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 piú 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 $(\Omega, \mathcal{F})$ 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 $(\Omega, \mathcal{F})$ sono necessariamente disgiunti. Se A_1, A_2, ..., A_n sono tutti gli atomi di $(\Omega, \mathcal{F})$ 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 $(\Omega, \mathcal{F})$ può essere ordinata parzialmente come segue. Date due partizioni finite π e π' di $(\Omega, \mathcal{F})$ 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 $(\Omega, \mathcal{F})$ 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 $(\Omega, \mathcal{F})$. 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 $(\Omega, \mathcal{F})$, 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 $(\Omega, \mathcal{F})$ 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 $\pi(B) \subset \mathcal{C}$; 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 $\pi(C)$ 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 $C_{i,j} \in \pi(\mathcal{C})$ $(i = 1, 2, \ldots, p; j = 1, 2, \ldots, s(i))$. Dunque B è in $\mathcal{C}$.

La (c) e la (d) sono ovvie. □

Se $\pi = \{A_1, A_2, \ldots, A_n\}$ e $\mathcal{B}$ sono, rispettivamente, una partizione di $(\Omega, \mathcal{F})$ e una sotto–tribú di $\mathcal{F}$, si pone, per $r \in \mathbf{Z}$,

$$T^r \pi := \{T^r A_1, T^r A_2, \ldots, T^r A_n\},$$
$$T^r \mathcal{B} := \{T^r A : A \in \mathcal{B}\}.$$

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 $(\Omega, \mathcal{F})$, $\mathcal{B}$ 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, $\pi \le \pi'$, allora $T^{-n} \pi'$ è un raffinamento di $T^{-n} \pi$, $T^{-n} \pi \le T^{-n} \pi'$;*
(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

$$B = \cap_{i=1}^{r} T^{-n} B_i = T^{-n} \left(\cap_{i=1}^{r} B_i\right),$$

con $B_i \in \mathcal{B}$ $(i = 1, 2, \ldots, r)$, 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

$$A = \cup_{j=1}^{r} T^{-n} A_j = T^{-n} \left(\cup_{j=1}^{r} A_j\right),$$

con $A_j \in \pi$ $(j = 1, 2, \ldots, r)$, 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

$$A = T^{-n}\,(B \cap C) = \left(T^{-n}\,B\right) \cap \left(T^{-n}\,C\right),$$

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

$$T^{-n}\,A = \cup_{j=1}^{n}\left(T^{-n}\,B_j\right),$$

sicché $T^{-n}\,\pi'$ è un raffinamento di $T^{-n}\,\pi$.

La (f) è banale. $\square$

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–tribú 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

$$H(\mathcal{A}) = H(\pi(\mathcal{A})) := -\sum_{j=1}^{n} \mu\left(A_j\right)\,\log\mu\left(A_j\right), \tag{2.1}$$

ove i logaritmi, qui e nel seguito, si intendono in base 2, e si adotta la convenzione

$$0\,\log 0 =: 0. \tag{2.2}$$

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 $H : \mathcal{P} \longrightarrow \mathbf{R}$ 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

$$\begin{aligned} H(\pi) &= -\sum_{j=1}^{n} \mu(A_j) \log \mu(A_j) = \sum_{j=1}^{n} \varphi(\mu(A_j)) \\ &= n \sum_{j=1}^{n} \frac{\varphi(\mu(A_j))}{n} \le n\,\varphi\left(\frac{\sum_{j=1}^{n} \mu(A_j)}{n}\right) \\ &= n\,\varphi\left(\frac{1}{n}\right) = \log n, \end{aligned}$$

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 $\mathcal{A} \overset{\circ}{=} \mathcal{B}$, 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 $\mathcal{A} \overset{\circ}{\subset} \mathcal{B}$.

PROPOSIZIONE 2. *Se $\mathcal{A}$ e $\mathcal{B}$ sono due sotto–tribú finite ed equivalenti di $\mathcal{F}$, $\mathcal{A} \overset{\circ}{=} \mathcal{B}$, 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 $j \ge k+1$. 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 $j \ge k$, 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. □

DEFINIZIONE 2.3. Date due sottotribú finite $\mathcal{A}$ e $\mathcal{B}$ di $\mathcal{F}$ talii 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 $H : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbf{R}_+$ definita da

$$H(\mathcal{B} \mid \mathcal{A}) := -\sum_{i=1}^{n} \mu(A_i) \sum_{j=1}^{s} \mu(B_j \mid A_i) \log \mu(B_j \mid A_i), \tag{2.3}$$

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, $H(\mathcal{A} \mid \mathcal{N}) = H(\mathcal{A})$;*
(b) *se $\mathcal{A}$ e $\mathcal{B}$ sono equivalenti, $\mathcal{A} \overset{\circ}{=} \mathcal{B}$, allora, per ogni $\mathcal{C} \in \mathcal{P}$, è $H(\mathcal{A} \mid \mathcal{C}) = H(\mathcal{B} \mid \mathcal{C})$;*
(c) *se $\mathcal{B}$ e $\mathcal{C}$ sono equivalenti, $\mathcal{B} \overset{\circ}{=} \mathcal{C}$, allora, per ogni $\mathcal{A} \in \mathcal{P}$, è $H(\mathcal{A} \mid \mathcal{B}) = H(\mathcal{A} \mid \mathcal{C})$.*

DIMOSTRAZIONE. La (a) è ovvia.

(b) In questo caso si ha $\pi(A) = \{A_1, A_2, \dots, A_n\}$, $\pi(\mathcal{B}) = \{B_1, B_2, \dots, B_s\}$, $\mu(A_j \Delta B_j) = 0$ 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

$$\begin{aligned} H(\mathcal{A} \mid \mathcal{C}) &= -\sum_{j=1}^{r} \mu(C_j) \sum_{i=1}^{n} \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \\ &= -\sum_{j=1}^{r} \mu(C_j) \sum_{i=1}^{n} \frac{\mu(A_i \cap C_j \cap B_i)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j \cap B_i)}{\mu(C_j)} \\ &= -\sum_{j=1}^{r} \mu(C_j) \sum_{i=1}^{n} \frac{\mu(B_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(B_i \cap C_j)}{\mu(C_j)} = H(\mathcal{B} \mid \mathcal{C}), \end{aligned}$$

poiché si ha $0 \leq \mu(B_i \cap C_j) \leq \mu(B_i) = 0$ per $i = n+1, n+2, \ldots, s$.

