Complessità Caos Informazione

Claudio Bonanno

2006

Indice

Capitolo 1

Sorgenti e informazione

L'approccio matematico alla teoria delle comunicazioni si può fare risalire all'articolo di C.E. Shannon apparso nel 1948 sul Bell System Technology Journal [Sh]. Nel corso degli anni tale articolo ha dimostrato la sua importanza non solo nell'ambito della teoria delle comunicazioni, ma ha anche fornito idee per lo sviluppo di un approccio statistico alla teoria dei sistemi dinamici. Infatti il concetto di "entropia", oggi comune nella teoria dei sistemi dinamici e fino ad allora usato esclusivamente nella meccanica statistica, è stato introdotto nella teoria dell'informazione in tale articolo, e poi da lì è stato successivamente esteso alla teoria dei sistemi dinamici. In questo primo capitolo seguiamo allora la nascita della teoria matematica delle comunicazioni secondo l'approccio di Shannon.

Per iniziare vanno chiariti due importanti concetti. Seguendo le parole di Shannon [Sh]

> "Il problema fondamentale della comunicazione è quello di riprodurre esattamente o approssimativamente in un punto un messaggio selezionato in un altro punto. Spesso tale messaggio ha un significato...Questi aspetti semantici della comunicazione non hanno rilevanza per il problema ingegneristico. L'aspetto significativo è che il messaggio è uno selezionato da un insieme di possibili messaggi. Il sistema deve essere creato per operare per ogni possibile scelta, non solo per quella effettivamente fatta poiché tale scelta è ignota al momento della creazione del sistema."

Il significato del messaggio diventa quindi irrilevante per la trasmissione dello stesso, ma ad avere importanza sono le caratteristiche "strutturali" del messaggio, quindi ad esempio la sua lunghezza, il tipo di simboli che

contiene, e così via. Un messaggio viene spesso scelto in un insieme finito di possibili messaggi, e questo permette un'interpretazione del contenuto di informazione del messaggio. Citando ancora Shannon [Sh]

> "Se il numero di messaggi nell'insieme (di possibili messaggi) è finito, allora questo numero o qualsiasi funzione monotona di tale numero può essere considerata come una misura dell'informazione prodotta quando un messaggio è scelto dall'insieme, essendo tutte le scelte equivalenti...La scelta più naturale è la funzione logaritmica."

Uno dei principali vantaggi nella scelta della funzione logaritmica è legato alla sua proprietà di trasformare prodotti in somme. Questo permette di realizzare formalmente l'idea intuitiva che se un messaggio è ottenuto come unione di due pezzi scelti in due insiemi (uguali o diversi), allora l'informazione del messaggio totale è la somma delle informazioni dei singoli pezzi.

Seguendo l'idea di Shannon il contenuto di informazione di ogni singola lettera dell'alfabeto italiano sarà allora $\log 21$ (una lettera è infatti un possibile messaggio scelto nell'insieme dato dall'alfabeto). Per una lettera dell'alfabeto inglese invece l'informazione è $\log 26$. Quindi possiamo concludere che parlando in inglese si fornisce più informazione. Ovviamente questa conclusione vale nelle ipotesi fissate prima, ossia nel caso di messaggi il cui significato non è importante e nel caso di scelte equivalenti, supponendo come messaggi le singole lettere di ogni parola. Questa conclusione vale quindi se ci consideriamo "produttori" di lettere casuali! Vedremo come da queste semplici considerazioni e dalla loro generalizzazione a "produttori" di parole e poi di frasi (inserendo quindi non un significato ma una correlazione nella successione di lettere) si ottiene il concetto di "entropia" di una sorgente di informazione.

Prima di chiarire il concetto di sorgente di informazione e di darne le caratteristiche principali per i nostri scopi, resta da chiarire un punto: la scelta della base nella funzione logaritmica. Tale scelta corrisponde alla scelta di un'unità di misura per l'informazione. Essendo abituati a considerare l'informazione trasmessa attraverso messaggi binari, scegliamo 2 come base del logaritmo. Tuttavia tale scelta è assolutamente irrilevante per lo sviluppo della teoria, e basta ricordare la semplice relazione

$$\log_b a = \frac{\log_c a}{\log_c b}$$

che vale per tre numeri positivi a, b, c.

1.1 Le sorgenti di informazione

Un sistema di comunicazione è formato da 5 parti:

- una *sorgente di informazione*, che produce un messaggio o una serie di messaggi da comunicare. La natura dei messaggi può essere varia. Noi ci limitiamo al caso di messaggi discreti, ossia una successione di simboli;
- un *trasmettitore*, che opera sul messaggio in modo da produrre un segnale che può essere trasmesso attraverso il *canale*;
- il *canale*, che rappresenta il mezzo fisico in cui viaggia il segnale, come ad esempio un cavo telefonico o una banda di frequenze radio;
- un *ricevitore*, che effettua l'operazione inversa del trasmettitore, ricostruendo il messaggio dal segnale;
- un *destinatario* del messaggio.

Supponiamo che la sorgente di informazione produca messaggi discreti e che il tempo necessario per la produzione di un simbolo sia fissato e uguale per tutti. In questo modo è possibile fissare come unità di tempo il tempo necessario per la produzione di un simbolo. Quindi il tempo diventa discreto e in n unità di tempo vengono prodotti n simboli. Supponiamo inoltre che non ci sia rumore nella trasmissione del segnale attraverso il canale, e che quindi le proprietà della comunicazione dipendano esclusivamente dalla sorgente di informazione. Parleremo quindi delle caratteristiche della sorgente di informazione, intendendo in generale quelle della comunicazione.

Enunciamo e dimostriamo innanzitutto un teorema che fornisce uno strumento fondamentale per la dimostrazione di molti risultati.

Teorema 1.1. *Sia $(a_n)_n$ una successione di numeri reali tale che $a_{m+n} \leq a_m + a_n$ per ogni m, n interi. Allora*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \left\{ \frac{a_n}{n} \right\}$$

Dimostrazione. Innanzitutto notiamo che per ogni $n \geq 1$ vale

$$a_n \leq a_1 + a_{n-1} \leq \cdots \leq n a_1$$

quindi la successione $\left(\frac{a_n}{n}\right)_n$ è limitata dall'alto da a_1. Sia $l = \inf_n \left\{ \frac{a_n}{n} \right\}$, allora per ogni $\varepsilon > 0$ esiste un $N \geq 1$ tale che $a_N \leq N(l + \varepsilon)$. Inoltre per

ogni $n \geq 1$ possiamo scrivere $n = kN + r$ dove $k \geq 0$ e $0 \leq r < N$. Allora

$$\frac{a_n}{n} \leq \frac{ka_N + a_r}{n} \leq \frac{kN(l+\varepsilon)}{kN+r} + \frac{\max\{a_1, \ldots, a_{N-1}\}}{kN+r}$$

da cui

$$l \leq \liminf_{n\to\infty} \frac{a_n}{n} \leq \limsup_{n\to\infty} \frac{a_n}{n} \leq l + \varepsilon$$

e il teorema segue dall'arbitrarietà di ε. □

Utilizziamo subito il teorema precedente per definire la capacità di una sorgente.

Definizione 1.1. Si definisce *capacità* C della sorgente di informazione il limite (quando esiste)

$$C = \lim_{n\to\infty} \frac{\log M(n)}{n}$$

dove $M(n)$ è il numero di possibili messaggi prodotti dalla sorgente in n unità di tempo.

L'esistenza del limite nella definizione di capacità si può ottenere applicando alla successione $(\log M(n))_n$ il Teorema 1.1. Basta infatti notare che basta supporre che $M(m+n) \leq M(m)M(n)$ per ogni m, n interi (infatti è ragionevole pensare che per una sorgente di informazione, parti di un possibile messaggio siano singoli possibili messaggi).

Sia $\mathcal{A}$ l'insieme di possibili simboli prodotti da una sorgente. Indichiamo con $d(\mathcal{A})$ la cardinalità di $\mathcal{A}$. Nel caso di una sorgente che produca messaggi in italiano, l'insieme $\mathcal{A}$ conterrà l'alfabeto italiano, più lo spazio e i simboli per la punteggiatura (virgole, punti, ecc.), e i vari simboli fonetici necessari (accenti, apostrofi, ecc.).

Per una sorgente di informazione senza restrizioni sui possibili messaggi prodotti si ottiene $C = \log d(\mathcal{A})$. Infatti tutti i possibili messaggi prodotti in n unità di tempo sono $d(\mathcal{A})^n$, tutti i messaggi di lunghezza n. Notiamo inoltre che $\log d(\mathcal{A})$ è anche la massima capacità per una sorgente che produca simboli dell'insieme $\mathcal{A}$.

Tuttavia possono esistere delle restrizioni sui possibili messaggi prodotti da una sorgente. Per esempio per sorgenti che producano messaggi di senso compiuto in una qualche lingua, o che seguano regole precise dettate da una codifica usata, come nel caso del telegrafo, in cui il simbolo spazio non può essere seguito da un altro spazio (in qualche caso il concetto di capacità deve essere definito tramite un limite superiore).

È necessario a questo punto introdurre una formalizzazione per le sorgenti di informazione. Nelle ipotesi che abbiamo fatto possiamo considerare una sorgente di informazione come un processo stocastico $\mathbb{Y} = (Y_n)_n$ discreto[1] e stazionario, in cui le variabili aleatorie assumono valori in $\mathcal{A}$. Le proprietà statistiche della sorgente sono quindi le proprietà statistiche del processo, e le restrizioni sui messaggi possibili dipendono dalle proprietà di correlazione delle variabili aleatorie Y_n. Tale formalizzazione per la sorgente di informazione permette l'applicazione diretta della teoria statistica dei sistemi dinamici. Rimandiamo quindi al Capitolo 2 la formalizzazione di questi concetti.

1.2 L'entropia di Shannon

In questa sezione affrontiamo il problema dell'introduzione del concetto di *entropia* nella teoria dell'informazione. Le proprietà dell'entropia sono studiate nel Capitolo 2, dove le sorgenti di informazione vengono studiate come caso particolare di sistemi dinamici.

Supponiamo di ricevere messaggi da una sorgente di informazione con le proprietà descritte nella sezione precedente. Ci possiamo chiedere quale sia l'informazione prodotta dalla sorgente, oppure equivalentemente quale sia la nostra incertezza sul messaggio che riceveremo.

Una prima risposta parziale si ottiene usando il concetto di capacità. Se infatti supponiamo che la nostra sorgente produca $M(n) \sim 2^{nC}$ messaggi equivalenti di lunghezza n, possiamo dire, usando il concetto di informazione contenuto in un messaggio introdotto da Shannon, che un messaggio lungo n ha un contenuto di informazione $\log M(n) \sim nC$. Quindi la sorgente produce C bits di informazione per unità di tempo.

Il ragionamento precedente vale nel caso in cui non conosciamo altro della sorgente di informazione, così come il concetto di informazione definito da Shannon valeva per messaggi equivalenti. Spesso invece è possibile dire qualcosa di più. Se per esempio la sorgente emette messaggi scritti con simboli dell'alfabeto $\mathcal{A} = \{1, \ldots, N\}$, spesso è possibile supporre di conoscere le probabilità $p_1, \ldots, p_N$ con cui questi simboli vengono generati. Questo è per esempio il caso dei linguaggi, per cui è possibile studiare le frequenze delle diverse lettere. Osserviamo che stiamo ancora considerando il caso in cui i simboli vengono emessi indipendentemente dal precedente (siamo ancora solo dei "produttori" di lettere).

[1] Il tempo è discretizzato usando l'unità di tempo per la produzione di un simbolo.

Se vogliamo ottenere una misura dell'informazione prodotta dalla produzione di un simbolo da parte di una tale sorgente, dobbiamo supporre l'esistenza di una funzione $H(p_1, \ldots, p_N)$ che abbia particolari caratteristiche. Shannon [Sh] impone che

(1) H sia continua nelle variabili;

(2) se tutte le probabilità sono uguali tra loro e quindi $p_i = \frac{1}{N}$ per ogni $i = 1, \ldots, N$, allora H deve essere una funzione crescente in N;

(3) se la scelta del messaggio si spezza in due scelte successive, allora H deve essere la media pesata delle H_i, le funzioni legate alla seconda scelta, più la funzione legata alla prima scelta.

Esempio 1.1. Consideriamo il caso di una stringa binaria $s = (s_1 s_2)$ (sull'alfabeto $\{0, 1\}$) di lunghezza 2, prodotta da una sorgente che abbia le seguenti probabilità sulle stringhe lunghe 2, che costituiscono i messaggi:

$$p_{00} = \frac{1}{12} \ , \ p_{01} = \frac{1}{4} \ , \ p_{10} = \frac{1}{3} \ , \ p_{11} = \frac{1}{3}$$

D'altra parte, queste stringhe possono considerarsi prodotte anche tramite la seguente operazione: si produce s_1 secondo le probabilità $p_0 = \frac{1}{3}$ e $p_1 = \frac{2}{3}$, e poi si produce s_2 con le probabilità condizionate $p(s_2 = 0|s_1 = 0) = \frac{1}{4}$, $p(s_2 = 1|s_1 = 0) = \frac{3}{4}$, e $p(s_2 = 0|s_1 = 1) = p(s_2 = 1|s_1 = 1) = \frac{1}{2}$.

Allora l'ipotesi (3) afferma che deve sussistere l'uguaglianza

$$H\left(\frac{1}{12}, \frac{1}{4}, \frac{1}{3}, \frac{1}{3}\right) = H\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{1}{3}\, H\left(\frac{1}{4}, \frac{3}{4}\right) + \frac{2}{3}\, H\left(\frac{1}{2}, \frac{1}{2}\right) \qquad \triangle$$

Teorema 1.2 ([Sh], Appendice 2)**.** *Sia $H(p_1, \ldots, p_N)$ una funzione che soddisfa le ipotesi (1)-(3). Allora deve essere*

$$H(p_1, \ldots, p_N) = -k \sum_{i=1}^{N} p_i \log p_i \tag{1.1}$$

con k costante positiva arbitraria.

Dimostrazione. Poniamo $h(N) = H(\frac{1}{N}, \ldots, \frac{1}{N})$. Per l'ipotesi (2), $h(N)$ è crescente in N. Per l'ipotesi (3) invece $h(N^m)$, ossia l'informazione relativa a N^m scelte equivalenti, è uguale a $mh(N)$. Basta infatti ripetere induttivamente l'uguaglianza

$$h(N^m) = h(N^{m-1}) + \sum_{j=1}^{N^{m-1}} \frac{1}{N^{m-1}}\, h(N)$$

Supponiamo che h sia una funzione definita su $\mathbb{R}^+ = (0, +\infty)$, crescente e tale che $h(t^n) = nh(t)$ per ogni $t \in \mathbb{R}^+$ e ogni $n \in \mathbb{N}$. Dati due numeri reali positivi s e t, scegliamo due interi m, n tali che $s^m \leq t^n < s^{m+1}$. Allora

$$\begin{array}{rcccl} m \log s & \leq & n \log t & < & (m+1) \log s \\ m\ h(s) & \leq & n\ h(t) & < & (m+1)\ h(s) \end{array}$$

da cui dividendo rispettivamente per $n \log s$ e per $nh(s)$, si ottiene

$$\begin{aligned} \left| \frac{m}{n} - \frac{\log t}{\log s} \right| &\leq \frac{1}{n} \\ \left| \frac{m}{n} - \frac{h(t)}{h(s)} \right| &\leq \frac{1}{n} \end{aligned}$$

Quindi, mettendo insieme le due equazioni si ottiene

$$\left| \frac{\log t}{\log s} - \frac{h(t)}{h(s)} \right| \leq \varepsilon$$

con ε arbitraria. Quindi $h(t) = k \log t$ con k costante positiva, per rispettare l'ipotesi (2).

Supponiamo di avere una scelta tra n oggetti equivalenti, in cui ci siano delle ripetizioni, e gli oggetti diversi siano $N < n$. Allora la scelta tra gli n oggetti equivalenti si può spezzare nella scelta tra N oggetti con probabilità $p_i = \frac{r_i}{n}$, per $i = 1, \ldots, N$, dove r_i è il numero di volte che compare l'oggetto i-esimo, e poi nella scelta tra r_i oggetti equivalenti, nel caso in cui al primo passo venga scelto l'i-esimo oggetto. Applicando a tale ragionamento l'ipotesi (3), deve valere

$$k \log n = H(p_1, \ldots, p_N) + k \sum_{i=1}^{N} p_i \log r_i$$

Da questa uguaglianza si ricava

$$H(p_1, \ldots, p_N) = -k \sum_{i=1}^{N} p_i \log p_i$$

per p_i razionali. L'ipotesi (1) di continuità permette poi l'estensione a tutti i reali positivi. □

La funzione H con $k = 1$ è chiamata *entropia*[2] di una sorgente di informazione che produce messaggi in cui ogni simbolo non è correlato al precedente, quindi di un processo stocastico stazionario $\mathbb{Y}$ con variabili aleatorie indipendenti.

