvee_{j=n-1}^{n+k-2} T^{-j} \mathcal{A} \right) + H \left[\left(\bigvee_{j=0}^{n-2} T^{-j} \mathcal{A} \right) \middle| \left(\bigvee_{j=n-1}^{n+k-2} T^{-j} \mathcal{A} \right) \right] \\
 &= H \left(\bigvee_{j=0}^{k-1} T^{-j} \mathcal{A} \right) + H \left[\left(\bigvee_{j=0}^{n-2} T^{-j} \mathcal{A} \right) \middle| \left(\bigvee_{j=n-1}^{n+k-2} T^{-j} \mathcal{A} \right) \right].
 \end{aligned}$$

Siccome si ha

$$\frac{1}{k} H \left[\left(\bigvee_{j=0}^{n-2} T^{-j} \mathcal{A} \right) \middle| \left(\bigvee_{j=n-1}^{n+k-2} T^{-j} \mathcal{A} \right) \right] \leq \frac{1}{k} H \left(\bigvee_{j=0}^{n-2} T^{-j} \mathcal{A} \right) \xrightarrow{k \rightarrow +\infty} 0,$$

la (4.8) è dimostrata. Come nel Teorema di KOLMOGOROV-SINAI basta provare che, per ogni sotto-tribú finita \mathcal{B} di \mathcal{F} , vale la (4.7). In virtù della (4.8) si ha

$$\begin{aligned}
 h(T, \mathcal{B}) &\leq h \left(T, \bigvee_{j=0}^n T^{-j} \mathcal{A} \right) + H \left[\mathcal{B} \middle| \left(\bigvee_{j=0}^n T^{-j} \mathcal{A} \right) \right] \\
 &= h(T, \mathcal{A}) + H \left[\mathcal{B} \middle| \left(\bigvee_{j=0}^n T^{-j} \mathcal{A} \right) \right]
 \end{aligned}$$

e la tesi segue come nel teorema precedente. \square

COROLLARIO 4.2. *Sia T una trasformazione invertibile che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e sia \mathcal{A} una sotto-tribú finita di \mathcal{F} tale che*

$$\bigvee_{n \in \mathbf{Z}_+} T^{-n} \mathcal{A} \overset{\circ}{=} \mathcal{F}.$$

Allora $h(T) = 0$.

DIMOSTRAZIONE. Per i Teoremi 4.4 e 3.3 risulta

$$h(T) = h(T, \mathcal{A}) = \lim_{n \rightarrow +\infty} H \left[\mathcal{A} \middle| \bigvee_{j=1}^n T^{-j} \mathcal{A} \right].$$

Ma

$$\bigvee_{n \in \mathbf{N}} T^{-n} \mathcal{A} \stackrel{\circ}{=} T \mathcal{F} = \mathcal{F}.$$

Posto

$$\mathcal{A}_n := \bigvee_{j=1}^n T^{-j} \mathcal{A},$$

la successione $\{\mathcal{A}_n : n \in \mathbf{N}\}$ è crescente, $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ per ogni $n \in \mathbf{N}$ e si ha

$$\bigvee_{n \in \mathbf{N}} \mathcal{A}_n = \mathcal{F},$$

onde

$$\lim_{n \rightarrow +\infty} H(\mathcal{A} | \mathcal{A}_n) = 0,$$

e, quindi, $h(T) = 0$. □

TEOREMA 4.5. *Siano T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e \mathcal{A} una sotto-tribú finita di \mathcal{F} tale che $\mathcal{F}(\mathcal{A}) \stackrel{\circ}{=} \mathcal{F}$. Allora*

$$h(T) = \sup \{h(T, \mathcal{B}) : \mathcal{B} \in \mathcal{P}, \mathcal{B} \subset \mathcal{A}\}. \quad (4.9)$$

DIMOSTRAZIONE. Poiché l'entropia $h(T)$ maggiora il secondo membro della (4.9), per stabilire la (4.9) basta far vedere che

$$h(T) \leq \sup \{h(T, \mathcal{B}) : \mathcal{B} \in \mathcal{P}, \mathcal{B} \subset \mathcal{A}\}. \quad (4.10)$$