La dimostrazione della (c) è del tutto analoga. □

Teorema 2.2. *Siano $\mathcal{A}$, $\mathcal{B}$ e $\mathcal{C}$ sotto–tribú finite di $\mathcal{F}$. Valgono le proprietà*

(a) *(additività condizionata forte)*

$$H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = H(\mathcal{A} \mid \mathcal{C}) + H(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C});$$

(b) *(additività forte)* $H(\mathcal{A} \vee \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A})$;
(c) $H(\mathcal{A} \mid \mathcal{C}) \leq H(\mathcal{B} \mid \mathcal{C})$ *se* $\mathcal{A} \subset \mathcal{B}$;
(d) $H(\mathcal{A}) \leq H(\mathcal{B})$ *se* $\mathcal{A} \subset \mathcal{B}$;
(e) $H(\mathcal{A} \mid \mathcal{B}) \geq H(\mathcal{A} \mid \mathcal{C})$ *se* $\mathcal{B} \subset \mathcal{C}$;
(f) $H(\mathcal{A} \mid \mathcal{C}) \leq H(\mathcal{A})$;
(g) *(subadditività condizionata)* $H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) \leq H(\mathcal{A} \mid \mathcal{C}) + H(\mathcal{B} \mid \mathcal{C})$;
(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 $\mathcal{C}$,

$$\pi(A) = \{A_1, A_2, \ldots, A_n\}, \qquad \pi(\mathcal{B}) = \{B_1, B_2, \ldots, B_s\}$$
$$\pi(\mathcal{C}) = \{C_1, C_2, \ldots, C_r\}.$$

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

$$H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C});$$

pertanto

$$\begin{aligned} H(\mathcal{A}\vee\mathcal{B}\mid\mathcal{C}) &= -\sum_{i,j,k}\mu(A_i\cap B_j\cap C_k)\,\log\frac{\mu(A_i\cap B_j\cap C_k)}{\mu(A_i\cap C_k)} \\ &\qquad -\sum_{i,j,k}\mu(A_i\cap B_j\cap C_k)\,\log\frac{\mu(A_i\cap C_k)}{\mu(C_k)} \\ &= H(\mathcal{B}\mid\mathcal{A}\vee\mathcal{C})-\sum_{i,k}\mu(A_i\cap C_k)\,\log\frac{\mu(A_i\cap C_k)}{\mu(C_k)} \\ &= H(\mathcal{A}\mid\mathcal{C})+H(\mathcal{B}\mid\mathcal{A}\vee\mathcal{C}). \end{aligned}$$

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

$$H(\mathcal{B}\mid\mathcal{C})=H(\mathcal{A}\vee\mathcal{B}\mid\mathcal{C})=H(\mathcal{A}\mid\mathcal{C})+H(\mathcal{B}\mid\mathcal{A}\vee\mathcal{C})\geq H(\mathcal{A}\mid\mathcal{C}).$$

La (d) scende prendendo $\mathcal{C}=\mathcal{N}$ nella (c).

(e) Si ponga

$$\alpha_k:=\frac{\mu(B_j\cap C_k)}{\mu(B_j)},\qquad x_k:=\frac{\mu(A_i\cap C_k)}{\mu(C_k)};$$

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

$$\begin{aligned} \sum_{k=1}^{s}\frac{\mu(A_i\cap C_k)\,\mu(B_j\cap C_k)}{\mu(C_k)} &= \sum_{k=1}^{s}\frac{\mu(A_i\cap C_k)}{\mu(C_k)}\,\delta_{k,j(h)}\,\mu(C_k) \\ &= \sum_{h\in I_j}\mu\left(A_i\cap C_{j(h)}\right)=\mu(A_i\cap B_j). \end{aligned}$$

Moltiplicando la (2.4) par $\mu(B_j)$ e sommando su i e j si ottiene

$$\begin{aligned}\sum_{i,j}\mu(A_i\cap B_j)\log\frac{\mu(A_i\cap B_j)}{\mu(B_j)}\\ &\le\sum_{i,j,k}\mu(B_j\cap C_k)\,\frac{\mu(A_i\cap C_k)}{\mu(C_k)}\,\log\frac{\mu(A_i\cap C_k)}{\mu(C_k)}\\ &=\sum_{i,k}\mu(A_i\cap C_k)\log\frac{\mu(A_i\cap C_k)}{\mu(C_k)},\end{aligned}$$

vale a dire

$$-H(\mathcal{A}\mid\mathcal{B})\le -H(\mathcal{A}\mid\mathcal{C}),$$

onde l'asserto.

La (f) segue prendendo $\mathcal{B}=\mathcal{N}$.

(g) Si usino la (a) e la (e), poiché $\mathcal{C}\subset\mathcal{A}\vee\mathcal{C}$,

$$H(\mathcal{A}\vee\mathcal{B}\mid\mathcal{C})=H(\mathcal{A}\mid\mathcal{C})+H(\mathcal{B}\mid\mathcal{A}\vee\mathcal{C})\le H(\mathcal{A}\mid\mathcal{C})+H(\mathcal{B}\mid\mathcal{C}).$$

(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) $H\left(T^{-1}\mathcal{A}\mid T^{-1}\mathcal{B}\right)=H(\mathcal{A}\mid\mathcal{B})$;
(b) $H\left(T^{-1}\mathcal{A}\right)=H(\mathcal{A})$.