Generalizzando il concetto di contenuto di informazione di un messaggio dato nel caso di messaggi equivalenti, possiamo dire che nel caso precedente, ogni lettera dell'alfabeto $\mathcal{A}$ ha un contenuto di informazione pari a $\log \frac{1}{p}$, dove p è la probabilità che la lettera sia prodotta dalla sorgente. Quindi maggiore è la probabilità che una lettera sia prodotta, minore è l'informazione che contiene, quindi minore è la sorpresa del destinatario del messaggio nel riceverla, e minore è l'incertezza sul messaggio che viene rimossa. Osserviamo che nel caso in cui le lettere siano equiprobabili si ricade nel concetto introdotto da Shannon. L'entropia H dell'equazione (1.1) si può quindi interpretare come l'informazione media di una lettera dell'alfabeto $\mathcal{A}$. Di nuovo nel caso di lettere equiprobabili, l'entropia è uguale alla capacità della sorgente.

Possiamo ora generalizzare il concetto di entropia a una qualsiasi sorgente, utilizzando le probabilità di emissione dei messaggi di lunghezza n. Supponiamo che l'insieme dei messaggi possibili di lunghezza n abbia cardinalità $M(n)$, e indichiamo con m_i^n i vari messaggi.

Definizione 1.2. Data una sorgente di informazione, si chiama *entropia di Shannon* della sorgente, il limite

$$h = \lim_{n\to\infty} \; -\frac{1}{n} \sum_{i=1}^{M(n)} p(m_i^n) \log p(m_i^n) \tag{1.2}$$

dove $p(m_i^n)$ indica la probabilità che il messaggio m_i^n sia prodotto.

L'esistenza del limite in (1.2) e le proprietà dell'entropia di Shannon verranno studiate in un ambito più generale nel Capitolo 2.

Notiamo che le funzioni

$$H_n = -\sum_{i=1}^{M(n)} p(m_i^n) \log p(m_i^n)$$

possono considerarsi le funzioni entropie di una sorgente che produca messaggi di lunghezza n con le probabilità $p(m_i^n)$ assegnate. Infatti tutte le H_n

[2]Il nome è stato scelto da Shannon per la somiglianza con la formula dell'entropia trovata da Boltzmann nella meccanica statistica.

sono uguali alla funzione H dell'equazione (1.1), in cui i messaggi di lunghezza n sostituiscono i simboli. Se quindi H_n è l'informazione media contenuta in un messaggio lungo n, allora $\frac{1}{n}H_n$ è l'informazione media contenuta in uno dei simboli di un messaggio lungo n. *L'entropia di Shannon è quindi il contenuto medio di informazione di un simbolo prodotto da una sorgente di informazione.*

Tornando all'equivalenza con i linguaggi, possiamo dire che le entropie H_n si riferiscono a "produttori" di frasi con n simboli, e quindi approssimano il vero contenuto di informazione di un linguaggio.

1.3 Codici

Nelle sezioni precedenti, siamo partiti dal problema pratico della comunicazione di messaggi e ci siamo spostati verso un approccio teorico, che ha perso il contatto con il problema reale. Il concetto di limite per il tempo che tende all'infinito è una pura idealizzazione, che non ha nessun contatto con la realtà! Adesso, prima di idealizzare ancora di più i concetti studiati entrando nel mondo dei sistemi dinamici, torniamo al problema concreto della trasmissione di segnali. Se dobbiamo trasmettere un messaggio, qual è il modo più economico di farlo? Nelle ipotesi che la comunicazione avvenga senza disturbi esterni nel canale, il compito di risparmiare sulla trasmissione di un messaggio è affidato al trasmettitore, che riceve il messaggio dalla sorgente e deve "trasformarlo" adeguatamente prima di inviarlo. Tale operazione è una *codifica* del messaggio, e il problema è quello di trovare il *codice* più adatto alla sorgente.

Sia $\{0,1\}^n$ l'insieme di tutte le stringhe di lunghezza n con simboli in $\{0,1\}$, e indichiamo con $\{0,1\}^*$ l'insieme di tutte le stringhe finite, ossia

$$\{0,1\}^* := \bigcup_{n=0}^{\infty} \{0,1\}^n$$

Indichiamo con $\lambda \in \{0,1\}^0$ la stringa vuota. Su $\{0,1\}^*$ esiste un'operazione binaria, detta *concatenazione*, che associa a una coppia (s,t) in $\{0,1\}^* \times \{0,1\}^*$ la stringa st in cui i simboli di s vengono seguiti da quelli di t. Tale operazione ha la proprietà associativa e ha un elemento neutro, la stringa vuota λ. Ovviamente la concatenazione non è commutativa.

Diamo adesso una nozione formale di codice. In seguito consideriamo solo codici costruiti su $\{0,1\}^*$.

Definizione 1.3. Dato un insieme M di messaggi (al più numerabile), un *codice* per M è una funzione $D : \{0,1\}^* \to M$. Le stringhe nel dominio di D sono dette *stringhe codice*, i messaggi nell'immagine di D sono detti *messaggi sorgente*. L'operazione $E := D^{-1}$ è la *codifica*, $E(m) = \{s \in \{0,1\}^* : D(s) = m\}$, che associa a un messaggio sorgente tutte le stringhe codice associate. La funzione D è quindi la *decodifica* di una stringa.

Supponiamo che l'insieme dei messaggi sia l'insieme $\mathbb{N}_0$ dei numeri naturali con in più lo 0. Data una stringa $s \in \{0,1\}^n$ indichiamo con s_i per $i = 1, \ldots, n$ i simboli di s, quindi $s = (s_1 s_2 \ldots s_n)$. Allora un possibile codice è dato dalla funzione

$$D(s) = 2^n - 1 + \sum_{i=1}^{n} s_{n-i+1}\ 2^{i-1} \qquad s \in \{0,1\}^n \tag{1.3}$$

con $D(\lambda) = 0$. Quindi per inviare il numero (5) per esempio, inviamo la stringa binaria $s = (10)$. Questo codice è bigettivo su $\mathbb{N}_0$, quindi la codifica associata è una funzione. Dato un numero naturale n indichiamo da ora in poi con $\bar{n}$ la stringa codice di n secondo questa codifica. Sia $|n|$ la lunghezza della stringa codice $\bar{n}$, si verifica che

$$|n| = \lfloor \log(n+1) \rfloor \tag{1.4}$$

Tuttavia il codice (1.3) crea un problema nella decodifica. Infatti la concatenazione di due stringhe si può interpretare anche come un'unica stringa. Così (101) si può interpretare sia come $(5, 2)$ sia come (12). Osserviamo come anche la rappresentazione decimale posizionale dei numeri naturali sia una codifica con questo problema, abbiamo infatti avuto bisogno della virgola per distinguere 51 da 5 e 1. Esistono codici che evitano questo problema.

Definizione 1.4. Date due stringhe $s, t \in \{0,1\}^*$, si dice che s è *prefisso* di t se esiste una stringa $w \in \{0,1\}^*$ tale che $t = sw$. Un sottoinsieme di $\{0,1\}^*$ si dice *libero da prefissi* se nessun suo elemento è prefisso di un altro. Un codice si dice *codice prefisso* se il suo dominio è libero da prefissi.

Un codice prefisso permette quindi di poter distinguere due stringhe concatenate senza bisogno di usare un simbolo extra come separatore (come la virgola nel sistema decimale). Usando un codice prefisso, il destinatario legge il messaggio sapendo quando finisce la codifica di un simbolo. Un esempio di codice prefisso per $\mathbb{N}_0$ è dato da $E(n) = 1^n 0$, ossia n simboli "1" seguiti da uno "0". Quando il destinatario legge un simbolo "0" capisce quindi che inizia una nuova stringa, quindi un nuovo numero codificato. La

codifica del messaggio $(3, 5, 0, 1)$ è la concatenazione delle diverse codifiche, ed è data da $s = (1110111110010)$. Questo codice non è tuttavia molto economico, infatti $|E(n)| = n + 1$, dove $|\cdot|$ indica la lunghezza binaria della stringa $E(n)$. Iterando questo tipo di codifica è possibile costruire una successione di codici sempre più economici per $\mathbb{N}_0$. Definiamo la seguente successione di codifiche

$$E_i(n) = \begin{cases} 1^n 0 & \text{se } i = 0 \\ E_{i-1}(|n|)\ \bar{n} & \text{se } i > 0 \end{cases} \tag{1.5}$$

Quindi $E_1(n) = E_0(|n|)\bar{n} = 1^{|n|}0\bar{n}$, e si ha

$$|E_1(n)| = 2|n| + 1 = 2\lfloor \log(n+1) \rfloor + 1$$

Supponiamo di ricevere il messaggio (111011011011) con il codice E_1, allora deduciamo che il messaggio originale è $(13, 6)$. Ancora più economico è il codice prefisso $E_2(n)$ per cui

$$\begin{aligned} E_2(n) &= E_1(|n|)\bar{n} = 1^{||n||}0\ \overline{|n|}\ \bar{n} \\ |E_2(n)| &= 2||n|| + 1 + |n| \sim \log(n) + 2\log\big(\log(n)\big) + 1 \end{aligned}$$

dove $||n||$ indica la lunghezza della stringa che rappresenta il numero naturale $|n|$ secondo il codice (1.3). Usando il codice E_2 il messaggio $(13, 6)$ viene codificato nella stringa (1100011010111). Notiamo che il codice E_2 è più economico di E_1 per numeri non troppo piccoli. Usare un codice prefisso è sicuramente meno economico del codice non prefisso di equazione (1.3), ma la differenza è abbastanza piccola, dell'ordine del logaritmo del logaritmo del numero naturale.

I codici che abbiamo considerato finora sui numeri naturali non tengono conto delle proprietà statistiche della sorgente di informazione. Supponiamo di voler costruire un codice per la lingua inglese. Un esempio noto è il *codice Morse*. Il codice Morse è un codice binario, in cui ad ogni lettera dell'alfabeto inglese viene associata in maniera univoca una successione binaria (punti e linee possono essere sostituiti da "0" e "1"). Nella lingua inglese si notano differenze nelle frequenze di uso delle lettere. Per esempio la lettera "e" ha una frequenza di 0.12, mentre la lettera "w" ha una frequenza di 0.02. Cercando di risparmiare sul numero di simboli da trasmettere, conviene quindi associare ad "e" una stringa breve, mentre a "w" si può associare una stringa più lunga. Da questo ragionamento si deduce che deve esserci un collegamento tra l'entropia di una sorgente e il codice più economico.

Esempio 1.2 (Codice di Shannon-Fano). Supponiamo di voler costruire un codice per un insieme di messaggi M, che contenga n elementi, e per una sorgente che produca i messaggi di M con probabilità $p_1, \dots, p_n$. Supponiamo che i messaggi siano ordinati in ordine descrescente di probabilità. Il codice di Shannon-Fano si ottiene tramite una codifica $E : M \to \{0,1\}^*$ tale che

$$-\log p_k \leq |E(m_k)| < 1 - \log p_k$$

prendendo i primi $|E(m_k)|$ simboli dell'espansione binaria di

$$P_k = p_1 + \cdots + p_{k-1}$$

per $k > 1$ e $P_1 = 0$. Così m_1, il messaggio più probabile, avrà la codifica più breve, e la lunghezza della codifica è non decrescente. Inoltre siamo sicuri che le stringhe codice siano tutte diverse, grazie al fatto che $P_j \geq P_k + 2^{-|E(m_k)|}$ per ogni $j \geq k+1$. Questo dimostra anche che P_k e P_j, per ogni $j \geq k+1$, differiscono in almeno un simbolo tra i primi $|E(m_k)|$, quindi il codice di Shannon-Fano è un codice prefisso.

Questo codice ha anche un'altra importante proprietà. La lunghezza media di una stringa codice per i messaggi di M è $L = \sum_{k=1}^{n} p_k |E(m_k)|$, che verifica

$$H(p_1, \dots, p_n) \leq L < H(p_1, \dots, p_n) + 1$$

Vedremo che questa stima è quanto di meglio si può ottenere per un codice prefisso. $\triangle$

Abbiamo osservato che un codice prefisso usa in generale stringhe codice più lunghe di altri codici. Per le lunghezze delle stringhe codice di un codice prefisso vale la seguente stima.

Teorema 1.3 (Disuguaglianza di Kraft)**.** *Sia $(l_n)_n$ una successione di numeri naturali. Esiste un codice prefisso le cui stringhe codice hanno lunghezze date dai numeri di questa successione se e solo se*

$$\sum_{n=1}^{\infty} 2^{-l_n} \leq 1$$

Dimostrazione. La dimostrazione si basa sulla rappresentazione binaria dei numeri reali dell'intervallo $[0,1)$. A una stringa binaria $s \in \{0,1\}^n$ possiamo far corrispondere l'intervallo $J_s = [0.s, 0.s + 2^{-n})$. Inoltre date due stringhe s, t di un codice prefisso si ha $J_s \cap J_t = \emptyset$.

Per un codice prefisso si ha quindi che le stringhe s individuano intervalli J_s a due a due disgiunti di lunghezza $2^{-|s|}$, quindi $\sum_{n=1}^{\infty} 2^{-l_n}$ è la lunghezza

dell'unione di tutti gli intervalli J_s, che sono tutti contenuti in $[0,1)$. Quindi vale la disuguaglianza.

Sia ora data una successione $(l_n)_n$ che verifica la disuguaglianza, e supponiamo che sia non decrescente. Costruiamo una successione $(I_n)_n$ di intervalli adiacenti contenuti in $[0,1)$ di lunghezza 2^{-l_n}, con I_1 con estremo sinistro in 0. Scegliamo come stringhe del codice prefisso i primi l_n simboli dell'espansione binaria della parte frazionaria degli estremi sinistri degli intervalli I_n. □

Sia $M = \{m_1, m_2, \dots\}$ un insieme di messaggi prodotti da una sorgente con probabilità $p_1, p_2, \dots$. Sia $D : \{0,1\}^* \to M$ un codice prefisso con codifica unica per ogni messaggio.

Definizione 1.5. Si chiama *lunghezza media delle stringhe codice* di un codice D, la media $L_D := \sum_{i=1}^{d(M)} p_i \, |D^{-1}(m_i)|$. Definiamo poi *minima lunghezza media delle stringhe codice* il numero

$$L := \min \{L_D : D \text{ è codice prefisso}\}$$

Un codice prefisso D per cui $L_D = L$ si chiama *ottimale* rispetto alla sorgente.

L'idea che sta dietro la realizzazione di un codice prefisso ottimale rispetto a una sorgente è quella del codice di Shannon-Fano.

Teorema 1.4 ([Sh], Noiseless Coding Theorem)**.** *Se $H = H(p_1, p_2, \dots)$ è la funzione entropia della sorgente che produce messaggi di M con probabilità p_i, allora la minima lunghezza media delle stringhe codice L verifica*

$$H \le L \le H + 1$$

Dimostrazione. Poniamo $l_n = \lceil -\log p_n \rceil$. Allora $1 \ge \sum_{i=1}^{d(M)} p_i \ge \sum_{i=1}^{d(M)} 2^{-l_i}$, quindi per il Teorema 1.3 esiste un codice prefisso con stringhe codice lunghe l_n. Quindi

$$L \le \sum_{i=1}^{d(M)} p_i \, l_i \le \sum_{i=1}^{d(M)} p_i(1 - \log p_i) = 1 + H$$

Viceversa sia D un codice prefisso con stringhe codice di lunghezza $(l_n)_n$. Allora per la concavità del logaritmo si ha

$$-\sum_{i=1}^{d(M)} p_i \log p_i \le -\sum_{i=1}^{d(M)} p_i \log \frac{2^{-l_i}}{\sum_i 2^{-l_i}} = L_D + \left(\sum_{i=1}^{d(M)} p_i \right) \log \sum_i 2^{-l_i}$$

Infine, essendo D un codice prefisso, vale la disuguaglianza di Kraft, quindi $\sum_i 2^{-l_i} \leq 1$. Si ottiene quindi che $H \leq L_D$ per ogni codice prefisso D, quindi $H \leq L$. □

Cosa succede se non conosciamo la distribuzione di probabilità dei messaggi? È possibile trovare un codice che sia ottimale per qualsiasi sorgente su un insieme di messaggi? Le risposte a queste domande vanno ben oltre lo scopo di queste note, rimandiamo quindi a [LV] per una discussione di questi problemi. Ci limitiamo a fornire un esempio di codifica che risulta importante nel seguito delle note.