Fissato arbitrariamente $\varepsilon > 0$, sia \mathcal{C} una sotto-tribú finita di \mathcal{F} . Per il Teorema 4.2 esiste un'altra sotto-tribú finita di \mathcal{F} , sia essa \mathcal{B}_ε con $\mathcal{B}_\varepsilon \subset \mathcal{A}$, tale che $H(\mathcal{C} | \mathcal{B}_\varepsilon) < \varepsilon$, sicché

$$h(T, \mathcal{C}) \leq h(T, \mathcal{B}_\varepsilon) + H(\mathcal{C} | \mathcal{B}_\varepsilon) < h(T, \mathcal{B}_\varepsilon) + \varepsilon,$$

e, perciò

$$h(T, \mathcal{C}) \leq \varepsilon + \sup \{H(T, \mathcal{B}) : \mathcal{B} \in \mathcal{P}, \mathcal{B} \subset \mathcal{A}\},$$

dalla quale scende la (4.10), per l'arbitrarietà di ε . □

TEOREMA 4.6. *Siano T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e $\{\mathcal{A}_n : n \in \mathbf{N}\}$ una successione crescente, $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ per ogni $n \in \mathbf{N}$ di sotto-tribú finite di \mathcal{F} tale che*

$$\bigvee_{n \in \mathbf{N}} \mathcal{A}_n \stackrel{\circ}{=} \mathcal{F}.$$

Allora

$$h(T) = \lim_{n \rightarrow +\infty} h(T, \mathcal{A}_n). \quad (4.11)$$

DIMOSTRAZIONE. Si sa che la successione $\{h(T, \mathcal{A}_n) : n \in \mathbf{N}\}$ è anch'essa crescente, di modo che il limite che compare nella (4.11) esiste. Posto $\mathcal{A} := \cup_{n \in \mathbf{N}} \mathcal{A}_n$, risulta, per ipotesi,

$$\mathcal{F}(\mathcal{A}) \stackrel{\circ}{=} \mathcal{F}.$$

Se \mathcal{B} è una sotto-tribú finita di \mathcal{F} contenuta in \mathcal{A} , essa sarà contenuta in una sotto-tribú della successione, $\mathcal{B} \subset \mathcal{A}_{n_0}$ per un opportuno indice n_0 . Perciò

$$h(T, \mathcal{B}) \leq h(T, \mathcal{A}_{n_0})$$

e, quindi,

$$h(T) \leq \lim_{n \rightarrow +\infty} h(T, \mathcal{A}_n).$$

Poiché la diseguglianza inversa è ovvia, ciò stabilisce la (4.11). □

TEOREMA 4.7. *Siano T_1 e T_2 due trasformazioni che conservano la misura sugli spazi di probabilità $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ rispettivamente. Allora*

$$h(T_1 \times T_2) = h(T_1) + h(T_2). \tag{4.12}$$

DIMOSTRAZIONE. Se \mathcal{A}_1 e \mathcal{A}_2 sono sotto-tribú finite rispettivamente di \mathcal{F}_1 e di \mathcal{F}_2 , allora si consideri $\mathcal{A}_1 \times \mathcal{A}_2 := \{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$ che è una sotto-tribú finita di $\mathcal{F}_1 \times \mathcal{F}_2$ che genera la partizione

$$\pi(\mathcal{A}_1 \times \mathcal{A}_2) = \{A_1 \times A_2 : A_1 \in \pi(\mathcal{A}_1), A_2 \in \pi(\mathcal{A}_2)\}.$$

Sia \mathcal{T}_0 un'algebra formata dalle unioni finite di rettangoli "misurabili", vale a dire dalle unioni finite di rettangoli $A_1 \times A_2$ con $A_1 \in \mathcal{F}_1$ e $A_2 \in \mathcal{F}_2$. Ma allora $\mathcal{F}(\mathcal{T}_0) = \mathcal{F}_1 \times \mathcal{F}_2$ e, per il Teorema 4.5, è

$$h(T_1 \times T_2) = \sup \{h(T_1 \times T_2, \mathcal{C}) : \mathcal{C} \subset \mathcal{T}_0 \text{ sotto-tribú finita di } \mathcal{F}_1 \times \mathcal{F}_2\}$$

Se $\mathcal{C} \subset \mathcal{T}_0$ è finita, si possono trovare due sotto-tribú finite \mathcal{A}_1 di \mathcal{F}_1 e \mathcal{A}_2 di \mathcal{F}_2 tali che $\mathcal{C} \subset \mathcal{A}_1 \times \mathcal{A}_2$, onde

$$h(T_1 \times T_2) = \sup \{h(T_1 \times T_2, \mathcal{A}_1 \times \mathcal{A}_2)\}$$

ove l'estremo superiore si intende calcolato sopra le sotto-tribú finite \mathcal{A}_j di \mathcal{F}_j ($j = 1, 2$).