DIMOSTRAZIONE. Basta osservare che, poiché T conserva la misura, si ha $\mu\left(T^{-1}A\cap T^{-1}B\right)=\mu(A\cap B)$. □

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) $H(\mathcal{A}\mid\mathcal{B})=0$.

DIMOSTRAZIONE. Si considerino le partizioni generate da $\mathcal{A}$ e da $\mathcal{B}$ rispettivamente,

$$\pi(\mathcal{A})=\{A_1,A_2,\dots,A_n\}\qquad\text{e}\qquad\pi(\mathcal{B})=\{B_1,B_2,\dots,B_s\}.$$

Non è restrittivo supporre che tutti gli insiemi di tali partizioni abbiano probabilità non nulla.

(a) $\Longrightarrow$ (b) L'ipotesi dice che, per ogni coppia di indici, j e k, è

$$\mu\,(A_j \cap B_k) = 0 \qquad \text{oppure} \qquad \mu\,(A_j \cap B_k) = \mu\,(B_k)\,,$$

sicché $H(\mathcal{A} \mid \mathcal{B}) = 0$.

(b) $\Longrightarrow$ (a) Ogni termine della somma che definisce $H(\mathcal{A} \mid \mathcal{B})$ 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–tribú 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) *$H(\mathcal{A} \mid \mathcal{B}) = H(\mathcal{A})$.*

DIMOSTRAZIONE. (a) $\Longrightarrow$ (b) Segue immediatamente dalla definizione di entropia condizionata.

(b) $\Longrightarrow$ (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*

$$\widetilde{\mathcal{P}} := \mathcal{P}/\stackrel{\circ}{=}$$

(in $\mathcal{P}$ si identificano due sotto–tribú finite $\mathcal{A}$ e $\mathcal{B}$ di $\mathcal{F}$ se $\mathcal{A} \stackrel{\circ}{=} \mathcal{B}$). Posto

$$d(\mathcal{A},\mathcal{B}) := H(\mathcal{A} \mid \mathcal{B}) + H(\mathcal{B} \mid \mathcal{A}),$$

la coppia $(\widetilde{\mathcal{P}}, d)$ è uno spazio metrico.

DIMOSTRAZIONE. È noto che $d(\mathcal{A},\mathcal{B}) \geq 0$; il segno d'eguaglianza vale se, e solo se, $\mathcal{A} \stackrel{\circ}{=} \mathcal{B}$. Inoltre la simmetria vale per definizione. Infine, se $\mathcal{C}$ è un'altra sotto–tribú finita di $\mathcal{F}$, è

$$H(\mathcal{A} \mid \mathcal{C}) \leq H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = H(\mathcal{B} \mid \mathcal{C}) + H(\mathcal{A} \mid \mathcal{B} \vee \mathcal{C}) \leq H(\mathcal{B} \mid \mathcal{C}) + H(\mathcal{A} \mid \mathcal{B});$$

similmente si ha $H(\mathcal{C} \mid \mathcal{A}) \leq H(\mathcal{B} \mid \mathcal{A}) + H(\mathcal{C} \mid \mathcal{B})$ 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

$$\lim_{n\to+\infty} \frac{a_n}{n} = \inf_{n\in\mathbf{N}} \frac{a_n}{n}. \tag{2.5}$$

DIMOSTRAZIONE. Quali che siano i numeri naturali k e n si ha

$$a_{kn} = a_{(k-1)n+n} \leq a_{(k-1)n} + a_n \leq \cdots \leq k\, a_n.$$

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

$$\limsup_{n\to+\infty} \frac{a_j}{j} \leq \frac{a_n}{n}$$

e, di qui

$$\limsup_{n\to+\infty} \frac{a_j}{j} \leq \inf_{n\in\mathbf{N}} \frac{a_n}{n}.$$

Poiché, d'altro canto,

$$\liminf_{n\to+\infty} \frac{a_j}{j} \geq \inf_{n\in\mathbf{N}} \frac{a_n}{n},$$

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–tribú finita di $\mathcal{F}$; allora esiste il limite*

$$\lim_{n \to +\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right).$$

DIMOSTRAZIONE. Posto

$$a_n := H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right),$$

si controlla immediatamente, in virtú 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–tribú finita $\mathcal{A}$ di $\mathcal{F}$* il limite

$$h(T, \mathcal{A}) := \lim_{n \to +\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right).$$

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–tribú 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 spazî 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 $\Phi : (\widetilde{\mathcal{F}}_2, \widetilde{\mu}_2) \longrightarrow (\widetilde{\mathcal{F}}_1, \widetilde{\mu}_1)$ di algebre di misura tale che $\Phi\widetilde{T}_2^{-1} = \widetilde{T}_1^{-1}\Phi$. 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 $\widetilde{B}_j = \Phi(\widetilde{A}_j)$ $(j = 1, 2, \dots, s)$. Sia $\mathcal{A}_1$ la sotto–tribú finita generata dalla partizione π_1. Ora

$$\mu_1 \left(\bigcap_{j=0}^{n-1} T_1^{-1} B_{k(j)} \right) = \mu_2 \left(\bigcap_{j=0}^{n-1} T_2^{-1} A_{k(j)} \right)$$

ove $k(j) \in \{1, 1, \dots, s\}$, poiché

$$\begin{aligned} \Phi \left[\bigcap_{j=0}^{n-1} \left(T_2^{-j} A_{k(j)} \right)^{\sim} \right] &= \Phi \left(\bigcap_{j=0}^{n-1} \widetilde{T}_2^{-j} \widetilde{A}_{k(j)} \right) \\ &= \bigcap_{j=0}^{n-1} \Phi\widetilde{T}_2^{-j} \widetilde{A}_{k(j)} = \bigcap_{j=0}^{n-1} \widetilde{T}_1^{-j} \Phi \left(\widetilde{A}_{k(j)} \right) \\ &= \bigcap_{j=0}^{n-1} \widetilde{T}_1^{-j} \widetilde{B}_{k(j)} = \bigcap_{j=0}^{n-1} \left(T_1^{-j} B_{k(j)} \right)^{\sim} . \end{aligned}$$

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 $h(T_1) = h(T_2)$. □

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) $h(T, \mathcal{A}) \leq h(T, \mathcal{B}) + H(\mathcal{A} \mid \mathcal{B})$.