Esempio 1.3 (Frequency coding). Sia t una stringa di lunghezza n scritta con simboli di un alfabeto finito $\mathcal{A}$. Supponiamo di non conoscere le proprietà statistiche della sorgente che ha prodotto t, ma usiamo per una codifica di t le proprietà statistiche dei simboli di $\mathcal{A}$ in t. Siano k i simboli di $\mathcal{A}$ e indichiamo con $n_1, \ldots, n_k$ il numero di volte che ciascun simbolo è ripetuto in t. Noti i numeri (n_j), supponiamo di ordinare in maniera lessicografica le stringhe di lunghezza n che hanno le stesse frequenze di apparizione di simboli di $\mathcal{A}$. Allora per specificare t basta inviare i numeri (n_j) e il numero d'ordine di t nell'insieme di stringhe con la stessa statistica. I numeri naturali n_j possono essere codificati tramite il codice E_1. Trasmettiamo quindi la stringa codice $E_1(n_1) \ldots E_1(n_k)$, concatenata con la stringa $\bar{h}$, dove h è il numero d'ordine di t. Quindi la stringa codice finale s ha lunghezza

$$|s| \leq k\big(2\lfloor \log(n+1) \rfloor + 1\big) + \lfloor \log(h+1) \rfloor$$

dove il numero h verifica

$$h \leq \left(\begin{array}{c} n \\ n_1 \; n_2 \ldots n_k \end{array} \right) = \frac{n!}{n_1! \; n_2! \ldots n_k!}$$

Facendo tendere n all'infinito, mantenendo costanti le frequenze $p_j = \frac{n_j}{n}$, si ottiene che la lunghezza binaria di s si comporta asintoticamente come $n(-\sum_{i=1}^{k} p_i \log p_1)$ (approssimando i fattoriali usando la formula di approssimazione di Stirling $n! \sim \sqrt{2\pi n}\; n^n e^{-n}$). Quindi abbiamo costruito un codice per cui la lunghezza media delle stringhe codice (la lunghezza di s fratto la lunghezza della stringa originale) è asintoticamente uguale alla funzione entropia $H(p_1, \ldots, p_k)$. Se quindi t è prodotta da una sorgente con probabilità per i simboli di $\mathcal{A}$ date dai numeri p_i, siamo riusciti a realizzare il miglior codice prefisso possibile senza conoscere le proprietà statistiche della sorgente. Tale codifica si chiama *frequency coding* perché si basa sulla trasmissione delle frequenze di apparizione dei vari simboli. △

Esercizi

1.1. Dimostrare che la funzione entropia dell'equazione (1.1) verifica l'uguaglianza dell'esempio 1.1.

1.2. Dimostrare che per una sorgente di informazione senza restrizioni e con variabili aleatorie indipendenti, l'entropia di Shannon h è uguale alla funzione entropia dell'equazione (1.1).

1.3. Dimostrare l'uguaglianza (1.4).

1.4. Trovare le codifiche dei numeri naturali 10, 15, 20, secondo i codici E_1 ed E_2.

1.5. Trovare la decodifica della stringa (111001011010) secondo E_1, e della stringa (11001101011000011) secondo E_2.

Capitolo 2

Approccio statistico ai sistemi dinamici

Introduciamo in questo capitolo alcuni concetti classici riguardanti lo studio dei sistemi dinamici. Per trattazioni più complete rimandiamo per esempio a [CFS], [Ga], [KH], [Pe], [PY], [Wa].

2.1 Definizioni ed esempi base

Sia $(X, \mathcal{B})$ uno spazio metrico compatto con $\mathcal{B}$ la σ-algebra di Borel. Nel seguito consideriamo sempre questa situazione, in caso contrario verrà specificato.

Definizione 2.1. Chiamiamo *sistema dinamico* su X l'azione di $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ tramite una funzione $T : X \to X$ misurabile. L'azione è data dall'iterazione della funzione T, ossia è definita da

$$\mathbb{N}_0 \times X \ni (n, x) \mapsto T^n(x) \in X$$

in cui T^n indica la composizione di T per n volte, e T^0 è definita uguale all'identità. Nel caso in cui T sia invertibile e con inversa misurabile, si può considerare l'azione analoga di $\mathbb{Z}$ su X.

Nel seguito richiamiamo alcune semplici definizioni di base, relative al caso più generale di una funzione T non invertibile. Spesso per il caso invertibile basterà sostituire $\mathbb{N}_0$ con $\mathbb{Z}$. Nel caso invertibile si suppone che anche T^{-1} sia misurabile.

Definizione 2.2. Dato un punto $x \in X$, si definisce *orbita* di x rispetto al sistema dinamico (X, T), l'insieme $\mathcal{O}(x) = \{T^n(x) \ : \ n \in \mathbb{N}_0\}$.

Definizione 2.3. Un punto x si dice *periodico* se esiste un intero $p(x) > 0$ tale che $T^{p(x)}(x) = x$. L'intero $p(x)$ si chiama *periodo* se vale anche che per ogni $1 \leq k < p(x)$ si ha $T^k(x) \neq x$. Un punto periodico di periodo $p(x) = 1$ si dice *fisso*, verificando $T(x) = x$.

Un punto x si dice *definitivamente periodico* se esiste un intero $n \geq 0$ tale che $T^n(x)$ è periodico di periodo p, quindi $T^{n+p}(x) = T^n(x)$ e $T^{n+k}(x) \neq T^n(x)$ per ogni $1 \leq k < p$.

Definizione 2.4. Un insieme misurabile $B \subset X$ si dice *invariante* per il sistema dinamico (X, T) se $T^{-1}(B) = B$.

Definiamo alcuni sistemi dinamici che costituiscono gli esempi chiave che useremo per introdurre i vari concetti.

Esempio 2.1 (Mappa del panettiere). Introduciamo uno degli esempi più classici di sistema dinamico *caotico*. Vedremo che ha una definizione semplicissima, e tuttavia presenta un comportamento delle orbite che ha la massima impredicibilità.

Sia $X = [0, 1]/(0 \sim 1) \cong \mathbb{R}/\mathbb{Z} \cong S^1$ con la metrica euclidea indotta da $\mathbb{R}$. Definiamo la seguente funzione

$$T(x) = 2x \pmod 1 = \begin{cases} 2x & \text{per } 0 \leq x < \frac{1}{2} \\ 2x - 1 & \text{per } \frac{1}{2} \leq x < 1 \end{cases} \tag{2.1}$$

La prima formulazione è una funzione continua su S^1, mentre la seconda ha una discontinuità in $\frac{1}{2}$ ed è definita in $[0, 1)$. Tuttavia le due formulazioni sono perfettamente equivalenti. △

Esempio 2.2 (Rotazioni del cerchio). Consideriamo nuovamente lo spazio $X = [0, 1]/(0 \sim 1) \cong S^1$ e sia α un numero reale in $(0, 1)$. Ruotare il cerchio S^1 di un angolo dato equivale ad aumentare di una quantità fissa l'angolo di ogni punto del cerchio. Se quindi scriviamo per i punti del cerchio $z = e^{2\pi i\vartheta}$ con $\vartheta \in X$, la rotazione $\tilde{T}_\alpha$ è definita dalla funzione

$$\tilde{T}_\alpha(z) = e^{2\pi i\alpha} z = e^{2\pi i(\vartheta+\alpha)}$$

che definisce su X il seguente sistema dinamico

$$T_\alpha : X \to X \qquad T_\alpha(x) = x + \alpha \pmod 1 \tag{2.2}$$

Vedremo in seguito che le proprietà di questo sistema dinamico dipendono fortemente dalle proprietà aritmetiche di α, in particolare risulta fondamentale l'appartenenza di α ai razionali $\mathbb{Q}$.

Come nel caso della mappa del panettiere si può scrivere una formulazione di T_α discontinua su $[0, 1)$. △

Esempio 2.3 (Famiglia quadratica). Un esempio particolarmente significativo per la teoria dei sistemi dinamici è dato dalla famiglia quadratica. Questo esempio di sistema dinamico fu introdotto da Sir R. May nel 1976 per simulare la dinamica di una popolazione di individui. Si suppone che il numero di individui di una popolazione dipenda esclusivamente dal tasso di riproduzione e dal tasso di mortalità. Quest'ultimo si suppone determinato esclusivamente dalla limitatezza delle risorse, per cui esiste un numero massimo di individui che una popolazione può avere. Si può allora studiare l'andamento del rapporto tra il numero di individui e il suo valore massimo. Questo rapporto è un numero $x \in [0,1]$, per cui se $x = 0$ non ci sono individui, e se $x = 1$ si ottiene il numero massimo di individui, che non possono sopravvivere e quindi muoiono. Il sistema dinamico che cerchiamo è quindi regolato da una funzione $T : X \to X$, dove $X = [0,1]$, per cui se la popolazione ha un rapporto x all'inizio dell'evoluzione, avrà un rapporto $T^n(x)$ dopo n unità di tempo[1]. Le ipotesi considerate impongono per T che valga $T(0) = T(1) = 0$. Si considera la funzione

$$T_\lambda(x) = \lambda\ x\ (1 - x) \qquad \lambda \in [0,4] \tag{2.3}$$

dove la scelta dell'intervallo per il parametro λ serve solo a imporre che l'immagine di T sia contenuta in X.

Al variare del parametro λ questa famiglia di mappe presenta una ricchezza di proprietà che ne hanno fatto un esempio chiave nella teoria dei sistemi dinamici. In particolare per λ minore di un valore $\lambda_\infty = 3.56994567\ldots$ tutte le condizioni iniziali convergono verso orbite periodiche, di periodo crescente al crescere di λ. È interessante notare che i periodi presenti sono solo potenze di 2, per cui questo fenomeno si chiama *period doubling* o *cascata di raddoppiamento di periodo*. Il fenomeno (più ricco di quello qui descritto) fu scoperto da M. Feigenbaum nel 1978 ([Fe]) ed è un fenomeno universale per una classe di sistemi dinamici non lineari. Vedi [De] per una trattazione approfondita. $\triangle$

Esempio 2.4 (Famiglia di Pomeau-Manneville). Di grande interesse nella teoria dei sistemi dinamici è anche la famiglia di sistemi di Pomeau-Manneville. La provenienza di tale famiglia di sistemi è di natura fisica. Uno dei fenomeni ancora non del tutto compresi in fisica è la turbolenza. Negli esperimenti in dinamica dei fluidi si osserva che un fluido sotto certe condizioni (importante risulta la viscosità) passa da un regime cosiddetto

[1] La misura del tempo in maniera discreta è il solo caso realistico per l'esperimento che si considera.

"laminare", in cui la velocità del fluido misurata ha un comportamento regolare e prevedibile, a un regime "turbolento", in cui invece i valori misurati della velocità sono caotici, e viceversa, continuando poi ad alternare i due regimi. Per studiare questo fenomeno di "intermittenza", Y. Pomeau e P. Manneville introdussero intorno al 1980 ([PM]) una famiglia di sistemi dinamici su $X = [0, 1]$ che presentavano lo stesso fenomeno, ossia l'orbita di un punto generico alterna fasi regolari e caotiche. La definizione di questa famiglia di sistemi dipende da un parametro reale z ed è data da

$$T_z(x) = x + x^z \pmod 1 \qquad z > 1 \tag{2.4}$$

Notiamo che per $z = 1$ si ottiene la mappa del panettiere. Questa famiglia di sistemi ha la caratteristica di avere un punto fisso per $x = 0$ in cui $T'_z(0) = 1$. Tuttavia le orbite di punti vicini all'origine si allontanano, anche se molto lentamente. Questa fase di allontanamento riproduce la fase laminare. Quando poi un'orbita si allontana abbastanza dall'origine allora la sua evoluzione non è dissimile da quella di un'orbita della mappa del panettiere, raggiunge quindi il massimo grado di caoticità. E questa rappresenta la fase turbolenta. L'alternanza delle due fasi è assicurata da una probabilità positiva per ogni orbita di tornare infinite volte vicino all'origine, e anche a una distanza piccola a piacere. $\triangle$

Esempio 2.5 (Dinamica simbolica). Descriviamo infine il sistema dinamico simbolico. Questo sistema dinamico ha una particolare importanza nella nostra teoria, in quanto permette il collegamento tra un sistema dinamico generico e la teoria delle sorgenti di informazione.

Dato un insieme finito $\mathcal{A} = \{1, \dots, N\}$, che consideriamo l'alfabeto, indichiamo con $\mathcal{A}^n$ l'insieme delle stringhe s di lunghezza n i cui simboli sono lettere di $\mathcal{A}$. Indichiamo poi con $\mathcal{A}^*$ l'insieme di tutte le stringhe finite su $\mathcal{A}$, ossia

$$\mathcal{A}^* := \bigcup_{n \in \mathbb{N}_0} \mathcal{A}^n \tag{2.5}$$

dove $\mathcal{A}^0 := \{\lambda\}$, essendo λ la parola vuota.

Sull'alfabeto $\mathcal{A}$ si possono considerare anche le stringhe infinite, che indicheremo con ω. Poniamo

$$\Omega := \mathcal{A}^{\mathbb{N}_0} = \{\omega = (\omega_i)_{i \in \mathbb{N}_0} \ : \ \omega_i \in \mathcal{A} \ \forall \ i \in \mathbb{N}_0\} \tag{2.6}$$

Si può definire anche l'insieme delle stringhe bi-infinite $\mathcal{A}^{\mathbb{Z}}$, in cui gli indici dei simboli di una stringa variano in $\mathbb{Z}$ anziché in $\mathbb{N}_0$. Nel seguito ci restringiamo però al caso di $\mathbb{N}_0$, essendo le varie estensioni facili esercizi.

L'insieme Ω può essere munito della seguente metrica

$$d(\omega, \bar{\omega}) := \sum_{i=0}^{\infty} \frac{\delta(\omega_i, \bar{\omega}_i)}{2^i}$$

dove δ è definita sulle lettere di $\mathcal{A}$ tramite $\delta(a, b) = 0$ se $a = b$, e $\delta(a, b) = 1$ se $a \neq b$, per cui due stringhe sono "vicine" se hanno "abbastanza" simboli iniziali uguali. Lo spazio Ω risulta essere uno spazio metrico compatto, in cui la topologia indotta dalla metrica d ha come pre-base gli insiemi della forma

$$C(\omega, k, n) = \{\bar{\omega} \in \Omega \ : \ \bar{\omega}_i = \omega_i \text{ per ogni } i = k, \ldots, k + n - 1\}$$

per una generica stringa ω. Gli insiemi di questa forma si chiamano *cilindri*. Su Ω consideriamo sempre la σ-algebra di Borel generata dalla topologia dei *cilindri*. Nel seguito, indichiamo con ω_m^n la sottostringa finita di ω data dagli $n-m$ simboli $(\omega_m, \omega_{m+1}, \ldots, \omega_{n-1})$. Poniamo in particolare $\omega^n := \omega_0^n$.

Osserviamo che l'insieme Ω si può considerare l'insieme delle stringhe infinite prodotte da una sorgente di informazione che utilizzi l'alfabeto $\mathcal{A}$. Per estendere questa similarità consideriamo su Ω l'azione di una funzione τ che faccia "scorrere" la stringa, simulando in questo modo la successiva produzione dei simboli della stringa. Questo definisce lo *shift* $\tau : \Omega \to \Omega$, per cui la stringa $\tau(\omega)$ ha simboli dati da

$$\big(\tau(\omega)\big)_i = \omega_{i+i} \qquad \forall\ i \geq 0 \tag{2.7}$$

Proposizione 2.1. *La mappa di shift τ verifica le seguenti proprietà:*

(i) è continua rispetto alla topologia dei cilindri;

(ii) ha un'infinità numerabile di orbite periodiche;

(iii) ha un'infinità più che numerabile di orbite non periodiche;

(iv) ha un'orbita densa;

(v) ha dipendenza sensibile dalle condizioni iniziali, ossia esiste $\eta > 0$ tale che per ogni stringa ω e per ogni ε, esistono una stringa $\bar{\omega}$ e un intero n tale che $d(\omega, \bar{\omega}) < \varepsilon$ e $d(\tau^n\omega, \tau^n\bar{\omega}) > \eta$.

Dimostrazione. (i) La continuità della mappa di shift segue facilmente dalla forma dei cilindri. Infatti essendo i cilindri una pre-base per la topologia, basta verificare che $\tau^{-1}(C(\omega, k, n))$ è un aperto per ogni $\omega \in \Omega$ e ogni k, n.

Basta quindi osservare che $\tau^{-1}(C(\omega,k,n)) = \cup_{i\in\mathcal{A}} C(i\omega,k+1,n)$, dove $i\omega$ indica la concatenazione della lettera i con ω, e quindi è un aperto.

(ii) Data una stringa $s \in \mathcal{A}^n$ con $n \geq 1$, è facile verificare che la stringa $\omega = (s\ s\ s\dots)$, ottenuta come una ripetizione di s, è una stringa periodica di periodo n. Le orbite periodiche sono quindi in bigezione con l'insieme $\mathcal{A}^*$.

(iii) Le stringhe periodiche sono un'infinità numerabile, quindi le stringhe non periodiche sono un'infinità più che numerabile, essendo $d(\Omega) = d(N_0^{\mathbb{N}})$. Le orbite non periodiche devono quindi essere un'infinità più che numerabile. Osserviamo infatti che ogni orbita comprende una quantità numerabile di stringhe di Ω.