Ora, se $A_{1,r}$ e $A_{2,s}$ sono insiemi che appartengono rispettivamente alle partizioni

$$\pi \left(\bigvee_{j=0}^{n-1} T_1^{-j} \mathcal{A}_1 \right) \quad \text{e} \quad \pi \left(\bigvee_{j=0}^{n-1} T_2^{-j} \mathcal{A}_2 \right),$$

si ha

$$\begin{aligned}
H \left[\bigvee_{j=0}^{n-1} (T_1 \times T_2)^{-j} (\mathcal{A}_1 \times \mathcal{A}_2) \right] \\
&= H \left[\left(\bigvee_{j=0}^{n-1} T_1^{-j} \mathcal{A}_1 \right) \times \left(\bigvee_{j=0}^{n-1} T_2^{-j} \mathcal{A}_2 \right) \right] \\
&= - \sum_{r,s} (\mu_1 \otimes \mu_2) (A_{1,r} \times A_{2,s}) \log (\mu_1 \otimes \mu_2) (A_{1,r} \times A_{2,s}) \\
&= - \sum_{r,s} \mu_1 (A_{1,r}) \mu_2 (A_{2,s}) \log \mu_1 (A_{1,r}) \mu_2 (A_{2,s}) \\
&= - \sum_r \mu_1 (A_{1,r}) \log \mu_1 (A_{1,r}) - \sum_s \mu_2 (A_{2,s}) \log \mu_2 (A_{2,s}) \\
&= H \left(\bigvee_{j=0}^{n-1} T_1^{-j} \mathcal{A}_1 \right) + H \left(\bigvee_{j=0}^{n-1} T_2^{-j} \mathcal{A}_2 \right).
\end{aligned}$$

Perciò $h(T_1 \times T_2, \mathcal{A}_1 \times \mathcal{A}_2) = h(T_1, \mathcal{A}_1) + h(T_2, \mathcal{A}_2)$, da cui segue l'asserto. \square

5. Esempi di calcolo dell'entropia

ESEMPIO 5.1. L'identità I su Ω , $I\omega = \omega$ per ogni $\omega \in \Omega$, ha entropia nulla, $h(I) = 0$. Infatti è

$$h(T, \mathcal{A}) = \lim_{n \rightarrow +\infty} \frac{1}{n} H(\mathcal{A}) = 0.$$

Si osservi che, se T è periodica di periodo k , $T^k = I$, allora $h(T^k) = 0$. Infatti il Teorema 4.1 dà

$$k h(T) = h(T^k) = h(I) = 0.$$

ESEMPIO 5.2. Le rotazioni $T_a z = az$ ($a \in \mathcal{K}$) della circonferenza unitaria in \mathbf{C} hanno entropia nulla, $h(T_a) = 0$.

Se $\{a^n : n \in \mathbf{Z}\}$ non è densa in \mathcal{K} , vale a dire se a è una radice dell'unità, per la quale esiste, quindi un naturale k tale che $a^k = 1$, allora $T_a^k z = a^k z = z$, sicché $h(T_a) = 0$, dall'esempio precedente.

Si supponga, invece, che $\{a^n : n \in \mathbf{Z}\}$ sia densa in \mathcal{K} ; allora è densa l'orbita $\{a^{-n} : n \in \mathbf{N}\}$. Si consideri la partizione di \mathcal{K} data $\pi := \{A_1, A_2\}$ dove

$$A_1 := \{e^{2\pi i x} : x \in [0, 1/2[\} \quad \text{e} \quad A_2 := \{e^{2\pi i x} : x \in [1/2, 1[\}$$

sono rispettivamente le semicirconferenze unitarie superiore e inferiore. Per ogni $n \in \mathbf{N}$, $T_a^{-n}\pi$ è composto dalle due semicirconferenze con punti iniziali in a^{-n} e $-a^{-n}$. Poiché $\{a^{-n} : n \in \mathbf{N}\}$ è densa in \mathcal{K} , ogni semicirconferenza appartiene a $\bigvee_{n \in \mathbf{Z}_+} T_a^{-n}(\pi)$. Di conseguenza ogni arco della circonferenza unitaria appartiene a $\bigvee_{n \in \mathbf{Z}_+} T_a^{-n}(\pi)$. Perciò

$$\mathcal{B} = \bigvee_{n \in \mathbf{Z}_+} T_a^{-n}(\pi).$$

In virtù del Corollario 4.2 ciò implica $h(T_a) = 0$.