DIMOSTRAZIONE. (a) Si ha

$$\frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right) \leq \frac{1}{n}\sum_{j=0}^{n-1} H\left(T^{-j}\mathcal{A}\right) = \frac{1}{n}\sum_{j=0}^{n-1} H(\mathcal{A}) = H(\mathcal{A}).$$

Per la (b) è

$$\begin{aligned}
H\left[\bigvee_{j=0}^{n-1} T^{-j}\,(\mathcal{A}\vee\mathcal{B})\right] &= H\left[\bigvee_{j=0}^{n-1}\left(T^{-j}\mathcal{A}\vee T^{-j}\mathcal{B}\right)\right] \\
&\leq H\left[\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right)\bigvee\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right)\right] \\
&\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right) + H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right).
\end{aligned}$$

(c) Dall'inclusione $\mathcal{A}\subset\mathcal{B}$ scende $\vee_{j=0}^{n-1}T^{-j}\mathcal{A}\subset\vee_{j=0}^{n-1}T^{-j}\mathcal{B}$ di modo che il risultato segue dalla (d) del teorema 2.2.

Infine, per quanto riguarda la (d), si ha

$$\begin{aligned}
H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right) &\leq H\left[\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right)\bigvee\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right)\right] \\
&\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right) + H\left[\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right)\middle|\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right)\right];
\end{aligned}$$

ma, per la subadditività condizionata risulta

$$\begin{aligned}
H\left[\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right)\middle|\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right)\right] &\leq \sum_{j=1}^{n-1} H\left[T^{-j}\mathcal{A}\middle|\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{B}\right)\right] \\
&\leq \sum_{j=1}^{n-1} H\left[T^{-j}\mathcal{A}\mid T^{-j}\mathcal{B}\right] = \sum_{j=1}^{n-1} H(\mathcal{A}\mid\mathcal{B}) = n\,H(\mathcal{A}\mid\mathcal{B}),
\end{aligned}$$

onde

$$H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right) \leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right) + n\,H(\mathcal{A}\mid\mathcal{B});$$

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

$$\begin{aligned} h\left(T, \bigvee_{j=-m}^{m} T^{-j} A\right) &= \lim_{k\to+\infty} \frac{1}{k} H\left[\bigvee_{j=0}^{k-1} T^{-j}\left(\bigvee_{i=-m}^{m} T^{-i} A\right)\right] \\ &= \lim_{k\to+\infty} \frac{1}{k} H\left(\bigvee_{j=-m}^{m+k-1} T^{-j} A\right). \end{aligned}$$

D'altra parte, risulta

$$\begin{aligned} H\left(\bigvee_{j=-m}^{m+k-1} T^{-j} A\right) &= H\left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A\right) \bigvee \left(\bigvee_{j=m}^{m+k-1} T^{-j} A\right)\right] \\ &= H\left(\bigvee_{j=m}^{m+k-1} T^{-j} A\right) + H\left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A\right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A\right)\right] \\ &= H\left[T^{-m}\left(\bigvee_{j=0}^{k-1} T^{-j} \mathcal{A}\right)\right] + H\left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A\right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A\right)\right] \\ &= H\left(\bigvee_{j=0}^{k-1} T^{-j} \mathcal{A}\right) + H\left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A\right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A\right)\right]. \end{aligned}$$

Per concludere la dimostrazione basterà far vedere che

$$\lim_{k\to+\infty} \frac{1}{k} H\left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A\right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A\right)\right] = 0.$$

La (f) del Teorema 2.2 dà

$$0 \le \frac{1}{k} H\left[\left(\bigvee_{j=-m}^{m-1} T^{-j} A\right) \middle| \left(\bigvee_{j=m}^{m+k-1} T^{-j} A\right)\right] \le \frac{1}{k} H\left(\bigvee_{j=-m}^{m-1} T^{-j} A\right),$$

ma quest'ultima espressione tende a zero la tendere di k a $+\infty$. □

Vi è una maniera alternativa di definire l'entropia.

TEOREMA 3.3. *Siano T una trasformazione che conserva la misura sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$ e $\mathcal{A}$ una sotto–tribú finita di $\mathcal{F}$; allora*

$$h(T, \mathcal{A}) = \lim_{n \to +\infty} H\left[\mathcal{A} \middle| \bigvee_{j=1}^{n} T^{-j}\mathcal{A}\right]. \tag{3.1}$$

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

$$H\left[\mathcal{A} \middle| \bigvee_{j=1}^{n} T^{-j}\mathcal{A}\right] \geq H\left[\mathcal{A} \middle| \bigvee_{j=1}^{n+1} T^{-j}\mathcal{A}\right]$$

di modo che il limite che compare nella (3.1) esiste finito.

Si mostrerà per induzione che

$$H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right) = H(\mathcal{A}) + \sum_{j=1}^{n-1} H\left[\mathcal{A} \middle| \bigvee_{i=1}^{j} T^{-i}\mathcal{A}\right]. \tag{3.2}$$