(iv) La dimostrazione di questo punto procede per costruzione. Bisogna mostrare l'esistenza di una stringa $\bar\omega$ che verifica

$$\forall\ \omega\in\Omega \quad \forall\ \varepsilon>0 \quad \exists\ n\in\mathbb{N} \quad t.c. \quad d(\omega,\tau^n(\bar\omega)) < \varepsilon$$

Dalla forma della metrica d segue che $d(\omega,\tau^n(\bar\omega)) < \varepsilon$ se $\omega_i = (\tau^n(\bar\omega))_i = \bar\omega_{i+n}$ per ogni $i = 0,\dots,\lfloor -\log_2\varepsilon\rfloor + 1$. Ne segue che basta trovare una stringa $\bar\omega$ tale che per ogni stringa finita $s\in\mathcal{A}^k$ esista un n tale che i primi k simboli di $\tau^n(\bar\omega)$ coincidano con s.

Per ogni k, vale $d(\mathcal{A}^k) = N^k$, e supponiamo di aver scelto una relazione d'ordine in $\mathcal{A}^k$ in modo da numerare le stringhe finite. Indichiamo con $s_1^k,\dots,s_{N^k}^k$ le stringhe di $\mathcal{A}^k$. La stringa $\bar\omega$ che cerchiamo si ottiene allora come la concatenazione di tutte le stringhe di $\mathcal{A}^1$ in ordine, concatenate con tutte le stringhe di $\mathcal{A}^2$ scelte in ordine, e così via.

(v) Scegliamo $\eta = \frac{1}{2}$. Data una qualsiasi stringa ω e un qualsiasi $\varepsilon > 0$, prendiamo la stringa $\bar\omega$ fatta in modo che $\bar\omega_i = \omega_i$ per ogni $i = 0,\dots,k$ con $k = \lfloor -\log_2\varepsilon\rfloor + 1$, e $\bar\omega_{k+1} \neq \omega_{k+1}$. Allora $d(\omega,\bar\omega) < \varepsilon$ e $d(\tau^{k+1}(\omega),\tau^{k+1}(\bar\omega)) > \eta$. □

Le proprietà della Proposizione 2.1 fanno della mappa di shift un esempio di sistema dinamico "caotico".

Continuando il parallelismo tra dinamica simbolica e sorgenti di informazione, osserviamo che finora abbiamo considerato solo il caso di una sorgente che non abbia restrizioni nelle stringhe prodotte. Supponiamo invece che ci siano restrizioni, ma che tali restrizioni si possano descrivere tramite una matrice M di dimensione $N\times N$ detta di *transizione* e definita nel modo seguente. Indichiamo con $(m_{ij})_{ij}$ gli elementi della matrice M e supponiamo che $m_{ij}\in\{0,1\}$. Definiamo allora l'insieme Ω_M delle stringhe *ammissibili* tramite

$$\Omega_M := \big\{\omega\in\Omega\ :\ m_{\omega_i\ \omega_{i+1}} = 1 \text{ per ogni } i\geq 0\big\} \tag{2.8}$$

quindi gli elementi di M uguali a 1 dicono quali sono le coppie di simboli di $\mathcal{A}$ che possono essere prodotte.

Si dimostra che Ω_M è chiuso rispetto all'azione della mappa di shift τ, che con la metrica indotta Ω_M è metrico compatto, e la σ-algebra di Borel corrisponde alla restrizione della σ-algebra dei cilindri. Se indichiamo con τ_M la restrizione di τ a Ω_M, abbiamo quindi che (Ω_M, τ_M) è un sistema dinamico nel senso della nostra Definizione 2.1. Tale sistema dinamico si chiama comunemente *subshift di tipo finito* o *topological Markov chain*.

Le proprietà dinamiche di un subshift di tipo finito dipendono dalle proprietà della matrice di transizione, che stabilisce le dipendenze tra i simboli di $\mathcal{A}$.

Definizione 2.5. *Una matrice di transizione M si dice* irriducibile *se per ogni coppia di simboli (i, j) esiste un intero k tale che $m_{ij}^k \neq 0$, dove $(m_{ij}^k)_{ij}$ sono gli elementi della matrice M^k, la k-esima potenza di M.*

L'irriducibilità di una matrice di transizione, assicura che esiste almeno una stringa finita s che collega i e j per ogni coppia di simboli di $\mathcal{A}$. Quindi entrambi i simboli possono essere presenti contemporaneamente in una stringa infinita (e in un ordine prestabilito).

Definizione 2.6. *Una matrice di transizione M si dice* irriducibile e aperiodica *se esiste un intero k tale che $m_{ij}^k \neq 0$ per ogni coppia di simboli (i, j).*

Proposizione 2.2. *Un subshift di tipo finito con matrice di transizione irriducibile e aperiodica verifica le proprietà (i)-(v) della Proposizione 2.1.*

Dimostrazione. La dimostrazione procede analogamente a quella della Proposizione 2.1. Bisogna solo prestare attenzione al fatto che non tutte le stringhe finite sono ammissibili. In particolare, per la dimostrazione del punto (ii), bisogna dimostrare l'esistenza di infinite stringhe finite che iniziano e finiscono con lo stesso simbolo. La concatenazione di queste stringhe forma le stringhe periodiche. □

Esistono sistemi dinamici che descrivono limitazioni più generali nella produzione di simboli, come ad esempio i sistemi sofici. Tuttavia noi ci restringiamo nel seguito al caso di subshift di tipo finito. △

2.2 Coniugio e rappresentazione simbolica

Due sistemi dinamici diversi possono avere le stesse caratteristiche. Finora l'unica struttura che abbiamo considerato sull'insieme di partenza è la

presenza di una metrica, che implica anche la σ-algebra di Borel. Dunque possiamo parlare di sistemi dinamici che sono equivalenti dal punto di vista al più topologico.

Definizione 2.7. Siano (X_1, T_1) e (X_2, T_2) due sistemi dinamici. Se esiste $h : X_1 \to X_2$ surgettiva e misurabile tale che $h \circ T_1 = T_2 \circ h$ si dice che T_2 è un *fattore* di T_1 e che h è un *semi-coniugio*. Se h è anche iniettiva con inversa misurabile allora h si chiama *coniugio* e T_1 e T_2 si dicono *coniugati*.

Se inoltre T_1 e T_2 sono continue, allora si richiede che anche h (e h^{-1}) sia continua, e si parla di *semi-coniugio (coniugio) topologico*.

Nel caso di un semi-coniugio si dice che T_2 è un fattore di T_1 perché in generale la dinamica di T_2 è meno "complicata". Infatti, mancando l'iniettività, h potrebbe mandare un'intero insieme invariante di X_1 in un punto di X_2, che quindi perderebbe la ricchezza della dinamica dell'insieme invariante. Si può addirittura pensare il caso estremo in cui X_2 sia un unico punto e h manda tutto X_1 in un punto.

Se invece i due sistemi dinamici sono coniugati, allora c'è completa equivalenza della dinamica (almeno a livello topologico). Per esempio punti fissi e orbite periodiche corrispondono nei due sistemi, così come gli insiemi invarianti.

Un coniugio topologico può essere visto come una rappresentazione topologica di un sistema dinamico tramite un'altro sistema di più facile interpretazione. È questo per esempio il caso dell'esercizio 2.2, in cui il sistema T_1, essendo lineare, è di più facile studio del sistema T_2.

Un caso particolare di rappresentazione di un sistema è la *rappresentazione simbolica*. Sia (X, T) un sistema dinamico e sia $Z = \{I_1, \dots, I_N\}$ una partizione finita e misurabile dello spazio X. Ossia gli insiemi $I_j \subset X$ sono misurabili e tali che $\cup_{j=1}^{N} I_j = X$ e $I_i \cap I_j = \emptyset$ per $i \neq j$. Alla partizione Z viene associato l'alfabeto $\mathcal{A} = \{1, \dots, N\}$ e sia $\Omega = \mathcal{A}^{\mathbb{N}_0}$.

Si definisce allora un'applicazione $\varphi_Z : X \to \Omega$ tramite

$$\varphi_Z(x) = \omega \quad \Longleftrightarrow \quad T^j(x) \in I_{\omega_j} \ \ \forall\, j \in \mathbb{N}_0 \tag{2.9}$$

La rappresentazione simbolica crea un collegamento tra il sistema dinamico (X, T) e la dinamica simbolica (Ω, τ). Infatti si dimostra facilmente che l'immagine di X tramite φ_Z è invariante per l'azione dello shift τ e che $\tau \circ \varphi_Z = \varphi_Z \circ T$. L'immagine $\varphi_Z(X)$ contiene le stringhe ω tali che $\cap_{i=0}^{\infty} T^{-i} I_{\omega_i} \neq \emptyset$ e in generale non è un subshift di tipo finito. Le proprietà topologiche di $\varphi_Z(X)$ in Ω dipendono fortemente dalla dinamica, per esempio non è detto che l'immagine sia un insieme chiuso.

L'applicazione φ_Z risulta misurabile per definizione, basta verificarlo sui cilindri di Ω, che generano la σ-algebra di Borel della dinamica simbolica. Infatti se $C(\omega, k, n)$ è un cilindro di Ω si ha

$$(\varphi_Z)^{-1}(C(\omega, k, n)) = \bigcap_{i=k}^{k+n-1} T^{-i} I_{\omega_i}$$

che è intersezione di insiemi misurabili.

Nel caso di mappe continue diventa importante la continuità di φ_Z. Si verifica che tranne per un insieme trascurabile di punti la rappresentazione simbolica è continua (vedi esempio 2.6).

Otteniamo quindi che la rappresentazione simbolica è un coniugio o un semi-coniugio (topologico) del sistema dinamico, e l'iniettività dell'applicazione φ_Z dipende dalla scelta della partizione. Le partizioni buone sono chiamate "generanti" per il sistema dinamico (vedi Definizione 2.12).

Esempio 2.6. Sia $X = [0,1]/(0 \sim 1)$ con la partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$ e l'azione della mappa T del panettiere (vedi l'equazione (2.1)). L'alfabeto associato alla partizione Z è l'alfabeto binario $\mathcal{A} = \{0, 1\}$ e la rappresentazione simbolica ha una facile interpretazione. Sia infatti

$$x = \sum_{i=1}^{\infty} \frac{x_i}{2^i}$$

l'espansione binaria di un numero reale $x \in [0, 1]$, dove $x_i \in \{0, 1\}$ per ogni i. È facile dimostrare che ad ogni punto x viene associata una singola stringa $\omega \in \mathcal{A}^{\mathbb{N}_0}$, tranne che per un insieme numerabile di punti, chiamati *punti diadici*. Questi punti sono quelli che corrispondono alle coppie di stringhe equivalenti (dal punto di vista dell'espansione binaria) ω^1 e ω^2 tali che: esiste $\bar{n}$ per cui $\omega^1_i = \omega^2_i$ per ogni $i = 0, \ldots, \bar{n} - 1$; $\omega^1_{\bar{n}} = 1$ e $\omega^1_i = 0$ per ogni $i > \bar{n}$; $\omega^2_{\bar{n}} = 0$ e $\omega^2_i = 1$ per ogni $i > \bar{n}$. Altrettanto semplicemente si verifica che la rappresentazione simbolica relativa a Z associa a un punto x la stringa data dalla sua espansione binaria, quindi $T^i(x) \in I_j$ se e solo se $x_{i+1} = j$ per ogni $i \in \mathbb{N}_0$, e nel caso dei punti diadici viene scelta la stringa della forma ω^1.

Lo studio di questa rappresentazione simbolica diventa quindi molto più semplice usando l'espansione binaria dei punti x. Si ottiene facilmente che $\varphi_Z(X) = \Omega$ e l'iniettività di φ_Z. Inoltre φ_Z è continua tranne che nei punti diadici, che sono un insieme numerabile (e quindi trascurabile rispetto alla misura di Lebesgue).

Si ottiene che la rappresentazione simbolica relativa alla partizione Z data è un coniugio topologico (a meno di un insieme di misura nulla e dalla

dinamica trascurabile) tra la mappa del panettiere e la mappa di shift su $\{0,1\}^{\mathbb{N}_0}$. $\triangle$

La rappresentazione simbolica di un sistema dinamico riveste una grande importanza per i nostri scopi, in quanto costituisce il mezzo di passaggio da un sistema dinamico a una sorgente di informazione.

2.3 Entropia topologica

Abbiamo visto che una classificazione delle sorgenti di informazione è possibile attraverso lo studio delle loro proprietà statistiche, come la capacità e l'entropia di Shannon (vedi Definizioni 1.1 e 1.2). Iniziamo adesso uno studio dei metodi di classificazione dei sistemi dinamici, studiando prima le proprietà topologiche.

Nel passaggio da sistemi dinamici a sorgenti di informazione, abbiamo visto che uno strumento fondamentale sono le partizioni dello spazio X su cui si svolge la dinamica, partizioni che ci permettono una "discretizzazione" della dinamica.

Sia (X,T) un sistema dinamico e sia $Z = \{I_1, \ldots, I_N\}$ una partizione finita e misurabile di X. Consideriamo solo il caso di partizioni con insiemi a parte interna non vuota. Indichiamo con $T^{-1}Z$ la partizione di X data dalle contro-immagini degli insiemi di Z, quindi $T^{-1}Z = \{T^{-1}I_1, \ldots, T^{-1}I_N\}$, dove ricordiamo che $T^{-1}I_j = \{x \in X \ : \ T(x) \in I_j\}$. Così per ricorrenza poniamo $T^{-n}Z := T^{-1}(T^{-(n-1)}Z)$ per ogni $n \geq 1$, dove $T^0 Z = Z$.

Definizione 2.8. Date due partizioni P e Q, si definisce *unione* di P e Q, e si indica con $P \vee Q$, la partizione data da tutte le possibili intersezioni degli insiemi di P e Q, ossia

$$P \vee Q := \{P_i \cap Q_j \ : \ P_i \in P,\ Q_j \in Q\}$$

La partizione P è contenuta in Q ($P \subset Q$) se per ogni $P_i \in P$ esiste un insieme $Q_j \in Q$ tale che $P_i \subset Q_j$.

Data la partizione Z, definiamo la *partizione iterata* Z_n come

$$Z_n := \bigvee_{i=0}^{n-1} T^{-i} Z \tag{2.10}$$

Gli insiemi della partizione Z_n sono della forma

$$I_{i_0} \cap T^{-1} I_{i_1} \cap \cdots \cap T^{-(n-1)} I_{i_{n-1}}$$

al variare degli indici i_j nell'insieme $\mathcal{A} = \{1, \dots, N\}$. I punti x che appartengono a un insieme della partizione iterata sono caratterizzati dal fatto che la loro stringa simbolica è tale che $(\varphi_Z(x))^n = (i_0\ i_1 \dots i_{n-1})$. Dato un punto $x \in X$ indichiamo con $Z_n(x)$ l'insieme della partizione iterata che lo contiene, quindi $x \in Z_n(x)$ per ogni $x \in X$.

Se consideriamo la sorgente di informazione che si ottiene dal sistema dinamico (X, T) tramite la rappresentazione simbolica φ_Z, la capacità di tale sorgente è legata alla cardinalità delle partizioni iterate Z_n. Usando il concetto di capacità, possiamo quindi definire l'analogo concetto per il sistema dinamico visto attraverso la partizione Z.

Definizione 2.9. Data una partizione finita Z di X, indichiamo con $d(Z)$ la cardinalità di Z, ossia il numero di insiemi di Z. Si definisce *entropia topologica di Z* il numero

$$H_{top}(Z) := \log(d(Z))$$

Data una dinamica T su X, si definisce *entropia topologica del sistema dinamico (X, T) relativa a una partizione Z* il rapporto di crescita dell'entropia topologica delle partizioni iterate,

$$h_{top}(T, Z) := \lim_{n\to\infty} \frac{H_{top}(Z_n)}{n}$$

Proposizione 2.3. *Siano P, Q e Z partizioni finite. Allora*

(i) $H_{top}(P) \geq 0$;

(ii) Se $P \subset Q$ *allora* $H_{top}(P) \geq H_{top}(Q)$;

(iii) $H_{top}(P \vee Q) \leq H_{top}(P) + H_{top}(Q)$;

(iv) $H_{top}(Z_{n+m}) \leq H_{top}(Z_n) + H_{top}(Z_m)$ *per ogni* $n, m \in \mathbb{N}$.

Dimostrazione. (i) Deriva dalla definizione.

(ii) Basta notare che se $P \subset Q$ allora necessariamente P non può avere meno insiemi di Q, e quindi la tesi segue dalla monotonia della funzione logaritmo.

(iii) Bisogna mostrare che $d(P \vee Q) \leq d(P)d(Q)$. Questo segue immediatamente dalla definizione di unione di due partizioni.