ESEMPIO 5.3. La traslazione bilatera $(p_0, p_1, \dots, p_{k-1})$ ha entropia data da

$$-\sum_{j=0}^{k-1} p_j \log p_j. \tag{5.1}$$

Se si pone

$$A_i := \{ \{x_n : n \in \mathbf{Z}\} : x_0 = i \},$$

allora $\pi := \{A_0, A_1, \dots, A_{k-1}\}$ costituisce una partizione di

$$\Omega = \{0, 1, \dots, k-1\}^{\mathbf{Z}}.$$

Sia $\mathcal{A} = \mathcal{A}(\pi)$ l'algebra generata da π . Per definizione della tribú prodotto \mathcal{B} , si ha

$$\bigvee_{n \in \mathbf{Z}} T^n \mathcal{A} = \mathcal{B}.$$

Il Teorema di KOLMOGOROV-SINAI 4.3 dà

$$h(T) = h(T, \mathcal{A}) = \lim_{n \rightarrow +\infty} \frac{1}{n} H(\mathcal{A} \vee T^{-1} \mathcal{A} \vee \dots \vee T^{-(n-1)} \mathcal{A}).$$

Gli elementi della partizione $\pi(\mathcal{A} \vee T^{-1} \mathcal{A} \vee \dots \vee T^{-(n-1)} \mathcal{A})$ hanno la forma

$$\begin{aligned} & A_{i(0)} \cap T^{-1} A_{i(1)} \cap \dots \cap T^{-(n-1)} A_{i(n-1)} \\ & = \{ \{x_n\} : x_0 = i(0), x_1 = i(1), \dots, x_{n-1} = i(n-1) \}, \end{aligned}$$

insieme che ha misura $p_{i(0)} p_{i(1)} \dots p_{i(n-1)}$. Perciò

$$\begin{aligned} & H(\mathcal{A} \vee T^{-1} \mathcal{A} \vee \dots \vee T^{-(n-1)} \mathcal{A}) \\ & = - \sum_{i(0), i(1), \dots, i(n-1)=0}^{n-1} p_{i(0)} p_{i(1)} \dots p_{i(n-1)} \log p_{i(0)} p_{i(1)} \dots p_{i(n-1)} \\ & = - \sum_{i(0), i(1), \dots, i(n-1)=0}^{n-1} p_{i(0)} p_{i(1)} \dots p_{i(n-1)} \times \\ & \quad \times \{ \log p_{i(0)} + \log p_{i(1)} + \dots + \log p_{i(n-1)} \} \\ & = -n \sum_{j=0}^{n-1} p_j \log p_j. \end{aligned}$$

Di qui scende la (5.1).

Si noti che la traslazione bilatera $(1/2, 1/2)$ ha entropia eguale a $\log 2$, mentre l'entropia della traslazione bilatera $(1/3, 1/3, 1/3)$ è $\log 3$, sicché queste due traslazioni non sono coniugate.

ESEMPIO 5.4. Lo stesso ragionamento dell'esempio precedente, con la sostituzione del Teorema 4.4 al Teorema 4.3, porta a stabilire che l'entropia della traslazione unilatera $(p_0, p_1, \dots, p_{k-1})$ è ancora data dalla

$$-\sum_{j=0}^{k-1} p_j \log p_j.$$

ESEMPIO 5.5. (entropia della traslazione markoviana bilatera $-(p, \Pi)$). Si ricordi la notazione dell'Esempio 3.8 e, come nell'esempio precedente, si ponga

$$A_i := \{ \{x_n : n \in \mathbf{Z}\} : x_0 = i \}.$$

Adesso l'elemento tipico della partizione $\pi (\mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{-(n-1)}\mathcal{A})$ ha misura eguale a $p_{i(0)} p_{i(0) i(1)} \dots p_{i(n-1) i(n)}$, sicché

$$\begin{aligned} & H(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{-(n-1)}\mathcal{A}) \\ &= - \sum_{i(0), i(1), \dots, i(n-1)=0}^{k-1} p_{i(0)} p_{i(0) i(1)} \dots p_{i(n-1) i(n)} \times \\ & \quad \times \log \{ p_{i(0)} p_{i(0) i(1)} \dots p_{i(n-1) i(n)} \} \\ &= - \sum_{i(0), i(1), \dots, i(n-1)=0}^{k-1} p_{i(0)} p_{i(0) i(1)} \dots p_{i(n-1) i(n)} \times \\ & \quad \times \{ \log p_{i(0)} + \log p_{i(0) i(1)} + \dots + \log p_{i(n-1) i(n)} \} \\ &= - \sum_{i(0), i(1), \dots, i(n-1)=0}^{k-1} p_{i(0)} p_{i(0) i(1)} \dots p_{i(n-1) i(n)} \log p_{i(0)} \\ & \quad - \sum_{i(0), i(1), \dots, i(n-1)=0}^{k-1} p_{i(0)} p_{i(0) i(1)} \dots p_{i(n-1) i(n)} \log p_{i(0) i(1)} - \dots \\ & \quad - \sum_{i(0), i(1), \dots, i(n-1)=0}^{k-1} p_{i(0)} p_{i(0) i(1)} \dots p_{i(n-1) i(n)} \log p_{i(n-1) i(n)} \\ &= - \sum_{i=0}^{k-1} p_i \log p_i - n \sum_{i=0}^{k-1} p_i \sum_{j=0}^{k-1} p_{ij} \log p_{ij}. \end{aligned}$$