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 virtú dell'ipotesi d'induzione, si ha, per $r + 1$,

$$\begin{aligned}
H\left(\bigvee_{j=0}^{r} T^{-j}\mathcal{A}\right) &= H\left[\left(\bigvee_{j=1}^{r-1} T^{-j}\mathcal{A}\right) \vee \mathcal{A}\right] \\
&= H\left(\bigvee_{j=1}^{r} T^{-j}\mathcal{A}\right) + H\left[\mathcal{A} \middle| \bigvee_{i=1}^{r} T^{-i}\mathcal{A}\right] \\
&= H\left(\bigvee_{j=0}^{r-1} T^{-j}\mathcal{A}\right) + H\left[\mathcal{A} \middle| \bigvee_{i=1}^{r} T^{-i}\mathcal{A}\right] \\
&= H(\mathcal{A}) + \sum_{j=1}^{r-1} H\left[\mathcal{A} \middle| \bigvee_{i=1}^{j} T^{-i}\mathcal{A}\right] + H\left[\mathcal{A} \middle| \bigvee_{i=1}^{r} T^{-i}\mathcal{A}\right] \\
&= H(\mathcal{A}) + \sum_{j=1}^{r} H\left[\mathcal{A} \middle| \bigvee_{i=1}^{j} T^{-i}\mathcal{A}\right],
\end{aligned}$$

sicché la (3.2) è dimostrata. Si divida ora la (3.2) per n e se ne consideri il limite per $n \to +\infty$. 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 mebro 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

$$\begin{aligned} H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) &= H(\mathcal{A}) + \sum_{j=1}^{n-1} H \left[\mathcal{A} \middle| \bigvee_{i=1}^{j} T^{-i} \mathcal{A} \right] \\ &\geq H \left[\mathcal{A} \middle| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \right] + (n-1) H \left[\mathcal{A} \middle| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \right] \\ &= n H \left[\mathcal{A} \middle| \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) \right] . \end{aligned}$$

Pertanto

$$n \left\{ H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \middle| \bigvee_{j=1}^{n} T^{-j} \mathcal{A} \right] \right\} \leq (n+1) H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right)$$

e, quindi,

$$\begin{aligned} n H \left(\bigvee_{j=0}^{n} T^{-j} \mathcal{A} \right) &= n \left\{ H \left(\bigvee_{j=1}^{n} T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \middle| \bigvee_{j=1}^{n} T^{-j} \mathcal{A} \right] \right\} \\ &= n \left\{ H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right) + H \left[\mathcal{A} \middle| \bigvee_{j=1}^{n} T^{-j} \mathcal{A} \right] \right\} \\ &\leq (n+1) H \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \right), \end{aligned}$$

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*

$$h\left(T^m\right) = m\, h(T).$$

Se, inoltre, la trasformazione T è anche invertibile, allora, per ogni $m \in \mathbf{Z}$, è

$$h\left(T^m\right) = |m|\, h(T).$$

DIMOSTRAZIONE. Si dimostrerà dapprima che, se $\mathcal{A}$ è una sotto–tribú finita di $\mathcal{F}$, è

$$h\left(T^m, \bigvee_{j=0}^{m-1} T^{-j}\mathcal{A}\right) = m\, h(T, \mathcal{A}). \tag{4.1}$$

Infatti

$$\begin{aligned}\lim_{k\to+\infty} \frac{1}{k} H\left[\bigvee_{i=0}^{k-1} T^{-mi}\left(\bigvee_{j=0}^{m-1} T^{-j}\mathcal{A}\right)\right] &= \lim_{k\to+\infty} \frac{m}{mk} H\left(\bigvee_{j=0}^{km-1} T^{-j}\mathcal{A}\right)\\ &= m\, h(T, \mathcal{A}).\end{aligned}$$

Scende dalla (4.1) che

$$\begin{aligned}m\, h(T) = m \sup\{h(T, \mathcal{A}) : \mathcal{A} \in \mathcal{P}\} &= \sup\left\{h\left(T^m, \bigvee_{j=0}^{m-1} T^{-j}\mathcal{A}\right) : \mathcal{A} \in \mathcal{P}\right\}\\ &\leq \sup\left\{h\left(T^m, \mathcal{B}\right) : \mathcal{B} \in \mathcal{P}\right\} = h\left(T^m\right),\end{aligned}$$

vale a dire

$$m\, h(T) \leq h\left(T^m\right).$$

D'altra parte, per la (c) del Teorema 3.1, è

$$h\left(T^m, \mathcal{A}\right) \leq h\left(T^m, \bigvee_{j=0}^{m-1} T^{-j}\mathcal{A}\right) = m\, h(T, \mathcal{A}),$$

onde

$$h\left(T^m\right) \leq m\, h(T)$$

che dimostra l'asserto.

Si supponga ora che T sia anche invertibile. In virtú di quanto precede, basta dimostrare che $h\left(T^{-1}\right) = h(T)$, e, a tal fine basta dimostrare che, per ogni sotto–tribú finita $\mathcal{A}$ di $\mathcal{F}$, si ha $h(T, \mathcal{A}) = h\left(T^{-1}, \mathcal{A}\right)$. 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. □

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 $\mathcal{F} \overset{\circ}{=} \mathcal{F}(\mathcal{A})$, e $\mathcal{B}$ una sotto–tribú finita di $\mathcal{F}$. Allora, per ogni $\varepsilon > 0$, esiste una sotto–tribú finita $\mathcal{C}_\varepsilon$ di $\mathcal{A}$ tale che*

$$H\left(\mathcal{C}_\varepsilon \mid \mathcal{B}\right) + H\left(\mathcal{B} \mid \mathcal{C}_\varepsilon\right) < \varepsilon. \tag{4.2}$$

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 $\left\{B_1, B_2, \dots, B_{s-1}, \cup_{j=s}^r B_j\right\}$ si avrebbe

$$H\left(\mathcal{B}' \mid \mathcal{C}_\varepsilon\right) = H\left(\mathcal{B} \mid \mathcal{C}_\varepsilon\right) \qquad \text{e} \qquad H\left(\mathcal{C}_\varepsilon \mid \mathcal{B}'\right) = H\left(\mathcal{C}_\varepsilon \mid \mathcal{B}\right).$$