(iv) Notiamo che $Z_{n+m} = Z_n \vee T^{-n} Z_m$ e che $d(T^{-n} Z_m) \leq d(Z_m)$, quindi usando il punto (iii), $d(Z_{n+m}) \leq d(Z_n)d(Z_m)$. □

Il punto (iv) e il Teorema 1.1 assicurano la buona definizione dell'entropia topologica $h_{top}(T, Z)$.

Per la classificazione dei sistemi dinamici in base alle loro proprietà statistiche sarebbe tuttavia importante avere dei concetti indipendenti dalla scelta di una particolare partizione. Tuttavia ci sono delle difficoltà a lavorare con le partizioni, che non sono certo lo strumento più adatto per studiare le proprietà topologiche di un sistema dinamico. Per introdurre quindi un concetto che sia una proprietà intrinseca del sistema dinamico, costruiamo l'analogo dell'entropia topologica di una partizione per ricoprimenti aperti e finiti dello spazio X.

Definizione 2.10. Dato uno spazio metrico compatto X, si definisce *ricoprimento aperto di* X una collezione α di insiemi aperti $\{A_1, \ldots, A_k\}$, tali che $X = \cup_{i=1}^{k} A_i$.

A differenza delle partizioni notiamo che i ricoprimenti aperti di X sono costituiti da insiemi che si possono intersecare.

Le definizioni di entropia topologica $H_{top}(\alpha)$ di un ricoprimento aperto, i concetti di unione di due ricoprimenti, di iterazione di un ricoprimento e di contenimento tra due ricoprimenti, e la definizione di entropia topologica $h_{top}(T, \alpha)$ di un sistema dinamico relativa a un ricoprimento, sono perfettamente analoghe al caso di una partizione. Così valgono allo stesso modo le relazioni della Proposizione 2.3.

Usando allora i ricoprimenti aperti si può definire

Definizione 2.11. L'*entropia topologica di un sistema dinamico* (X, T) è

$$h_{top}(T) := \sup \{h_{top}(T, \alpha) \ : \ \alpha \text{ ricoprimento aperto finito}\}$$

La definizione di entropia topologica non è certo adatta al calcolo nei casi concreti. A questo scopo si dimostra che basta scegliere un ricoprimento "buono" e calcolare l'entropia topologica relativa.

Definizione 2.12. Un ricoprimento aperto e finito α è *generante* per il sistema dinamico (X, T) se

$$\lim_{n \to \infty} diam(\alpha_n) = 0$$

dove $diam(\alpha_n) = \sup \{diam(A) \ : \ A \in \alpha_n\}$[2].

Teorema 2.4. *Sia α ricoprimento aperto generante per il sistema dinamico (X, T), allora $h_{top}(T) = h_{top}(T, \alpha)$.*

[2]Si definisca $diam(A) = \sup \{d(x, y) \ : \ x, y \in A\}$

Dimostrazione. Innanzitutto dalla definizione dell'entropia topologica, segue che $h_{top}(T, \alpha) \leq h_{top}(T)$.

Dimostriamo ora la diseguaglianza opposta. Fissato $\varepsilon > 0$, troviamo, per le proprietà dell'estremo superiore di un insieme, un ricoprimento aperto finito α' tale che $h_{top}(T, \alpha') + \varepsilon > h_{top}(T)$. Ma essendo α ricoprimento generante, esiste $N > 0$ tale che $\alpha_N \subset \alpha'$, e quindi

$$h_{top}(T, \alpha_N) = \lim_{n\to\infty} \frac{H_{top}((\alpha_N)_n)}{n} \geq \lim_{n\to\infty} \frac{H_{top}(\alpha'_n)}{n} = h_{top}(T, \alpha')$$

D'altra parte dalla relazione $H_{top}((\alpha_N)_n) = H_{top}(\alpha_{N+n})$ segue

$$h_{top}(T, \alpha_N) = \lim_{n\to\infty} \frac{H_{top}(\alpha_{N+n})}{n} = \lim_{n\to\infty} \frac{N+n}{n} \frac{H_{top}(\alpha_{N+n})}{N+n} = h_{top}(T, \alpha)$$

Quindi per ogni $\varepsilon > 0$ vale $h_{top}(T, \alpha) > h_{top}(T) - \varepsilon$, e la tesi segue dall'arbitrarietà di ε. □

Proposizione 2.5. *Dati due sistemi dinamici (X_1, T_1) e (X_2, T_2), se T_2 è un fattore topologico di T_1, allora $h_{top}(T_2) \leq h_{top}(T_1)$. Se T_1 e T_2 sono coniugati topologicamente allora vale l'uguaglianza.*

Per i nostri scopi sarà sufficiente considerare l'entropia topologica di un sistema dinamico relativamente a una partizione finita e misurabile, quindi ci restringiamo ad analizzare i sistemi dinamici di base introdotti nel Paragrafo 2.1 usando particolari partizioni. Calcolare l'entropia topologica di un sistema dinamico in generale richiede definizioni equivalenti, su cui non ci soffermiamo. Vedremo comunque che nei casi che affrontiamo, le partizioni che scegliamo sono abbastanza "buone" da far ottenere l'entropia topologica.

Esempio 2.7 (Mappa del panettiere, Esempio 2.1). Consideriamo, sulla base dell'Esempio 2.6, la partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$. È immediato verificare che $d(Z_n) = 2^n$, quindi

$$h_{top}(T, Z) = \lim_{n\to\infty} \frac{\log 2^n}{n} = 1$$

La sorgente di informazione associata ha quindi capacità $C = 1$. Osserviamo che $h_{top}(T) = h_{top}(T, Z)$. △

Esempio 2.8 (Rotazioni irrazionali, Esempio 2.2). Consideriamo la rotazione $T_\alpha(x) = x + \alpha(mod 1)$ su $X = [0, 1]/(0 \sim 1)$. Supponiamo che $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Proposizione 2.6. *Data la partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$ vale $d(Z_n) = 2n$.*

Dimostrazione. Basta osservare che un insieme di Z_{n-1} interseca due elementi di $T_\alpha^{-(n-1)}Z$ se e solo uno dei punti $\frac{1}{2}-(n-1)\alpha$ e $1-(n-1)\alpha$ appartiene all'insieme in questione (modulo 1). Ne risulta che solo esattamente due insiemi di Z_{n-1} saranno divisi in due nella formazione di Z_n. □

Applicando la proposizione si ottiene $h_{top}(T_\alpha, Z) = 0$. Osserviamo che nuovamente $h_{top}(T) = 0$. △

Esempio 2.9 (Famiglia quadratica, Esempio 2.3). Studiare il comportamento di $h_{top}(T_\lambda, Z)$ con $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1]\right\}$ al variare di $\lambda \in [0, 4]$ comporta notevoli difficoltà, e usa tecniche che sono al di fuori dello scopo di questo corso. A titolo di esempio possiamo sfruttare l'esercizio 2.2, per ottenere $h_{top}(T_4, Z) = 1 = h_{top}(T_4)$. Il calcolo dell'entropia topologica della mappa tenda è analogo al caso della mappa del panettiere. △

Esempio 2.10 (Famiglia di Pomeau-Manneville, Esempio 2.4). Usando l'esercizio 2.5, si ottiene $h_{top}(T_z, Z) = 1 = h_{top}(T_z)$ per ogni $z > 1$, con $Z = \{[0, x_z), [x_z, 1)\}$, dove x_z è il numero reale in $(0, 1)$ soluzione di $x_z + x_z^z = 1$. △

Esempio 2.11 (Dinamica simbolica, Esempio 2.5). Sia $\mathcal{A} = \{1, \ldots, N\}$ l'alfabeto e M la matrice di transizione. Nel caso in cui $m_{ij} = 1$ per ogni coppia di simboli (i, j) (ossia consideriamo la mappa shift), si verifica facilmente che

$$h_{top}(\tau) = \log N$$

Infatti in questo caso i cilindri $C(j, 0, 1)$ al variare di j in $\mathcal{A}$ costituiscono una partizione Z finita e misurabile, ma anche un ricoprimento aperto generante, e si verifica che $d(Z_n) = N^n$. Basta infatti contare tutte le stringhe possibili di lunghezza n. Un po' più complesso risulta il conto nel caso di una matrice di transizione con alcuni zeri (ossia per un subshift di tipo finito). In questo caso bisogna usare

Teorema 2.7 (Frobenius-Perron). *Sia M matrice di transizione irriducibile e aperiodica. Allora esiste $\lambda_1 > 0$, tale che λ_1 è autovalore di M con autospazio di dimensione 1, e tutti gli altri autovalori verificano $|\lambda_i| < \lambda_1$.*

Proposizione 2.8. *Sia M matrice di transizione irriducibile e aperiodica, allora se λ_1 è l'autovalore massimale di M dato dal Teorema di Frobenius-Perron, si ha*

$$h_{top}(\tau_M) = \log \lambda_1$$

Dimostrazione. Scegliamo il ricoprimento aperto generante α dato dai cilindri $C(j, 0, 1)$ al variare di j in $\mathcal{A}$. Si verifica facilmente che $d(\alpha_n)$ è uguale al

numero di stringhe ammissibili di lunghezza n e quindi per ogni $n \geq 2$ vale

$$H_{top}(\alpha_n) = \log \left[\sum_{i,j=1}^{N} m_{i,j}^{n-1} \right]$$

dove $m_{i,j}^{n-1}$ sono gli elementi della matrice M^{n-1}.

Usando il Teorema di Frobenius-Perron si ottiene che esiste una matrice invertibile U e una matrice D diagonale a blocchi, i blocchi di Jordan, $D = diag(B_1, \ldots, B_k)$, tale che $M = UDU^{-1}$ e $B_1 = (\lambda_1)$ è il blocco relativo all'autovalore massimale, che ha autospazio di dimensione 1. Si dimostra allora, usando $M^n = UD^nU^{-1}$, che

$$\sum_{i,j=1}^{N} m_{i,j}^{n} = C\lambda_1^n + E(n)$$

dove C è una costante che dipende solo dalla matrice U e $E(n) = o(\lambda_1^n)$. Quindi

$$h_{top}(\tau_M) = \lim_{n\to\infty} \frac{\log C + (n-1)\log\lambda_1 + \log\left(1 + \frac{E(n-1)}{C\lambda_1^{n-1}}\right)}{n}$$

da cui la tesi. □

Notiamo che il calcolo dell'entropia topologica per subshift di tipo finito, può essere utile per avere stime dell'entropia topologica per sistemi semiconiugati topologicamente con le loro rappresentazioni simboliche, nel caso queste siano subshift di tipo finito. △

2.4 Misure invarianti ed entropia metrica

Per continuare la classificazione dei sistemi dinamici è necessario considerare le proprietà statistiche dei sistemi. Per questo motivo, come abbiamo già visto per le sorgenti di informazione, è necessario considerare misure definite sullo spazio delle fasi del sistema dinamico.

Sia $(X, \mathcal{B}, T)$ un sistema dinamico definito su uno spazio metrico compatto X, con σ-algebra di Borel $\mathcal{B}$. Sia μ una misura su $(X, \mathcal{B})$ di probabilità o che renda X spazio σ-finito. Supponiamo che μ sia *non-singolare* per la mappa T, ossia $\mu(T^{-1}(B)) = 0$ se e solo se $\mu(B) = 0$, per ogni insieme misurabile $B \subset X$.

Una proprietà importante che si richiede a una misura μ è l'invarianza rispetto all'azione del sistema dinamico. Ossia chiediamo che l'utilizzo della misura μ per il calcolo delle proprietà statistiche del sistema, prescinda dal numero di iterazioni già compiute dalla mappa T. Questo concetto corrisponde all'ipotesi di stazionarietà per le sorgenti di informazione (vedi fine Paragrafo 1.1).

Definizione 2.13. Una misura μ su $(X, \mathcal{B})$ è *invariante rispetto all'azione di una mappa misurabile T (T-invariante)* se vale

$$\mu(T^{-1}(B)) = \mu(B) \qquad \forall\ B \in \mathcal{B}$$

oppure, il che è lo stesso,

$$\int_X f(T(x))\ d\mu(x) = \int_X f(x)\ d\mu(x) \qquad \forall\ f \in L^1(X, \mu)$$

Teorema 2.9 (Krylov-Bogolubov)**.** *Sia $(X, \mathcal{B}, T)$ sistema dinamico con T mappa continua. Allora esiste almeno una misura μ di probabilità T-invariante.*

Dimostrazione. Sia μ_0 una misura di probabilità su X. Definiamo una misura di probabilità $\mu_1 := T^*\mu_0$ data da $\mu_1(B) = \mu_0(T^{-1}(B))$ per ogni $B \in \mathcal{B}$. Possiamo allora definire per ricorsione una successione di misure di probabilità $(\mu_n)_n$ a partire da μ_0. Considerando la topologia debole sullo spazio delle misure di probabilità, la funzione T^* è continua grazie alla continuità di T. Consideriamo poi la successione delle medie, ossia

$$\nu_n := \frac{1}{n+1} \sum_{k=0}^{n} \mu_k$$

Per la compattezza debole dello spazio delle misure di probabilità, esiste una sotto-successione $(\nu_{n_j})_j$ convergente debole. Sia ν il limite di una tale sotto-successione. Allora

$$T^*\nu = \lim_{j\to\infty} T^*\nu_{n_j} = \lim_{j\to\infty} \left(\frac{1}{n_j+1} \sum_{k=0}^{n_j} \mu_k + \frac{\mu_{n_j+1} - \mu_0}{n_j+1} \right) = \lim_{j\to\infty} \nu_{n_j} = \nu$$

dove la prima uguaglianza segue dalla continuità di T^*. Ne segue che ν è misura di probabilità invariante. □

Nel caso semplice di un'orbita periodica, la dimostrazione del Teorema di Krylov-Bogolubov fornisce l'idea per ottenere una misura di probabilità

T-invariante legata all'orbita periodica. Se infatti $x \in X$ è periodico di periodo p, allora, ponendo $\mu_0 = \delta_x$, la funzione generalizzata "delta di Dirac" centrata in x, la misura di probabilità limite ν invariante è

$$\mu_x = \frac{1}{p} \sum_{i=0}^{p-1} \delta_{T^i(x)} \tag{2.11}$$

In particolare nel caso di punti fissi, si ottengono le misure centrate nel punto. Notiamo che in questo caso basta supporre T misurabile.

La mappa del panettiere (Esempio 2.1) e le rotazioni del cerchio (Esempio 2.2) preservano la misura di Lebesgue m normalizzata su $[0, 1]$ (vedi esercizio 2.6).

Esempio 2.12 (Famiglia quadratica). Studiamo il caso dell'azione della mappa $T(x) = 4x(1-x)$ su $[0, 1]$. Cerchiamo in particolare una misura invariante μ che sia assolutamente continua rispetto alla misura di Lebesgue. Ossia cerchiamo una μ della forma $d\mu(x) = f(x)\, dx$, per una funzione $f(x)$ misurabile e non-negativa. Se la funzione $f(x)$ è anche in $L^1([0, 1], m)$ allora la misura μ è di probabilità (basta rinormalizzare usando $\|f\|_{L^1}$).

Per trovare la funzione f, usiamo la Definizione 2.13. Dato $x \in [0, 1]$, considerando l'insieme $B = [0, x]$, si ottiene che la f deve verificare

$$\int_B f(t)\, dt = \int_{T^{-1}(B)} f(t)\, dt$$

Dall'equazione di T si ricava che

$$T^{-1}([0, x]) = [0, y_1] \cup [y_2, 1]$$

dove $y_1 < y_2$ sono le due soluzioni di $4y(1 - y) = x$. Quindi

$$\int_0^x f(t)\, dt = \int_0^{y_1} f(t)\, dt + \int_{y_2}^1 f(t)\, dt$$

e derivando entrambi i membri rispetto a x si ottiene l'equazione funzionale

$$f(x) = \frac{1}{4\sqrt{1-x}} \left(f\left(\frac{1-\sqrt{1-x}}{2}\right) + f\left(\frac{1+\sqrt{1-x}}{2}\right)\right)$$

Notiamo poi che la funzione $f(x)$ deve essere simmetrica rispetto al punto $x = \frac{1}{2}$, per cui si ottiene

$$f(x) = \frac{1}{2\sqrt{1-x}}\, f\left(\frac{1+\sqrt{1-x}}{2}\right)$$

che ha come soluzione le funzioni della forma

$$f(x) = \frac{c}{4\sqrt{x(1-x)}} \tag{2.12}$$

con c costante positiva. La misura μ invariante e di probabilità si ottiene per $c = \frac{4}{\pi}$. $\triangle$

Esempio 2.13 (Famiglia di Pomeau-Manneville). La trattazione in questo caso è più complicata, per cui ci limitiamo a enunciare i risultati.