Ne consegue che l'entropia della traslazione markoviana bilatera $-(p, \Pi)$ è eguale a

$$-\sum_{i=0}^{k-1} p_i \sum_{j=0}^{k-1} p_{ij} \log p_{ij}.$$

Nota bibliografica

Prima di chiudere questi appunti, mi sia consentito un ricordo personale. Molti anni fa, quando ero uno studente che scriveva la tesi di laurea, entrai in contatto con la teoria ergodica attraverso il mio relatore, PIETRO BOCCHIERI, che mi fece conoscere il bellissimo libro di KHINCHIN. Da allora, il mio interesse per tale teoria non è mai venuto meno. Subito dopo la laurea, intraprendendo gli studi di Dottorato, ebbi la fortuna di avere la guida di BRUNO FORTE, che aveva contribuito alla teoria e che, proprio allora, stava scrivendo i suoi importanti lavori sull'entropia. A questi Maestri vada qui la mia commossa gratitudine.

La letteratura sulla meccanica statistica è veramente vastissima; quella indicata di seguito è solo un minuscolo assaggio. Tra i molti eccellenti libri di testo, dedicati soprattutto agli aspetti fisici, segnalo quelli di HUANG [31], di WANNIER [71], di LANDAU & LIFCHITZ [42], di PATHRIA [54]. Più recentemente alcune monografie sono state dedicate agli aspetti matematiche della meccanica statistica; l'antesignano di queste è l'oramai classico libro di RUELLE [62]. Si vedano poi i libri di THOMPSON [69] e di MINLOS [47].

Tra i libri dedicati alla Teoria ergodica molti sono eccellenti. Citerò in primo luogo quello di HALMOS [26] e poi quelli di FRIEDMAN [20], di BROWN [11], di WALTERS [70], PARRY [53], CORNFELD *et al* [13], di PETERSEN [57], di MAÑÉ [45] e di NADKARMI [49]. Una buona introduzione non strettamente tecnica, ma culturalmente impegnativa è costituita dal lungo articolo di MACKAY [44], che tratta anche dei legami con l'analisi armonica, argomento che nelle lezioni è stato soltanto sfiorato in un paio di esempi, ma non citato esplicitamente. Esistono poi libri dedicati ad aspetti particolari della Teoria ergodica, come quello di SHIELDS [67] sulla dimostrazione dei teoremi di isomorfismo di ORNSTEIN e quello di ORNSTEIN stesso [52].

1.1 Per i richiami di meccanica analitica, ci si può sempre rifare al classico testo di LEVI-CIVITA & AMALDI [43, Capitolo X], oppure a [41]. Per la dimostrazione del teorema di LIOUVILLE si può consultare il libro di

KHINCHIN [37]. Ai problemi ergodici della meccanica classica è dedicato il libro di ARNOLD & AVEZ [5]; per tutto l'approccio alla meccanica statistica basato sulla teoria ergodica si potrà citare, innanzi tutto, il già citato libro di KHINCHIN [37] e quelli di FARQUHAR [15] e JANCEL [32]. Qualche fisico sostiene anche che l'approccio alla meccanica statistica attraverso la teoria ergodica è oramai "fuori moda", *old fashioned*; si veda per esempio l'articolo di PENROSE [56].

1.3 Anche i processi stazionari a tempo continuo derivano da un semigruppato di trasformazioni che conservano la misura; si veda l'inizio del Capitolo XI del libro di DOOB [14].

1.6 Si noti che l'operatore U_T definito da $U_T f := f \circ T$, dove, per il momento, si può supporre che U_T agisca su $L^\infty := L^\infty(\Omega, \mathcal{F}, \mu)$ è un operatore di MARKOV, nel senso che esso è lineare, positivo, $U_T f \geq 0$ se $f \geq 0$, lascia invarianti le funzioni costanti, $U_T 1 = 1$ e lascia invariata la speranza, $E(U_T f) = E(f)$. Inoltre, per operatori di MARKOV U_T indotti da una trasformazione T che conserva la misura, vale la relazione

$$U_T(f \cdot g) = (U_T f) (U_T g).$$

Per gli operatori di MARKOV in generale si vedano [10, 51].

1.7–1.8 Le date di pubblicazione dei Teoremi di BIRKHOFF e di VON NEUMANN sono rispettivamente il 1931 e il 1932 ([8, 50]), ma il Teorema di VON NEUMANN di convergenza in L^2 precede quello di convergenza quasi certa. La dimostrazione del Teorema di BIRKHOFF che ho dato qui è dovuta a RIESZ [59].