In virtú della continuità della funzione $\varphi : [0,1] \longrightarrow \mathbf{R}_+$ 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 $\pi = \{C_1, C_2, \dots, C_r\}$ soddisfà alle diseguaglianze

$$\mu\left(B_j \Delta C_j\right) \leq \delta := \min\left\{\delta_0 \, \frac{\mu(B_i)}{2} : i = 1, 2, \dots, r\right\} \qquad (j = 1, 2, \dots, r), \tag{4.3}$$

allora risulta, se $\mathcal{C}_\varepsilon = \mathcal{F}(\pi)$,

$$H(\mathcal{B} \mid \mathcal{C}_\varepsilon) < \frac{\varepsilon}{2}. \tag{4.4}$$

Infatti, se la partizione π soddisfà alla (4.3), si ha

$$\begin{aligned}\mu\left(B_j\right) &= \mu\left(B_j \setminus C_j\right) + \mu\left(B_j \cap C_j\right) \leq \mu\left(B_j \Delta C_j\right) + \mu\left(C_j\right) \\ &\leq \delta + \mu\left(C_j\right) < \frac{\mu\left(B_j\right)}{2} + \mu\left(C_j\right),\end{aligned}$$

relazione che dà $\mu(B_j) < 2\,\mu(C_j)$ e, quindi,

$$\mu(C_j) - \mu(B_j \cap C_j) \le \mu(B_j \Delta C_j) \le \delta < \delta_0\, \mu(C_j).$$

Perciò

$$\mu(B_j \mid C_j) = \frac{\mu(B_j \cap C_j)}{\mu(C_j)} \ge 1 - \delta_0 \qquad (j = 1, 2, \dots, r),$$

onde

$$\begin{aligned} H(\mathcal{B} \mid \mathcal{C}_\varepsilon) &= -\sum_{j=1}^{r} \mu(C_j) \sum_{k=1}^{r} \mu(B_j \mid C_j) \log \mu(B_j \mid C_j) \\ &< \sum_{j=1}^{r} \mu(C_j) \sum_{k=1}^{r} \frac{\varepsilon}{2r}, \end{aligned}$$

ciò che dimostra la (4.4). La (4.4) continua a valere se la (4.3) è sostituita dalla diseguaglianza piú forte

$$\mu(B_j \Delta C_j) \le \frac{\delta_0}{4} \min\{\mu(B_i) : i = 1, 2, \dots, r\} \qquad (j = 1, 2, \dots, r). \quad (4.5)$$

Ma, se vale la (4.5), risulta

$$\mu(B_j \Delta C_j) < \frac{\mu(B_j)}{4}$$

e, quindi,

$$\mu(B_j) = \mu(B_j \setminus C_j) + \mu(B_j \cap C_j) \le \mu(B_j \Delta C_j) + \mu(C_j) < \frac{\mu(B_j)}{4} + \mu(C_j),$$

onde

$$\mu(C_j) > \frac{3}{4}\, \mu(B_j).$$

Infine da quest'ultima diseguaglianza 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

$$H(\mathcal{C}_\varepsilon \mid \mathcal{B}) < \frac{\varepsilon}{2}.$$

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

$$\mu(B_j \Delta C_j) \le \frac{\delta_0}{4} \min\{\mu((B_j) : j = 1, 2, \dots, r\} =: \tau \qquad (4.6)$$

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 $k \ne j$, 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 $j = 1, 2, \ldots, r-1$ e $C_r := \left(\cup_{j=1}^{r-1} C_j\right)^c$, sicché $\pi' := \{C_1, C_2, \ldots, C_r\}$ è una partizione finita di Ω in insiemidi $\mathcal{A}$. Per $j = 1, 2, \ldots, r-1$ 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. □

Se $\{\mathcal{A}_n : n \in \mathbf{N}\}$ è una successione di algebre contenute in $\mathcal{F}$ si indicherà con $\vee_{n \in \mathbf{N}} \mathcal{A}_n$ 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*

$$\lim_{n \to +\infty} H(\mathcal{B} \mid \mathcal{A}_n) = 0.$$

DIMOSTRAZIONE. Si ponga $\mathcal{A} := \cup_{n \in \mathbf{N}} \mathcal{A}_n$; 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 $\mathcal{C}_\varepsilon$ di $\mathcal{A}$ tale che $H(\mathcal{B} \mid \mathcal{C}_\varepsilon) < \varepsilon$. Poiché $\mathcal{C}_\varepsilon$ 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

$$H(\mathcal{B} \mid \mathcal{A}_n) \leq H(\mathcal{B} \mid \mathcal{A}_{n_0}) \leq H(\mathcal{B} \mid \mathcal{C}_\varepsilon) < \varepsilon,$$

cioè l'asserto. □

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

$$h(T, \mathcal{B}) \leq h(T, \mathcal{A}). \tag{4.7}$$

Ora

$$\begin{aligned} h(T, \mathcal{B}) &\leq h\left(T, \bigvee_{j=-n}^{n} T^j \mathcal{A}\right) + H\left(\mathcal{B} \middle| \bigvee_{j=-n}^{n} T^j \mathcal{A}\right) \\ &= h(T, \mathcal{A}) + H\left(\mathcal{B} \middle| \bigvee_{j=-n}^{n} T^j \mathcal{A}\right), \end{aligned}$$

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, $\mathcal{B} \subset \vee_{n\in\mathbf{N}} \mathcal{A}_n$, sicché

$$\lim_{n\to+\infty} H\left(\mathcal{B} \mid \mathcal{A}_n\right) = 0$$

per il corollario 4.1. La (4.7) è cosí completamente dimostrata. □

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 $n \in \mathbf{N}$,

$$h(T, \mathcal{A}) = h\left(T, \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}\right). \tag{4.8}$$