Proposizione 2.10. *Per ogni $z > 1$ esistono infinite misure invarianti di probabilità concentrate sulle orbite periodiche e sul punto fisso $x = 0$. Per ogni $z > 1$ esiste una misura T_z-invariante assolutamente continua rispetto a Lebesgue, con funzione misurabile $f_z(x)$ tale che $f_z \in L^1([0,1], m)$ se e solo se $z < 2$.* $\triangle$

Esempio 2.14 (Dinamica simbolica). Sia $\mathcal{A} = \{1, \ldots, N\}$ l'alfabeto su cui si costruisce uno shift simbolico. Sia $\{p(1), \ldots, p(N)\}$ una distribuzione di probabilità su $\mathcal{A}$. Definiamo una misura di probabilità μ su $\Omega := \mathcal{A}^{\mathbb{N}_0}$ che sia τ-invariante, ponendo

$$\mu(C(\omega, k, n)) := \prod_{i=0}^{n-1} p(\omega_{k+i}) \tag{2.13}$$

per ogni cilindro. La misura invariante su un insieme misurabile qualsiasi si definisce usando la proprietà dei cilindri di generare la σ-algebra di Borel di Ω. Si verifica facilmente che μ è una misura di probabilità e per la verifica che μ è una misura τ-invariante basta osservare che

$$\tau^{-1}(C(\omega, k, n)) = \{\bar{\omega} \ : \ \bar{\omega}_{k+i+1} = \omega_{k+i} \ \forall \ i = 0, \ldots, n-1\}$$

Questo implica che la misura μ è τ-invariante sui cilindri e si ragiona come prima per estendere la proprietà a tutti gli insiemi misurabili.

Osserviamo che la misura μ è la misura indotta su Ω da un processo stazionario di variabili aleatorie a valori in $\mathcal{A}$, indipendenti e distribuite tutte secondo la distribuzione $\{p(1), \ldots, p(N)\}$.

Più delicata è la situazione per un subshift di tipo finito. Sia M la matrice di transizione $N \times N$ irriducibile e aperiodica e sia Ω_M il sottospazio delle stringhe ammissibili. Associamo a M una *matrice stocastica* Π definita come segue. La matrice Π è una matrice $N \times N$ tale che $\pi_{ij} \geq 0$ per ogni coppia (i, j), che verifica $\pi_{ij} > 0$ se e solo se $m_{ij} = 1$ e $\sum_{j=1}^{N} \pi_{ij} = 1$ per ogni $i = 1, \ldots, N$. È chiaro che ad una matrice di transizione M si possono associare diverse matrici stocastiche. Enunciamo adesso senza dimostrazione un importante risultato per subshift di tipo finito.

Proposizione 2.11. *Se M è matrice di transizione irriducibile e aperiodica, allora per ogni matrice stocastica Π associata a M esiste un unico vettore $p = (p_1, \ldots, p_N)$, detto distribuzione stazionaria, tale che: $p_i \geq 0$ per ogni $i = 1, \ldots, N$; $\sum_{i=1}^{N} p_i = 1$; $p\Pi = p$.*

Data una matrice stocastica Π associata a M, sia p la sua distribuzione stazionaria su $\mathcal{A}$. Allora definiamo *misura di Markov* μ_Π la misura di probabilità definita sui cilindri tramite

$$\mu_\Pi(C(\omega, k, n)) = \left(\prod_{i=0}^{n-2} \pi_{\omega_{k+i}\omega_{k+i+1}} \right) p_{\omega_k} \tag{2.14}$$

e poi estesa a tutta la σ-algebra dei cilindri.

Verifichiamo che le misure di Markov sono τ_M-invarianti. Dalle definizioni si ottiene per ogni cilindro

$$\mu_\Pi(\tau_M^{-1}(C(\omega, k, n))) = \mu_\Pi \left\{ \bar{\omega} \in \Omega_M \ : \ \tau_M(\bar{\omega}) \in C(\omega, k, n) \right\} =$$

$$= \sum_{j=1}^{N} \pi_{j\omega_k} \left(\prod_{i=0}^{n-2} \pi_{\omega_{k+i}\omega_{k+i+1}} \right) p_j = \left(\prod_{i=0}^{n-2} \pi_{\omega_{k+i}\omega_{k+i+1}} \right) \sum_{j=1}^{N} \pi_{j\omega_k} p_j =$$

$$= \left(\prod_{i=0}^{n-2} \pi_{\omega_{k+i}\omega_{k+i+1}} \right) p_{\omega_k} = \mu_\Pi(C(\omega, k, n))$$

Osserviamo che potendo considerare lo shift τ come un subshift di tipo finito con matrice di transizione M tale che $m_{ij} = 1$ per ogni coppia (i, j), possiamo costruire le misure di Markov anche per lo shift. Nel caso in cui Π abbia tutte le righe uguali allora la misura di Markov associata corrisponde alla misura definita nell'equazione (2.13). $\triangle$.

Nello studio dei sistemi dinamici, la scelta di una misura invariante corrisponde alla scelta di un possibile tipo di stazionarietà per il processo stocastico associato al sistema, ossia di un tipo di stazionarietà per le rappresentazioni simboliche, e quindi per le sorgenti di informazione associate. In pratica, una volta scelta una misura invariante μ, si riduce lo studio del sistema dinamico allo studio delle proprietà statistiche che la misura μ "vede", ossia che si sviluppano nel suo supporto. Notiamo che l'invarianza di μ assicura che tali proprietà non dipendono dall'istante temporale in cui iniziamo a studiare il sistema dinamico.

Procedendo nello studio dei sistemi dinamici parallelamente a quanto fatto per le sorgenti di informazione, introduciamo un concetto di *entropia* per i sistemi dinamici.

Sia (X, T, μ) sistema dinamico, con μ misura di probabilità T-invariante. Sia $Z = \{I_1, \ldots, I_N\}$ partizione finita e misurabile di X.

Definizione 2.14. Si definisce *entropia metrica della partizione* Z il valore

$$H_\mu(Z) := -\sum_{i=1}^{N} \mu(I_i) \log(\mu(I_i))$$

con la convenzione $0 \log(0) = 0$.

I valori $\mu(I_i)$ si possono interpretare come la probabilità che un punto di X scelto a caso, secondo la distribuzione di probabilità μ, appartenga a I_i. Ne segue che $(-\log(\mu(I_i)))$ si può interpretare come l'informazione contenuta nella frase "$x \in I_i$". L'entropia metrica di una partizione Z è quindi l'informazione media che si ottiene conoscendo a quali insiemi di Z appartengono i punti di X. In quest'ottica si definisce una funzione di informazione $Inf_Z : X \to \mathbb{R}$ relativa alla partizione Z tramite

$$Inf_Z(x) := -\log(\mu(Z(x))) \tag{2.15}$$

dove $Z(x)$ indica l'insieme di Z che contiene x. Si ottiene

$$H_\mu(Z) = \int_X Inf_Z(x)\ d\mu(x)$$

Supponiamo ora di avere due partizioni finite e misurabili P e Q di X. Supponiamo di conoscere a quali insiemi di Q appartengono i punti di X. Ci chiediamo quale sia l'informazione necessaria a specificare in più a quali insiemi di P appartengono i punti. Si definisce così il concetto di entropia metrica condizionata.

Definizione 2.15. Date due partizioni finite e misurabile P e Q di X, si definisce *entropia metrica di* P *condizionata a* Q l'espressione

$$H_\mu(P|Q) := -\sum_{Q_j \in Q} \mu(Q_j) \left(\sum_{P_i \in P} \mu(P_i|Q_j) \log(\mu(P_i|Q_j)) \right)$$

dove $\mu(P_i|Q_j) = \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)}$.

Proposizione 2.12. *Siano P, Q e Z partizioni finite e misurabili di X. Allora*

(i) $0 \leq H_\mu(P) \leq \log(d(P))$;

(ii) $0 \leq H_\mu(P|Q) \leq H_\mu(P)$. *Inoltre* $H_\mu(P|Q) = H_\mu(P)$ *se e solo se* P *e* Q *sono indipendenti* ($\mu(P_i \cap Q_j) = \mu(P_i)\mu(Q_j)$ *per ogni* $P_i \in P$ *e* $Q_j \in Q$), *e* $H_\mu(P|Q) = 0$ *se e solo se* $Q \subset P$;

(iii) *Se* $Q \subset Z$ *allora* $H_\mu(P|Q) \leq H_\mu(P|Z)$;

(iv) $H_\mu(P \vee Q|Z) = H_\mu(P|Z) + H_\mu(Q|P \vee Z)$. *In particolare* $H_\mu(P \vee Q) = H_\mu(P) + H_\mu(Q|P)$ *e* $H_\mu(P \vee Q) \leq H_\mu(P) + H_\mu(Q)$;

(v) $H_\mu(P|Q) \leq H_\mu(P|Z) + H_\mu(Z|Q)$.

Dimostrazione. (i) Dalla definizione risulta evidente che $H_\mu(P) \geq 0$, essendo $\mu(P_i) \leq 1$ per ogni i. Si ottiene inoltre facilmente $H_\mu(P) = 0$ se e solo se tutti gli insiemi di P hanno misura nulla, tranne uno di misura totale. La seconda parte si dimostra studiando i massimi della funzione $H(p_1, \ldots, p_N)$ vincolati sul simplesso $\sum_i p_i = 1$, con $p_i \geq 0$. Si ottiene facilmente che il massimo si ha nel punto $p_1 = \cdots = p_N = \frac{1}{N}$, da cui la tesi.

(ii) La diseguaglianza $H_\mu(P|Q) \geq 0$ segue di nuovo facilmente dal fatto che $\mu(P_i|Q_j) \leq 1$ per ogni coppia i, j. Inoltre $H_\mu(P|Q) = 0$ se e solo se $\sum_i \mu(P_i|Q_j) \log(\mu(P_i|Q_j)) = 0$ per ogni j, il che equivale per ogni j (vedi punto (i)) a $Q_j \subset P_i$ per qualche i.

Osserviamo poi che $f(x) = -x \log x$ è una funzione concava su $[0, 1]$, quindi dati $(p_1, \ldots, p_N)$ non negativi tali che $\sum_j p_j = 1$, si ottiene, per ogni N-upla di numeri $(x_1, \ldots, x_N)$ in $[0, 1]$,

$$-\sum_j p_j x_j \log x_j \leq -(\sum_j p_j x_j) \log(\sum_j p_j x_j)$$

e l'uguaglianza vale se e solo se $x_j = \sum_j p_j x_j$ per ogni j. Ponendo $p_j = \mu(Q_j)$ e $x_j = \mu(P_i|Q_j)$ otteniamo allora

$$H_\mu(P|Q) = \sum_i \left(-\sum_j \mu(Q_j)\mu(P_i|Q_j) \log(\mu(P_i|Q_j)) \right) \leq$$

$$\leq -\sum_i \left(\sum_j \mu(P_i \cap Q_j) \right) \log \left(\sum_j \mu(P_i \cap Q_j) \right) = H_\mu(P)$$

Inoltre l'uguaglianza vale se e solo se $\mu(P_i|Q_j) = \mu(P_i)$ per ogni i, j.

(iii) La diseguaglianza segue ancora dalla concavità della funzione $f(x) = -x \log x$. Essendo $Q \subset Z$, per ogni Q_j esiste Z_k tale che $Q_j \subset Z_k$, in

particolare per ogni j vale

$$\mu(Q_j) = \sum_k \mu(Q_j \cap Z_k)$$

Applichiamo questa relazione alla definizione di entropia condizionata, e scriviamo

$$H_\mu(P|Q) = -\sum_{i,k} \mu(Z_k) \sum_j \mu(Q_j|Z_k)\mu(P_i|Q_j) \log(\mu(P_i|Q_j))$$

Osservando che per ogni k vale $1 = \sum_j \mu(Q_j|Z_k)$, otteniamo

$$H_\mu(P|Q) \leq -\sum_{i,k} \mu(Z_k) \left(\sum_j \frac{\mu(P_i \cap Q_j)}{\mu(Z_k)} \right) \log \left(\sum_j \frac{\mu(P_i \cap Q_j)}{\mu(Z_k)} \right)$$

che eguaglia $H_\mu(P|Z)$, notando che $\sum_j \mu(P_i \cap Q_j) = \mu(P_i \cap Z_k)$ per un certo k.

(iv) La prima parte si ottiene con le seguenti uguaglianze

$$H_\mu(P \vee Q|Z) = -\sum_k \mu(Z_k) \sum_{i,j} \mu(Q_j|P_i \cap Z_k)\mu(P_i|Z_k) \log(\mu(Q_j|P_i \cap Z_k)) -$$

$$-\sum_k \mu(Z_k) \sum_{i,j} \mu(Q_i|P_i \cap Z_k)\mu(P_i|Z_k) \log(\mu(P_i|Z_k)) =$$

$$= -\sum_{i,k} \mu(P_i \cap Z_k) \sum_j \mu(Q_j|P_i \cap Z_k) \log(\mu(Q_j|P_i \cap Z_k)) -$$

$$-\sum_{i,k} \mu(P_i \cap Z_k) \log(\mu(P_i|Z_k))(\sum_j \mu(Q_j|P_i \cap Z_k)) = H_\mu(Q|P \vee Z) + H_\mu(P|Z)$$

In particolare se $Z = \{X, \emptyset\}$ si ottiene $H_\mu(P \vee Q|Z) = H_\mu(P \vee Q)$ e $H_\mu(Q|P \vee Z) = H_\mu(Q|P)$, da cui segue la seconda relazione. Per la terza, basta usare (ii).

(v) Scriviamo $Z \vee Q \subset Z$ e applichiamo (iii) e (iv), ottenendo

$$H_\mu(P|Z) \geq H_\mu(P|Z \vee Q) = H_\mu(P \vee Z|Q) - H_\mu(Z|Q) \geq H_\mu(P|Q) - H_\mu(Z|Q)$$

e quindi la tesi. □

Sia come sopra φ_Z la rappresentazione simbolica associata alla partizione $Z = \{I_1, \ldots, I_N\}$ di X e a valori nello spazio $\Omega := \mathcal{A}^{\mathbb{N}_0}$, dove $\mathcal{A} = \{1, \ldots, N\}$.

Allora è immediato verificare che se $p_i := \mu(I_i)$ per ogni $i = 1, \ldots, N$, allora l'entropia metrica di Z, $H_\mu(Z)$, è uguale alla funzione entropia $H(p_1, \ldots, p_n)$ (vedi equazione (1.1)) della sorgente di informazione associata. È quindi evidente che il prossimo passo sarà lo studio della crescita dell'entropia delle partizioni iterate.

Definizione 2.16. Sia $\{Z_n\}_{n\geq 1}$ la successione di partizioni iterate generate da una partizione Z finita e misurabile di X. L'*entropia metrica* $h_\mu(T, Z)$ del sistema dinamico (X, T, μ) *relativa a* Z e dipendente dalla misura di probabilità μ T-invariante è data da

$$h_\mu(T, Z) := \lim_{n\to\infty} \frac{H_\mu(Z_n)}{n}$$

L'esistenza del limite è garantita dalla Proposizione 2.12,(iv) (da cui otteniamo $H_\mu(Z_{n+m}) \leq H_\mu(Z_n) + H_\mu(Z_m)$) e dal Teorema 1.1.

Notiamo che questa definizione coincide con la Definizione 1.2, data per la sorgente di informazione generata dalla rappresentazione simbolica φ_Z.

Enunciamo ora due proprietà dell'entropia metrica relativa a una partizione che saranno usate nel seguito. La prima proposizione lega l'entropia metrica al rapporto di informazione tra presente, passato e futuro delle orbite di un sistema dinamico. La seconda stabilisce invece la vicinanza per l'entropia metrica di due partizioni "vicine".

Proposizione 2.13. *Per una partizione Z vale $h_\mu(T, Z) = h_\mu(T, Z_k)$ per ogni $k \geq 1$, e*

$$h_\mu(T, Z) = \lim_{n\to\infty} H_\mu(Z|T^{-1}Z_n) = \lim_{n\to\infty} H_\mu(T^{-n}Z|Z_n)$$

Dimostrazione. Osserviamo innanzitutto che $T^{-n}Z_k = \vee_{i=n}^{n+k-1} T^{-i}Z$, quindi si ottiene

$$h_\mu(T, Z_k) = \lim_{n\to\infty} \frac{H_\mu((Z_k)_n)}{n} = \lim_{n\to\infty} \frac{H_\mu(Z_{n+k-1})}{n} = h_\mu(T, Z)$$

Dimostriamo poi la prima definizione equivalente di entropia metrica. Applicando la Proposizione 2.12,(iv) ripetutamente otteniamo

$$H_\mu(Z_n) = H_\mu(T^{-1}Z_{n-1}) + H_\mu(Z|T^{-1}Z_{n-1}) = H_\mu(Z) + \sum_{j=0}^{n-1} H_\mu(Z|T^{-1}Z_j)$$

da cui si ricava

$$h_\mu(T, Z) = \lim_{n\to\infty} \frac{H_\mu(Z) + \sum_{j=0}^{n-1} H_\mu(Z|T^{-1}(Z_j))}{n} = \lim_{n\to\infty} H_\mu(Z|T^{-1}Z_n)$$

dove abbiamo applicato il primo criterio di Cesàro, considerando il fatto che $H_\mu(Z|T^{-1}Z_n)$ è una successione descrescente (vedi Proposizione 2.12,(iii)).