La prima versione del teorema ergodico massimale è dovuta a WIENER [73] e a YOSIDA & KAKUTANI [75]. In seguito E. HOPF [30] ne diede la versione per operatori che appare qui; la dimostrazione che ne dò, e che è oramai quella usuale, è dovuta a GARSIA [21].

1.8. Un teorema di RÉNYI ([58]) mostra che una trasformazione che conserva la misura su $(\Omega, \mathcal{F}, \mu)$ è fortemente mescolante se, e solo se, per ogni insieme A di \mathcal{F} , si ha

$$\lim_{n \rightarrow +\infty} \mu(T^{-n} A \cap A) = \mu^2(A),$$

e che essa è debolmente mescolante se, e solo se, per ogni insieme A di \mathcal{F} , si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=1}^n |\mu(T^{-j} A \cap A) - \mu^2(A)| = 0.$$

1.9. Il Teorema 9.4 è dovuto a KOOPMAN e VON NEUMANN [40]. Nel Corollario 9.1 si è provato che per ogni coppia A e B di insiemi misurabili esiste un insieme $J(A, B)$ di densità nulla per il quale vale

$$\lim_{\substack{n \rightarrow +\infty \\ n \notin J(A, B)}} \mu(T^{-n} A \cap B) = \mu(A) \mu(B).$$

Si può dimostrare che in effetti esiste un *unico* insieme J di densità nulle tale che sia

$$\lim_{\substack{n \rightarrow +\infty \\ n \notin J}} \mu(T^{-n} A \cap B) = \mu(A) \mu(B),$$

per ogni coppia A e B di insiemi misurabili.

1.12 Il primo teorema di categoria è in realtà posteriore al secondo; il primo teorema fu dimostrato da ROKHLIN [60], mentre il secondo è dovuto a HALMOS [24]. In [4] si mostra come ottenere simultaneamente i due teoremi di categoria.

1.15 Per maggiori informazioni sulle misure invarianti si veda il libro di HALMOS [26].

2.1 Per le algebre di BOOLE si veda [29].

3.2 Come si è già detto, l'introduzione dell'entropia è dovuto a KOLMOGOROV. Si veda a tal proposito [39] e la discussione di SINAI in [68] alle pp. 247–250. Si vedano anche [28] e [7].

Il concetto di entropia viene dalla meccanica statistica, ma il suo uso più “moderno” avvenne nella Teoria dell'Informazione per la quale si veda, ad esempio [6]. Si deve riflettere sul fatto che nella teoria dell'informazione, sono state introdotte molte diverse rappresentazioni dell'entropia, si veda, all'uopo, [2]; questo pone immediatamente il problema di quale sia l'entropia “corretta” da adottare. Per tale problema si veda [72]. L'adozione dell'entropia (2.1), detta *entropia di SHANNON*, è giustificata adeguatamente, in

un ambito che si richiama esplicitamente alla meccanica statistica dai lavori di FORTE [18, 3].

Incidentalmente, vale la pena di riferire che esiste un approccio alla meccanica statistica attraverso la teoria dell'informazione; si vedano, tal proposito, i lavori pionieristici di JAYNES [33, 34] e poi [17, 35, 65, 19, 66].

3.4 Del Teorema 4.6 esiste un'altra dimostrazione basata sui teoremi di convergenza delle martingale; la si può leggere, per esempio, in [57, p. 241].