Infatti

$$h\left(T, \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}\right) = \lim_{n\to+\infty} \frac{1}{k} H\left[\bigvee_{s=0}^{k-1} T^{-s} \left(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}\right)\right],$$

ma

$$\begin{aligned}
H\left[\bigvee_{s=0}^{k-1} T^{-s}\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right)\right] &= H\left(\bigvee_{j=0}^{n+k-2} T^{-j}\mathcal{A}\right)\\
&= H\left[\left(\bigvee_{j=0}^{n-2} T^{-j}\mathcal{A}\right)\bigvee\left(\bigvee_{j=n-1}^{n+k-2} T^{-j}\mathcal{A}\right)\right]\\
&= H\left(\bigvee_{j=n-1}^{n+k-2} T^{-j}\mathcal{A}\right) + H\left[\left(\bigvee_{j=0}^{n-2} T^{-j}\mathcal{A}\right)\middle|\left(\bigvee_{j=n-1}^{n+k-2} T^{-j}\mathcal{A}\right)\right]\\
&= H\left(\bigvee_{j=0}^{k-1} T^{-j}\mathcal{A}\right) + H\left[\left(\bigvee_{j=0}^{n-2} T^{-j}\mathcal{A}\right)\middle|\left(\bigvee_{j=n-1}^{n+k-2} T^{-j}\mathcal{A}\right)\right].
\end{aligned}$$

Siccome si ha

$$\frac{1}{k}H\left[\left(\bigvee_{j=0}^{n-2} T^{-j}\mathcal{A}\right)\middle|\left(\bigvee_{j=n-1}^{n+k-2} T^{-j}\mathcal{A}\right)\right] \le \frac{1}{k}H\left(\bigvee_{j=0}^{n-2} T^{-j}\mathcal{A}\right) \xrightarrow[k\to+\infty]{} 0,$$

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 virtú della (4.8) si ha

$$\begin{aligned}
h(T,\mathcal{B}) &\le h\left(T, \bigvee_{j=0}^{n} T^{-j}\mathcal{A}\right) + H\left[\mathcal{B}\middle|\left(\bigvee_{j=0}^{n} T^{-j}\mathcal{A}\right)\right]\\
&= h(T,\mathcal{A}) + H\left[\mathcal{B}\middle|\left(\bigvee_{j=0}^{n} T^{-j}\mathcal{A}\right)\right]
\end{aligned}$$

e la tesi segue come nel teorema precedente. □

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

$$h(T) = h(T,\mathcal{A}) = \lim_{n\to+\infty} H\left[\mathcal{A}\middle|\bigvee_{j=1}^{n} T^{-j}\mathcal{A}\right].$$

Ma

$$\bigvee_{n\in\mathbf{N}} T^{-n}\mathcal{A} \overset{\circ}{=} T\mathcal{F} = \mathcal{F}.$$

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

$$\lim_{n\to+\infty} H\left(\mathcal{A} \mid \mathcal{A}_n\right) = 0,$$

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 $\mathcal{F}(\mathcal{A}) \overset{\circ}{=} \mathcal{F}$. Allora*

$$h(T) = \sup\left\{h(T, \mathcal{B}) : \mathcal{B} \in \mathcal{P},\ \mathcal{B} \subset \mathcal{A}\right\}. \tag{4.9}$$

DIMOSTRAZIONE. Poiché l'entropia $h(T)$ maggiora il secondo membro della (4.9), per stabilire la (4.9) basta far vedere che

$$h(T) \leq \sup\left\{h(T, \mathcal{B}) : \mathcal{B} \in \mathcal{P},\ \mathcal{B} \subset \mathcal{A}\right\}. \tag{4.10}$$

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 $\mathcal{B}_\varepsilon$ con $\mathcal{B}_\varepsilon \subset \mathcal{A}$, tale che $H\left(\mathcal{C} \mid \mathcal{B}_\varepsilon\right) < \varepsilon$, sicché

$$h(T, \mathcal{C}) \leq h\left(T, \mathcal{B}_\varepsilon\right) + H\left(\mathcal{C} \mid \mathcal{B}_\varepsilon\right) < h\left(T, \mathcal{B}_\varepsilon\right) + \varepsilon,$$

e, perciò

$$h(T, \mathcal{C}) \leq \varepsilon + \sup\left\{H(T, \mathcal{B}) : \mathcal{B} \in \mathcal{P},\ \mathcal{B} \subset \mathcal{A}\right\},$$

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*

$$\bigvee_{n\in\mathbf{N}} \mathcal{A}_n \overset{\circ}{=} \mathcal{F}.$$

Allora

$$h(T) = \lim_{n\to+\infty} h\left(T, \mathcal{A}_n\right). \tag{4.11}$$

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

$$h(T, \mathcal{B}) \leq h\left(T, \mathcal{A}_{n_0}\right)$$

e, quindi,

$$h(T) \leq \lim_{n\to+\infty} h\left(T, \mathcal{A}_n\right).$$

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 spazî di probabilità $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ rispettivamente. Allora*

$$h\left(T_1 \times T_2\right) = h\left(T_1\right) + h\left(T_2\right). \tag{4.12}$$

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

$$\pi\left(\mathcal{A}_1 \times \mathcal{A}_2\right) = \left\{A_1 \times A_2 : A_1 \in \pi(\mathcal{A}_1),\ A_2 \in \pi(\mathcal{A}_2)\right\}.$$

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 $\mathcal{F}\left(\mathcal{T}_0\right) = \mathcal{F}_1 \times \mathcal{F}_2$ e, per il Teorema 4.5, è

$$h\left(T_1 \times T_2\right) = \sup\left\{h\left(T_1 \times T_2, \mathcal{C}\right) : \mathcal{C} \subset \mathcal{T}_0 \text{ sotto–tribú finita di } \mathcal{F}_1 \times \mathcal{F}_2\right\}$$

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

$$h\left(T_1 \times T_2\right) = \sup\left\{h\left(T_1 \times T_2, \mathcal{A}_1 \times \mathcal{A}_2\right)\right\}$$

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

$$\pi\left(\bigvee_{j=0}^{n-1} T_1^{-j}\mathcal{A}_1\right) \qquad \text{e} \qquad \pi\left(\bigvee_{j=0}^{n-1} T_2^{-j}\mathcal{A}_2\right),$$