Scriviamo infine, usando nuovamente il punto (iv) della Proposizione 2.12,

$$H_\mu(Z_{n+1}) = H_\mu(Z_n) + H_\mu(T^{-n}Z|Z_n)$$

da cui si ricava per ricorrenza

$$\frac{H_\mu(Z_{n+1})}{n+1} = \frac{H_\mu(Z) + \sum_{k=1}^{n} H_\mu(T^{-k}Z|Z_k)}{n+1}$$

Come prima per ottenere la tesi, basta allora dimostrare che la successione $H_\mu(T^{-k}Z|Z_k)$ ha limite, per poter applicare il criterio di Cesàro. Scriviamo

$$H(T^{-k}Z|Z_k) \leq H_\mu(T^{-k}Z| \vee_{i=1}^{k-1} T^{-i}Z) = H_\mu(T^{-(k-1)}Z|Z_{k-1})$$

dove abbiamo applicato la diseguaglianza in Proposizione 2.12,(iii). □

Proposizione 2.14. *Date due partizioni finite e misurabili P e Q vale*

$$|h_\mu(T,P) - h_\mu(T,Q)| \leq H_\mu(P|Q) + H_\mu(Q|P)$$

Dimostrazione. Applicando ripetutamente la Proposizione 2.12,(iii)-(iv), si ottengono le relazioni

$$H_\mu(P_n) \leq H_\mu(P_n \vee Q_n) = H_\mu(Q_n) + H_\mu(P_n|Q_n)$$

$$H_\mu(P_n|Q_n) = H_\mu(P|Q_n) + H_\mu(T^{-1}P_{n-1}|P \vee Q_n) \leq$$
$$\leq H_\mu(P|Q) + H_\mu(T^{-1}P_{n-1}|T^{-1}Q_{n-1})$$

Queste implicano

$$\frac{H_\mu(P_n)}{n} \leq \frac{H_\mu(Q_n)}{n} + H_\mu(P|Q)$$

da cui si ottiene $h_\mu(T,P) - h_\mu(T,Q) \leq H_\mu(P|Q)$. Scambiando i ruoli di P e Q, si ottiene la tesi. □

È importante avere un concetto di entropia metrica di un sistema dinamico indipendente dalla scelta di una partizione. A differenza del caso dell'entropia topologica, è possibile farlo usando le partizioni.

Definizione 2.17. Sia μ una misura di probabilità T-invariante sullo spazio X. L'*entropia metrica* $h_\mu(T)$ del sistema dinamico (X,T) dipendente da μ è definita come

$$h_\mu(T) := \sup\{h_\mu(T,Z) \ : \ Z \text{ partizione finita}\}$$

Per il calcolo effettivo dell'entropia metrica di un sistema serve l'analogo di un ricoprimento aperto generante per le partizioni.

Definizione 2.18. Diciamo che un partizione finita e misurabile Z è *generante* per un sistema dinamico (X, T, μ) se $\vee_{i=0}^{\infty} T^{-i} Z$ genera la σ-algebra di Borel di X.

Possiamo allora enunciare il risultato che ci aspettiamo anche per l'entropia metrica.

Teorema 2.15 (Kolmogorov-Sinai)**.** *Se Z è una partizione generante per il sistema dinamico (X, T, μ), allora $h_\mu(T) = h_\mu(T, Z)$.*

A questo punto ci si può far tentare dalla voglia di liberarsi anche della dipendenza dalla misura μ. Tuttavia questo corrisponde allo studio delle proprietà topologiche del sistema.

Teorema 2.16 (Principio variazionale)**.** *Sia (X, T) sistema dinamico con T continua su X metrico compatto. Allora*

$$h_{top}(T) = \sup \{h_\mu(T) \ : \ \mu \text{ probabilità } T\text{-invariante}\}$$

Prima di studiare l'entropia metrica dei sistemi dinamici che abbiamo usato come esempi di base, estendiamo il concetto di coniugio ai sistemi dinamici con misura invariante.

Definizione 2.19. Due sistemi dinamici (X_1, T_1, μ_1) e (X_2, T_2, μ_2) si dicono *isomorfi* se esiste un coniugio misurabile $h : X_1 \to X_2$ che sia invertibile con inversa misurabile, e tale che $\mu_2(B) = \mu_1(h^{-1}(B))$ per ogni insieme $B \subset X_2$ misurabile.

Il concetto di isomorfismo è analogo al concetto di coniugio topologico per sistemi dinamici misurabili.

Dalla definizione di entropia metrica di un sistema dinamico segue la seguente proposizione.

Proposizione 2.17. *L'entropia metrica di un sistema dinamico è invariante per isomorfismo.*

Esempio 2.15 (Rotazioni del cerchio)*.* Ricordiamo che la misura di probabilità invariante di T_α è la misura di Lebesgue m. Sfruttando il Principio Variazionale, possiamo subito ottenere $h_m(T_\alpha) = 0$ per ogni $\alpha \in \mathbb{R}$. Vedi esercizio 2.8 per il calcolo diretto. $\triangle$

Esempio 2.16 (Mappa del panettiere). Abbiamo dimostrato che la mappa del panettiere T è coniugata topologicamente con lo shift τ su $\Omega = \{0,1\}^{\mathbb{N}_0}$ usando la rappresentazione simbolica φ_Z, indotta dalla partizione $Z = \left\{[0,\frac{1}{2}),[\frac{1}{2},1)\right\}$ (vedi esempio 2.6). Data su $X = [0,1]$ la misura di Lebesgue, che è invariante per la mappa del panettiere, consideriamo la misura ν su Ω indotta da φ_Z tramite $\nu(C) = m(\varphi_Z^{-1}(C))$ per ogni cilindro C di Ω. Notiamo che ν è τ-invariante, infatti

$$\nu(\tau^{-1}(C)) = m(\varphi_Z^{-1}(\tau^{-1}(C))) = m(T^{-1}(\varphi_Z^{-1}(C))) = m(\varphi_Z^{-1}(C)) = \nu(C)$$

Quindi la mappa del panettiere è isomorfa a (Ω,τ,ν). Per il calcolo dell'entropia metrica $h_m(T)$ possiamo quindi rifarci alla Proposizione 2.17 e usare l'esempio 2.19.

Analogamente si può usare il Teorema di Kolmogorov-Sinai e la definizione di entropia metrica relativa a una partizione (vedi esercizio 2.9). $\triangle$

Esempio 2.17 (Famiglia quadratica). Consideriamo la mappa T_4. Abbiamo dimostrato che la misura di probabilità invariante è data da $d\mu(x) = f(x)dx$, con $f(x)$ definita nell'equazione (2.12). Enunciamo il seguente risultato senza dimostrazione,

$$h_\mu(T_4) = \int_0^1 \log(|(T_4)'(x)|)\; d\mu(x) = 1$$

Il risultato $h_\mu(T_4) = 1$ si può ottenere anche usando l'esercizio 2.2, estendendo il coniugio topologico a un isomorfismo. $\triangle$.

Esempio 2.18 (Famiglia di Pomeau-Manneville). La misura invariante per T_z è della forma $d\mu_z(x) = f_z(x)dx$, dove $f_z \in L^1$ se e solo se $1 < z < 2$. Per $z \geq 2$ possiamo quindi calcolare l'entropia metrica solo sulle misure di probabilità invarianti che sono singolari rispetto alla misura di Lebesgue. In particolare possiamo farlo per le misure indotte dalle orbite periodiche e dal punto fisso nell'origine. Per queste misure l'entropia metrica è ovviamente nulla. Tuttavia, applicando il Principio Variazionale, sappiamo che esiste una misura di probabilità con entropia metrica uguale a $h_{top}(T_z) = 1$.

Nel caso $1 < z < 2$, vale invece di nuovo

$$h_{\mu_z}(T_z) = \int_0^1 \log(|(T_z)'(x)|)\; d\mu_z(x) > 0$$

e si ottiene sperimentalmente che $h_{\mu_z}(T_z) \to 0$ per $z \to 2$. $\triangle$

Esempio 2.19 (Dinamica simbolica). Per calcolare l'entropia metrica dello shift e degli shift di tipo finito su un alfabeto $\mathcal{A} = \{1, \dots, N\}$ rispetto a una misura di Markov, usiamo il Teorema di Kolmogorov-Sinai applicato alla partizione generante $Z = \{C_1, \dots, C_N\}$, dove $C_i := C(i, 0, 1)$.

Proposizione 2.18. *Sia μ_Π una misura di Markov associata a una matrice di transizione M irriducibile e aperiodica. Allora*

$$h_{\mu_\Pi}(\tau_M) = \sum_{j=1}^{N} p_j \left(-\sum_{i=1}^{N} \pi_{ji} \log(\pi_{ji}) \right)$$

dove $p = (p_j)$ rappresenta la distribuzione stazionaria di Π.

Dimostrazione. Per il calcolo dell'entropia, usiamo la definizione equivalente data nella Proposizione 2.13, ossia $h_{\mu_\Pi}(\tau_M) = \lim_{n\to\infty} H_{\mu_\Pi}(Z|\tau_M^{-1} Z_n)$ dove Z è la partizione generante dei cilindri C_i. Otteniamo

$$H_{\mu_\Pi}(Z|\tau_M^{-1} Z_n) = - \sum_{i_0, i_1, \dots, i_{n-1}, j} \mu_\Pi(C_{j, i_0, i_1, \dots, i_{n-1}}) \log \left(\frac{\mu_\Pi(C_{j, i_0, i_1, \dots, i_{n-1}})}{\mu_\Pi(C_{i_0, i_1, \dots, i_{n-1}})} \right)$$

dove $C_{j, i_0, i_1, \dots, i_{n-1}} = C_j \cap \tau_M^{-1} C_{i_0} \cap \dots \cap \tau_M^{-n} C_{i_{n-1}}$. Quindi

$$H_{\mu_\Pi}(Z|\tau_M^{-1} Z_n) = - \sum_{i_0, i_1, \dots, i_{n-1}, j} p_j \pi_{j i_0} \pi_{i_0 i_1} \dots \pi_{i_{n-2} i_{n-1}} \log \left(\frac{p_j \pi_{j i_0}}{p_{i_0}} \right) =$$

$$= - \sum_{i,j} p_j \pi_{ji} \log \left(\frac{p_j \pi_{ji}}{p_i} \right) = \sum_j p_j \left(- \sum_i \pi_{ji} \log(\pi_{ji}) \right)$$

sfruttando le uguaglianze $\sum_i \pi_{ji} = 1$ e $\sum_j p_j \pi_{ji} = p_i$. □

Notiamo che nel caso particolare dello shift, l'entropia metrica delle misure di probabilità invarianti μ del tipo dato nell'equazione (2.13), si ottiene dalla proposizione ponendo $\pi_{ji} = p_i$ per ogni j, dove $p_i = p(i)$. △

Esempio 2.20 (Rappresentazioni simboliche). Sia (X, T, μ) un sistema dinamico e sia data una partizione finita e misurabile $Z = \{I_1, \dots, I_N\}$, la rappresentazione simbolica φ_Z ha come immagine un sottospazio di $\Omega = \mathcal{A}^{\mathbb{N}_0}$, dove $\mathcal{A}$ è l'alfabeto associato a Z. Tale sottospazio è invariante per l'azione dello shift, e quindi definisce un sistema dinamico. Solitamente questo sistema dinamico non è un subshift di tipo finito, ma è comunque possibile

definire su $\varphi_Z(X)$ una misura ν_Z invariante indotta da φ_Z. Dati i cilindri della forma

$$C(\omega, k, n) = \{\bar{\omega} \in \varphi_Z(X) \ : \ \bar{\omega}_{k+i} = \omega_{k+i} \ \forall \ i = 0, \dots, n-1\}$$

si definisce ν_Z sui cilindri e poi si estende a tutta la σ-algebra. Si pone

$$\nu_Z(C(\omega, k, n)) = \mu \left(\bigcap_{i=0}^{n-1} T^{-(k+i)} I_{\omega_{k+i}} \right)$$

Quindi ne segue che $H_\mu(Z_n) = H_{\nu_Z}(\tilde{Z}_n)$ per ogni $n \geq 1$, dove $\tilde{Z}_n$ indica la partizione iterata secondo τ di $\tilde{Z} = \{C(1), \dots, C(N)\}$. Questo implica che $h_\mu(T, Z) = h_{\nu_Z}(\tau)$. Quindi la rappresentazione simbolica ha la stessa entropia metrica del sistema dinamico (relativamente alla partizione usata). $\triangle$

Esempio 2.21 (Processi stocastici). Dato un sistema dinamico (X, T, μ), le orbite $(T^n(x))_{n \in \mathbb{N}_0}$ del sistema associate ai punti $x \in X$, si possono considerare come realizzazioni di un processo stocastico stazionario discreto. Si può quindi studiare l'entropia metrica di un processo stocastico in generale. Facciamo adesso l'esempio di un caso particolare. Sia $\mathbb{Y} = (Y_n)_n$ un processo stocastico discreto e stazionario, con spazio base $(\Omega, \mathcal{P}, \nu)$ e a valori nello spazio $(X, \mathcal{B}, \mu)$. La stazionarietà del processo implica che $\mu = (Y_n)^* \nu$ per ogni n. Supponiamo in particolare che $X = [0, 1]$, che μ sia la misura di Lebesgue m normalizzata, e che il processo $\mathbb{Y}$ sia costituito da variabili aleatorie indipendenti. Tale processo viene chiamato *rumore bianco*. Dimostriamo che l'entropia metrica $h_m(\mathbb{Y})$ del rumore bianco è infinita. Consideriamo infatti la partizione

$$Z = \left\{ \left[0, \frac{1}{N}\right), \left[\frac{1}{N}, \frac{2}{N}\right), \dots, \left[\frac{N-1}{N}, 1\right] \right\}$$

Allora $H_m(Z) = \log N$, e per ogni n risulta che Z_n è costituita da tutte le possibili stringhe lunghe n. Usando l'indipendenza delle Y_n si ottiene $m(I) = N^{-n}$ per ogni $I \in Z_n$, quindi $H_m(Z_n) = nH_m(Z)$, da cui $h_m(\mathbb{Y}, Z) = \log N$. La dimostrazione si conclude considerando l'estremo superiore al variare delle partizioni finite.

Abbiamo quindi mostrato che è possibile avere sistemi dinamici a entropia metrica infinita. Al rumore bianco è infatti possibile associare uno shift sullo spazio $[0, 1]^{\mathbb{N}_0}$ con misura data dalla misura prodotto.

Osserviamo infine che per processi stocastici a valori in uno spazio X finito, vale invece $h_\mu(\mathbb{Y}) \leq \log(d(X))$. $\triangle$

2.5 Cenni di teoria ergodica

Fino ad ora ci siamo limitati alla scelta di una misura invariante di probabilità, per poter parlare di entropia metrica. Tuttavia non è sempre possibile trovare misure invarianti finite che abbiano anche un "senso fisico" per il nostro problema. Per esempio, nel caso delle mappe di Pomeau-Manneville, abbiamo visto che per valori del parametro $z \geq 2$, la misura invariante equivalente alla misura di Lebesgue è infinita. Quindi se vogliamo studiare le proprietà statistiche di orbite di punti "tipici" rispetto alla misura di Lebesgue, dobbiamo usare le proprietà della misura infinita, che in questo caso ha "senso fisico". In questo paragrafo diamo una definizione formale di misura con "senso fisico", un concetto che ci permette di dare una definizione di *sistema dinamico caotico*.

Enunciamo inoltre due risultati classici della *teoria ergodica*, con lo scopo di avere uno strumento forte per capire se lo studio di una singola orbita di un sistema dinamico può essere utile per ricavare proprietà globali del sistema.

Abbiamo visto che per un sistema dinamico possono esistere molte misure invarianti diverse e tra loro singolari. Diventa allora importante avere un modo di selezionarne una in particolare da usare per la classificazione del sistema. Noi ci occuperemo del caso di spazi X immersi in uno spazio euclideo, per cui la scelta naturale è quella di una misura che in qualche modo racchiuda le orbite di un insieme di punti di misura di Lebesgue positiva. Una nozione analoga si può dare nel caso generale di uno spazio di Lebesgue.

Definizione 2.20. Sia X un sottoinsieme misurabile di uno spazio euclideo $\mathbb{R}^k$, e sia m_k la misura di Lebesgue k-dimensionale su $\mathbb{R}^k$. Sia T trasformazione misurabile di X in sè. Diciamo che una misura di probabilità μ è una *misura fisica* per il sistema dinamico (X, T) se esiste un insieme $S \subset X$ tale che $m_k(S) > 0$ e per ogni $x \in S$ valga, a meno di sottosuccessioni,

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x)}$$

nel senso della topologia debole* nello spazio delle misure di probabilità, dove δ_y indica la Delta di Dirac centrata in y.