Bibliografia

- [1] J. Aaronson, *An introduction to infinite ergodic theory*, Amer. Math. Soc., Providence, RI; (1997) Mathematical Surveys and Monographs 50.
- [2] J. Aczél, Z. Daróczy, *On measures of information and their characterization*, Academic Press, New York – San Francisco – London, 1975.
- [3] J. Aczél, B. Forte, C.T. Ng, Why the Shannon and Hartley entropies are 'natural', *Adv. Appl. Probab.* **6**, 131–146 (1974).
- [4] S. Alpern, Approximation to and by measure preserving homeomorphisms, *J. London Math. Soc.* **18**, 305–315 (1978).
- [5] V.I. Arnold, A. Avez, *Ergodic problems of classical mechanics*, Benjamin, New York (1968).
- [6] R.B. Ash, *Information theory*, Wiley–Interscience, New York, 1965; ristampa, Dover, New York, 1990.
- [7] P. Billingsley, *Ergodic theory and information*, Wiley, New York, 1965; ristampa, Krieger, New York, 1978.
- [8] G.D. Birkhoff, Proof of the ergodic theorem, *Proc. Nat. Acad. Sci. USA* **17**, 6546–660 (1931).
- [9] F. Blume, *Ergodic theory*, in E. Pap, ed., *Handbook of measure theory*, vol. II, pp. 1185–1295, Elsevier, Amsterdam (2002).
- [10] J.R. Brown, Approximation theory for Markov operators, *Pacific J. Math.*, **16**, 13–23 (1965).
- [11] J.R. Brown, *Ergodic theory and topological dynamics*, Academic Press, New York – San Francisco – London (1976).
- [12] S. Chapman, T.G. Cowling, *The mathematical theory of non-uniform gases*, Cambridge University Press, 1970 (3rd ed.).
- [13] I.P. Cornfeld, S.V. Fomin, Ya.G. Sinai, *Ergodic theory*, Springer, New York (1982).
- [14] J.L. Doob, *Stochastic processes*, New York: Wiley, 1953.
- [15] I.E. Farquhar, *Ergodic theory in statistical mechanics*, Wiley, London–New York, 1964.
- [16] E. Fermi, *Termodinamica*, Boringhieri, Torino, 1958.
- [17] B. Forte, Il principio di minima certezza e le sue applicazioni alla meccanica statistica, *Ann. Univ. Ferrara*, Sez. VII, **12**, 7–17 (1966).
- [18] B. Forte, Why Shannon entropy, *Symp. Math.* **18**, (1973).
- [19] B. Forte, C. Sempì, Maximizing conditional entropies: a derivation of quantum statistics, *Rend. Mat. (6)* **9**, 551–566 (1976).
- [20] N.A. Friedman, *Introduction to ergodic theory*, Van Nostrand Reinhold, New York (1970).
- [21] A.M. Garsia, A simple proof of Hopf's maximal ergodic theorem, *J. Math. Mech.*, **14**, 381–382 (1965).
- [22] J.W. Gibbs, *Elementary principles in statistical mechanics*, Yale University Press, 1902; ristampato da Dover, New York, 1960.
- [23] D. ter Haar, Foundations of statistical physics, *Rev. Mod. Phys.* **27**, 289–338 (1955).

- [24] P.R. Halmos, In general a measure-preserving transformation is mixing, *Ann. Math.* (2) **45**, 786–792 (1944); ristampato in [64, pp. 68–74].
- [25] P.R. Halmos, *Measure theory*, Van Nostrand Reinhold, New York (1950).
- [26] P.R. Halmos, *Lectures on ergodic theory*, The Mathematical Society of Japan, Tokyo (1956).
- [27] P.R. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity*, Chelsea, New York, 1957 (2nd ed.).
- [28] P.R. Halmos, *Entropy in ergodic theory*, (1959), in *P.R. Halmos, Selecta. Expository writing*, D.E. Saranson, L. Gillman, Eds., Springer, New York, 1983.
- [29] P.R. Halmos, *Lectures on Boolean algebras*, Van Nostrand, New York, 1963; ristampa, Pringer, New York, 1974.
- [30] E. Hopf, The generally temporally discrete Markoff process, *J. Rat. Mech. Anal.* **3**, 13–45 (1954).
- [31] K. Huang, *Statistical Mechanics*, Wiley, New York, 1963.
- [32] R. Jancel, *Les fondaments de la mécanique classique et quantique*, Gauthier-Villars, Paris, 1963; traduzione inglese, *Foundations of classical and quantum statistical mechanics*, Pergamon Press, Oxford, 1969.
- [33] E.T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* **106**, 620–630 (1967).
- [34] E.T. Jaynes, Information theory and statistical mechanics II, *Phys. Rev.* **108**, 171–190 (1967).
- [35] A. Katz, *Principles of statistical mechanics. The information theory approach*, Freeman, San Francisco, 1967.
- [36] J.G. Kemeny, J.L. Snell, *Finite Markov chains*, Van Nostrand, Princeton NJ, 1960; ristampa, Springer, New York–Heidelberg–Berlin, 1976.
- [37] A.I. Khinchin, *Mathematical foundations of statistical mechanics*, Dover, New York 1949.
- [38] J.F.C. Kingman, S.J. Taylor, *Introduction to measure and probability*, Cambridge University Press, 1966.
- [39] A.N. Kolmogorov, New metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces, *Dokl. Akad. Nauk SSSR*, **119**, No. 5, 861–864 (1958), riprodotto in [68, pp. 52–61]
- [40] B.O. Koopman, J. von Neumann, Hamiltonian systems and transformations in Hilbert spaces, *Proc. Nat. Acad. Sci. USA*, **17**, 315–318 (1931).
- [41] L.D. Landau, E.M. Lifchitz, *Mécanique*, Mir, Moscou, 1966.
- [42] L.D. Landau, E.M. Lifchitz, *Physique statistique*, Mir, Moscou, 1967.
- [43] T. Levi-Civita, U. Amaldi, *Lezioni di Meccanica razionale. Vol. 2, Parte 2: Dinamica dei sistemi con un numero finito di gradi di libertà*, Zanichelli, Bologna, 1927; ristampa, *ibidem*, 1974.
- [44] G.W. Mackey, Ergodic theory and its significance for statistical mechanics and probability theory, *Adv. Math.* **12**, 178–268 (1974).
- [45] R. Mañé, *Ergodic theory and differentiable dynamics*, Springer, Berlin–Heidelberg (1987).
- [46] N.F.G. Martin, J.W. England, *Mathematical theory of entropy*, Addison–Wesley, Reading MA (1981) (Encyclopedia of Mathematics and Its Applications, vol. 12).
- [47] R.A. Minlos, *Introduction to mathematical statistical physics*, American Mathematical Society, Providence RI, 2000 (University Lecture Series 19).
- [48] L. Nachbin, *The Haar integral*, Van Nostrand, Princeton, 1965.
- [49] M.G. Nadkarni, *Basic ergodic theory*, Birkhäuser, Berlin (1998).
- [50] J. von Neumann, Proof of the quasi-ergodic hypothesis, *Proc. Acad. Nat. Sci. USA*, **18**, 70–82 (1932).
- [51] E.T. Olsen, W.F. Darsow, B. Nguyen, Copulas and Markov operators, in L. Rüschendorf, B. Schweizer, M.D. Taylor, Eds., *Distribution with fixed marginals and related*