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. Esempî 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

$$h(T, \mathcal{A}) = \lim_{n \to +\infty} \frac{1}{n} \, H(\mathcal{A}) = 0.$$

Si osservi che, se T è periodica di periodo k, $T^k = I$, allora $h\left(T^k\right) = 0$. Infatti il Teorema 4.1 dà

$$k\, h(T) = h\left(T^k\right) = h(I) = 0.$$

Esempio 5.2. Le rotazioni $T_a\, z = a\, z$ $(a \in \mathcal{K})$ 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

$$A_1 := \left\{ e^{2\pi i x} : x \in [0, 1/2[\right\} \quad \text{e} \quad A_2 := \left\{ e^{2\pi i x} : x \in [1/2, 1[\right\}$$

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 miniziali in a^{-n} e $-a^{-n}$. Poiché $\{a^{-n} : n \in \mathbf{N}\}$ è densa in $\mathcal{K}$, ogni semicirconferenza appartiene a $\vee_{n\in\mathbf{Z}_+} T_a^{-n}(\pi)$. Di conseguenza ogni arco della circonferenza unitaria appartiene a $\vee_{n\in\mathbf{Z}_+} T_a^{-n}(\pi)$. Perciò

$$\mathcal{B} = \vee_{n\in\mathbf{Z}_+} T_a^{-n}(\pi).$$

In virtú del Corollario 4.2 ciò implica $h(T_a) = 0$.

ESEMPIO 5.3. La traslazione bilatera $(p_0, p_1, \ldots, p_{k-1})$ ha entropia data da

$$-\sum_{j=0}^{k-1} p_j \log p_j. \tag{5.1}$$

Se si pone

$$A_i := \big\{ \{x_n : n \in \mathbf{Z}\} : x_0 = i \big\},$$

allora $\pi := \{A_0, A_1, \ldots, A_{k-1}\}$ costituisce una partizione di

$$\Omega = \{0, 1, \ldots, k-1\}^{\mathbf{Z}}.$$

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

$$h(T) = h(T, \mathcal{A}) = \lim_{n\to+\infty} \frac{1}{n} H\left(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-(n-1)}\mathcal{A}\right).$$

Gli elementi della partizione $\pi\left(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-(n-1)}\mathcal{A}\right)$ hanno la forma

$$\begin{aligned} &A_{i(0)} \cap T^{-1}A_{i(1)} \cap \cdots \cap T^{-(n-1)}A_{i(n-1)} \\ &\qquad = \big\{\{x_n\} : x_0 = i(0), x_1 = i(1), \ldots, x_{n-1} = i(n-1)\big\}, \end{aligned}$$

insieme che ha misura $p_{i(0)}\, p_{i(1)} \cdots p_{i(n-1)}$. Perciò

$$\begin{aligned} &H\big(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \cdots \vee T^{-(n-1)}\mathcal{A}\big) \\ &\quad = -\sum_{i(0),i(1),\ldots,i(n-1)=0}^{n-1} p_{i(0)}\, p_{i(1)} \cdots p_{i(n-1)} \log p_{i(0)}\, p_{i(1)} \cdots p_{i(n-1)} \\ &\quad = -\sum_{i(0),i(1),\ldots,i(n-1)=0}^{n-1} p_{i(0)}\, p_{i(1)} \cdots p_{i(n-1)} \times \\ &\qquad\qquad \times \big\{ \log p_{i(0)} + \log p_{i(1)} + \cdots + \log p_{i(n-1)} \big\} \\ &\quad = -n \sum_{j=0}^{n-1} p_j \log p_j. \end{aligned}$$

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

$$A_i := \big\{ \{x_n : n \in \mathbf{Z}\} : x_0 = i \big\}.$$

Adesso l'elemento tipico della partizione $\pi\left(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{-(n-1)}\mathcal{A}\right)$ ha misura eguale a $p_{i(0)}\, p_{i(0)\,i(1)} \cdots p_{i(n-1)\,i(n)}$, sicché

$$\begin{aligned}
&H\big(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{-(n-1)}\mathcal{A}\big) \\
&\quad = - \sum_{i(0), i(1), \dots, i(n-1) = 0}^{k-1} p_{i(0)}\, p_{i(0)\,i(1)} \cdots p_{i(n-1)\,i(n)} \times \\
&\qquad\qquad \times \log \left\{ p_{i(0)}\, p_{i(0)\,i(1)} \cdots p_{i(n-1)\,i(n)} \right\} \\
&\quad = - \sum_{i(0), i(1), \dots, i(n-1) = 0}^{k-1} p_{i(0)}\, p_{i(0)\,i(1)} \cdots p_{i(n-1)\,i(n)} \times \\
&\qquad\qquad \times \left\{ \log p_{i(0)} + \log p_{i(0)\,i(1)} + \dots + \log p_{i(n-1)\,i(n)} \right\} \\
&\quad = - \sum_{i(0), i(1), \dots, i(n-1) = 0}^{k-1} p_{i(0)}\, p_{i(0)\,i(1)} \cdots p_{i(n-1)\,i(n)} \log p_{i(0)} \\
&\qquad - \sum_{i(0), i(1), \dots, i(n-1) = 0}^{k-1} p_{i(0)}\, p_{i(0)\,i(1)} \cdots p_{i(n-1)\,i(n)} \log p_{i(0)\,i(1)} - \dots \\
&\qquad - \sum_{i(0), i(1), \dots, i(n-1) = 0}^{k-1} p_{i(0)}\, p_{i(0)\,i(1)} \cdots p_{i(n-1)\,i(n)} \log p_{i(n-1)\,i(n)} \\
&\quad = - \sum_{i=0}^{k-1} p_i \log p_i - n \sum_{i=0}^{k-1} p_i \sum_{j=0}^{k-1} p_{ij} \log p_{ij} .
\end{aligned}$$

Ne consegue che l'entropia della traslazione markoviana bilatera $-(p, \Pi)$ è eguale a

$$-\sum_{i=0}^{k-1} p_i \sum_{j=0}^{k-1} p_{ij} \log p_{ij} .$$