- topics*, Institute of Mathematical Statistics, Lecture Notes—Monograph Series, vol. 28, Hayward CA, 1996, pp.244–259.
- [52] D. Ornstein, *Ergodic theory, randomness, and dynamical systems*, Yale University Press, New Haven CT, 1974.
- [53] W. Parry, *Topics in ergodic theory*, Cambridge University Press, 1981.
- [54] R.K. Pathria, *Statistical mechanics*, Pergamon, Oxford – New York – Toronto, 1972.
- [55] W. Pauli, *Termodinamica e teoria cinetica dei gas*, Boringhieri, Torino, 1964.
- [56] O. Penrose, Foundations of statistical mechanics, *Rep. Prog. Phys.* **42**, 1937–2006 (1979).
- [57] K. Petersen, *Ergodic theory*, Cambridge University Press (1983).
- [58] A. Rényi, On mixing sequences of sets, *Acta Math. Acad. Sci. Hung.*, **9**, 215–228 (1958); anche in *Selected papers of Alfréd Rényi*, P. Turán, Ed., Akadémiai Kiadó, 1976, vol. 2, pp.159–172.
- [59] F. Riesz, Sur la théorie ergodique, *Comm. Math. helvetici* **17**, 221-239 (1945).
- [60] V.A. Rokhlin, La “generica” trasformazione che conserva la misura non è mescolante (in russo), *Dokl. Akad. Nauk SSSR* **60**, 349–351 (1948).
- [61] H.L. Royden, *Real analysis*, Macmillan, New York, 1968 (2nd ed.).
- [62] D. Ruelle, *Statistical Mechanics. Rigorous results*, Benjamin, New York–Amsterdam, 1969.
- [63] D.E. Saranson, L. Gillman, Eds., *P.R. Halmos, Selecta. Expository writing*, Springer, New York, 1983.
- [64] D.E. Saranson, L. Gillman, Eds., *P.R. Halmos, Selecta. Research contributions*, Springer, New York, 1983.
- [65] C. Sempì, *Entropies in information theory and statistical mechanics. Related methods*, Ph. D. Dissertation, University of Waterloo, Ontario, Canada, 1974.
- [66] C. Sempì, A proposito del principio di massima entropia, *Statistica (Bologna)* **46**, 59–66 (1986).
- [67] P.C. Shields, *The theory of Bernoulli shifts*, University of Chicago Press, 1973.
- [68] A.N. Shyryayev, Ed., *Selected works of A.N. Kolmogorov. Vol. III: Information theory and the theory of algorithms*, Kluwer, Dordrecht, 1993.
- [69] C.J. Thompson, *Mathematical statistical mechanics*, Macmillan, New York, 1972.
- [70] P. Walters, *An introduction to ergodic theory*, Springer, New York – Heidelberg – Berlin (1982).
- [71] G.H. Wannier, *Statistical physics*, Wiley, New York, 1966.
- [72] A. Wehrl, General properties of entropy, *Rev. Mod. Phys.* **50**, 221–260.
- [73] N. Wiener, The ergodic theorem, *Duke Math. J.* **5**, 1–18 (1939).
- [74] A.H. Wilson, *Thermodynamics and statistical mechanics*, Cambridge University Press, 1996.
- [75] K. Yosida, S. Kakutani, Birkhoff’s ergodic theorem and the maximal ergodic theorem, *Proc. Japan Acad.*, **15**, 165–168 (1939).