Osserviamo che una misura fisica risulta invariante per l'azione di T.

Scelte allora le misure invarianti da usare per lo studio di un sistema dinamico, possiamo dare una definizione di sistema caotico.

Definizione 2.21. Diciamo che un sistema dinamico (X, T) è *caotico* se esiste una misura fisica μ per cui $h_\mu(T) > 0$. Se invece il sistema dinamico non è periodico ma per ogni misura fisica vale $h_\mu(T) = 0$, allora diciamo che il sistema è *debolmente caotico*.

Esempi di sistemi caotici sono la mappa del panettiere, la mappa T_4 della famiglia quadratica e le mappe di Pomeau-Manneville con $1 < z < 2$. Esempi di sistemi debolmente caotici sono le rotazioni del cerchio irrazionali, la mappa T_{λ_∞} della famiglia quadratica e le mappe di Pomeau-Manneville per $z \geq 2$.

La definizione di sistema debolmente caotico come data sopra è molto generale, e comprende molti sistemi che non sono generalmente considerati sistemi caotici, come per esempio sistemi che hanno orbite periodiche come attrattori. Tuttavia questa definizione va considerata nell'ottica di chi studia un sistema dinamico senza conoscerne a priori le caratteristiche, e misura l'entropia metrica usando delle orbite campione con i metodi che esporremo nei prossimi capitoli. Basta pensare all'analisi di una serie temporale sperimentale. Lo scopo è quello di classificare l'imprevedibilità del sistema che l'ha generata, e questa prima divisione in sistemi caotici e debolmente caotici è un primo passo, legato agli indicatori classici. Nei prossimi capitoli vedremo come migliorare questa classificazione.

Proprio allo scopo di studiare sistemi dinamici senza conoscerne le caratteristiche a priori, è importante chiedersi se le proprietà di una singola orbita rappresentano quelle del sistema nella sua globalità. Questa caratteristicà è legata al concetto di *misura ergodica*, che esporremo, ed è una proprietà che si suppone sempre verificata nello studio di serie sperimentali.

Definizione 2.22. Una misura μ si dice *ergodica* per un sistema dinamico (X, T) se per ogni insieme misurabile $B \subset X$ invariante ($T^{-1}B = B$) vale $\mu(B) = 0$ oppure $\mu(X \setminus B) = 0$.

Per la trattazione di misure invarianti μ infinite, bisogna supporre che il sistema (X, T, μ) sia *conservativo*, ossia che non esista un insieme $W \subset X$ con $\mu(W) > 0$ tale che gli insiemi $(T^{-n}W)_n$ siano tutti disgiunti al variare di $n \geq 0$. L'ipotesi di conservatività è soddisfatta automaticamente nel caso di una misura invariante di probabilità. Un esempio di sistema non conservativo è dato da $X = \mathbb{R}$, $\mu = m$ e $T : x \mapsto x + 1$. Basta prendere $W = (0, 1)$. Nel seguito supponiamo che la conservatività sia sempre soddisfatta.

Vediamo ora alcune conseguenze dell'ergodicità di una misura. Diciamo che una proprietà vale μ-quasi ovunque o per μ-quasi ogni x (μ-q.o.), se vale ovunque a meno di un insieme di misura nulla rispetto a μ.

Proposizione 2.19. *Sia (X, T, μ) un sistema dinamico con μ misura T-invariante. La misura μ è ergodica se e solo se una funzione $f : X \to \mathbb{R}$ misurabile è T-invariante ($f(x) = f(T(x))$ per μ-q.o. x) se e solo se è costante μ-q.o.*

Dimostrazione. Supponiamo che esista una funzione misurabile, T-invariante e non costante. Allora esiste un valore $c \in \mathbb{R}$ tale che $L_c := \{x : f(x) \geq c\}$ ha misura positiva e lo stesso vale per il suo complementare. Inoltre poiché f è T-invariante anche l'insieme L_c è T-invariante, quindi μ non è ergodica.

Viceversa, se μ non è ergodica, allora esiste un insieme B di misura positiva, con complementare di misura positiva e T-invariante. Allora la funzione caratteristica di B è misurabile, T-invariante e non costante. □

Teorema 2.20 (Birkhoff)**.** *Sia (X, T, μ) un sistema dinamico con μ misura T-invariante ed ergodica. Allora per ogni funzione $f : X \to R$ nello spazio $L^1(X, \mu)$, vale*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \begin{cases} \int_X f(x)\ d\mu(x) & \text{se } \mu(X) = 1 \\ 0 & \text{se } \mu(X) = \infty \end{cases}$$

per μ-q.o. $x \in X$.

Nel caso di una misura di probabilità si ottengono quindi più informazioni dal Teorema di Birkhoff. Inoltre è possibile dare un legame puntuale tra la misura degli insiemi delle iterate di una partizione e l'entropia metrica relativa a questa partizione. Ricordiamo che se Z è una partizione finita misurabile, indichiamo con $(Z_n)_n$ la successione delle sue partizioni iterate e con $Z_n(x)$ l'elemento della partizione Z_n che contiene x.

Teorema 2.21 (Shannon-McMillan-Breiman)**.** *Sia (X, T, μ) un sistema dinamico con μ misura di probabilità invariante ed ergodica. Sia Z una partizione finita e misurabile, allora*

$$\lim_{n\to\infty} -\frac{1}{n} \log(\mu(Z_n(x))) = h_\mu(T, Z)$$

dove il limite vale sia nel senso della convergenza μ-q.o. sia nel senso della convergenza in $L^1(X, \mu)$.

Il Teorema di Shannon-McMillan-Breiman implica in particolare, che un tipico insieme della partizione Z_n ha misura che tende a zero con velocità esponenziale e coefficiente dato dall'entropia metrica relativa alla partizione. Tale risultato si esprime formalmente come la proprietà di equipartizione.

Corollario 2.22 (Proprietà di equipartizione)**.** *Nelle ipotesi del teorema precedente, data una partizione Z finita e misurabile, per ogni $\varepsilon > 0$ esiste un intero n_0 tale che per ogni $n \geq n_0$ la partizione Z_n si può scrivere come l'unione di due gruppi di insiemi $Z_n = T_n \cup N_n$ (insiemi "Typical" e "Negligible") tali che*

$$\mu\left(\bigcup_{B \in N_n} B\right) < \varepsilon$$

$$\forall\ B \in T_n \qquad 2^{-(n+1)(h_\mu(T,Z)+\varepsilon)} < \mu(B) < 2^{-(n+1)(h_\mu(T,Z)-\varepsilon)}$$

Esempi di misure ergodiche sono la misura di Lebesgue per la mappa del panettiere e per le rotazioni sul cerchio irrazionali, le misure di Markov per subshift di tipo finito con matrice stocastica irriducibile e aperiodica, la misura definita nell'equazione (2.12) per la mappa T_4 della famiglia quadratica e le misure invarianti $d\mu_z(x) = f_z(x)dx$ equivalenti alla misura di Lebesgue per le mappe T_z di Pomeau-Manneville per ogni $z > 1$.

Esercizi

2.1. Dimostrare che $h(x) = 4\left(x - \frac{1}{2}\right)$ è un coniugio topologico tra i sistemi dinamici $([0,1], T_1)$ e $([-1,1], T_2)$, dove T_1 è il sistema della famiglia quadratica per $\lambda = 4$ e $T_2(y) = 2 - y^2$.

2.2. Dimostrare che $h(x) = \left[\sin\left(\frac{\pi}{2}x\right)\right]^2$ è un coniugio topologico tra i sistemi con $X_1 = X_2 = [0,1]$ dati da

$$T_1(x) = \begin{cases} 2x & 0 \leq x \leq \frac{1}{2} \\ 2(1-x) & \frac{1}{2} \leq x \leq 1 \end{cases}$$

e $T_2(y) = 4y(1-y)$. La mappa T_1 è detta *mappa tenda.*

2.3. Dimostrare la buona definizione dell'espansione binaria dei punti $x \in [0,1]$ tranne che per i punti diadici, e le proprietà del coniugio topologico costruito sulla mappa del panettiere nell'esempio 2.6.

2.4. Dimostrare la Proposizione 2.5.

2.5. (*****) Dimostrare che le mappe T_z della famiglia di Pomeau-Manneville sono, per ogni $z > 1$, coniugate topologicamente alla mappa del panettiere.

2.6. Dimostrare che la mappa del panettiere e le rotazioni sul cerchio preservano la misura di Lebesgue m normalizzata su $[0,1]$. Dire per quale dei sistemi esistono misure di probabilità invarianti indotte da orbite periodiche come nell'equazione (2.11).

2.7. Dimostrare la Proposizione 2.17.

2.8. Dimostrare che per le rotazioni sul cerchio T_α vale $h_m(T_\alpha) = 0$ per ogni $\alpha \in \mathbb{R}$, usando la definizione di entropia e la Proposizione 2.12,(i) per la partizione generante.

2.9. Data la partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$ per la mappa del panettiere, calcolare $h_m(T, Z)$ per la misura di Lebesgue m.

2.10. Dimostrare, usando l'esercizio 2.2, che per la mappa T_4 della famiglia quadratica vale $h_\mu(T_4) = 1$.

2.11. Dimostrare che per un processo stocastico $\mathbb{Y}$ discreto e stazionario a valori in uno spazio di probabilità (X, μ) finito, vale $h_\mu(\mathbb{Y}) \leq \log(d(X))$.

2.12. Dimostrare che la misura di Lebesgue è ergodica per le rotazioni sul cerchio T_α con α irrazionale. Utilizzare la Proposizione 2.19 e lo sviluppo in serie di Fourier per funzioni misurabili in $[0,1]$.

Capitolo 3

La complessità di Kolmogorov-Chaitin

Nei primi due capitoli, abbiamo studiato le principali proprietà delle sorgenti di informazione e dei sistemi dinamici discreti, soffermandoci esclusivamente sull'aspetto statistico. In particolare, abbiamo introdotto i concetti di entropia di Shannon (Definizione 1.2) per le sorgenti, e di entropia metrica relativa a una partizione e assoluta (Definizioni 2.16 e 2.17) per un sistema dinamico con misura di probabilità invariante fissata. Abbiamo inoltre sottolineato come, grazie al concetto di rappresentazione simbolica, sia possibile considerare l'entropia metrica di un sistema dinamico, relativa a una partizione, come l'entropia di Shannon della sorgente di informazione associata al sistema rispetto alla partizione fissata. L'entropia metrica fornisce quindi, un utile strumento per misurare il contenuto di informazione medio per le orbite di un sistema dinamico.

Quest'approccio richiede tuttavia la conoscenza della misura invariante di probabilità del sistema dinamico, ossia della statistica generata dalla sorgente di informazione. Se vogliamo invece analizzare serie temporali (ossia successioni di dati) che provengono dal mondo reale, come misurazioni fatte su esperimenti, dobbiamo cercare di sviluppare una teoria che ci permetta di trarre informazioni circa il sistema che stiamo studiando, esclusivamente dalla serie temporale di cui disponiamo. Vogliamo cioè poter dire dall'analisi delle misurazioni fatte su un esperimento, per esempio di origine fisica, se il segnale che abbiamo registrato proviene da un sistema caotico o no, o addirittura riuscire a quantificare il grado di caoticità del sistema.

In questo capitolo, seguendo [LV], trattiamo dei concetti provenienti dalla teoria dell'informazione che possono essere utili a questo scopo. Nel

prossimo capitolo, applicheremo questi concetti ai sistemi dinamici.

3.1 Il concetto di casualità

Volendo analizzare una singola serie di dati, dobbiamo innanzitutto specificare cosa intendiamo per *casualità* per una singola stringa. Abbiamo infatti dato la definizione di sistema dinamico "casuale" o caotico, e per estensione, di sorgente di informazione caotica, ma queste nozioni si basano su una proprietà media del sistema, e non ci dicono in che senso parlare di casualità per una singola stringa.

Supponiamo infatti di considerare le possibili realizzazioni di un semplice esperimento: il lancio di una moneta. Si tratta di un processo stocastico stazionario, le cui variabili aleatorie sono indipendenti e tutte distribuite secondo la distribuzione $p(Testa) = p(Croce) = \frac{1}{2}$. Consideriamo allora una serie temporale ottenuta con N lanci della moneta. La teoria della probabilità ci dice che ogni possibile successione di N simboli dall'alfabeto $\mathcal{A} = \{C, T\}$ ha esattamente probabilità $\frac{1}{2^N}$. Se quindi $N = 10$, avremo per esempio

$$\begin{aligned} Prob(CCCCCCCCCC) &= \frac{1}{2^{10}} \\ Prob(CTCTCTCTCT) &= \frac{1}{2^{10}} \\ Prob(CCTTCTTCTC) &= \frac{1}{2^{10}} \end{aligned}$$

Analogamente, considerando la teoria dell'informazione secondo Shannon, essendo tutte le possibili stringhe lunghe N equiprobabili, esse avranno lo stesso contenuto di informazione, dato da $\log_2(2^N) = N$. Così le nostre stringhe lunghe 10, avranno tutte e tre contenuto di informazione uguale a 10 bits.

Se guardiamo però attentamente le tre stringhe dell'esempio, appare evidente che qualche differenza tra le tre possibili realizzazioni del nostro esperimento c'è. Anzi, questa differenza è talmente evidente, che sembra sorprendente non riuscire a coglierla utilizzando il contenuto di informazione definito da Shannon. La prima stringa è infatti semplicemente "la ripetizione di 10 simboli C", la seconda è "la ripetizione 5 volte della coppia CT", mentre la terza stringa è "CCTTCTTCTC"! La differenza sta dunque nella facilità di "descrizione" della stringa. Vediamo qualche altro esempio. Consideriamo due numeri naturali, 10000 e 19654. I due numeri

sono dello stesso ordine di grandezza, eppure il primo si può esprimere in maniera più compatta in varie forme. Ad esempio possiamo dire: "diecimila" e "diciannovemila, seicento cinquantaquattro"; oppure scriverli come: 10^5 e 1.9654×10^5. Notiamo che se avessimo scelto numeri di un ordine di grandezza molto maggiore, per esempio dell'ordine di 10^{100}, allora la differenza tra le due descrizioni sarebbe stata molto più accentuata. Così accadrebbe lo stesso per i risultati del nostro esperimento di lancio di moneta.

Ma a questo punto sorge un altro interrogativo: quanti e quali modi per esprimere una stringa dobbiamo considerare? In sostanza, è possibile che ci sia un modo per esprimere i numeri 10000 e 19654, in modo che la descrizione del secondo sia più breve di quella del primo? e quante e quali descrizioni dobbiamo provare? Chiariamo questo punto con un altro esempio. Siano dati i due numeri naturali 682 e 805. Sono numeri dello stesso ordine di grandezza, e se usiamo le descrizioni precedenti, non troviamo una grande differenza. Se invece proviamo a scrivere i due numeri in base 2, allora ci accorgiamo che il primo diventa 1010101010, mentre il secondo diventa 1100100101. La descrizione di 682 è quindi più semplice di quella di 805 (notiamo che le due stringhe binarie corrispondono al secondo e al terzo esempio di lanci di moneta, in cui a C è stato sostituito 1 e a T 0).

Vediamo un esempio più interessante. Consideriamo il numero reale trascendente π. La sua espressione in base decimale è infinita e aperiodica, e, come tutti sanno, si ha

$$\pi = 3.14159265358979323846264338328\ldots$$

Ma quanto è lunga la descrizione della stringa numerica delle cifre decimali di π? Ci aspetteremmo, in considerazione del fatto che non riusciamo a scorgere alcuna legge nel ripetersi delle cifre decimali, che la risposta sia che la descrizione migliore è semplicemente la ripetizione della stringa, così come accadeva per la terza stringa di lanci della moneta. La risposta a questa domanda, si ottiene invece notando che le cifre decimali di π possono essere ottenute tramite diversi algoritmi (uno di questi è stato usato dal mio pc per calcolare le prime 30 cifre scritte sopra). Per esempio, sappiamo che $\pi = 4 \arctan 1$, quindi basta osservare che

$$\frac{\pi}{4} = \arctan 1 = \sum_{n=0}^{\infty} \frac{D^n(\arctan x)(0)}{n!} = 1 - \frac{1}{3} + \frac{1}{5}\ldots$$

Quest'uguaglianza ci fornisce uno degli algoritmi per il calcolo delle cifre decimali di π. Altri algoritmi possono essere ottenuti usando le altre mirabili espressioni in serie di π. La risposta alla domanda quanto sia lunga

la descrizione della stringa delle cifre decimali di π, è quindi che tale descrizione è lunga quanto la descrizione dell'algoritmo che voglio usare, più la descrizione del numero naturale che esprime quante cifre decimali voglio conoscere. La prima parte è una costante, e la dipendenza dalla lunghezza N della stringa numerica che voglio ottenere è solo nella seconda parte della risposta. Dimostreremo più avanti che tale dipendenza è dell'ordine $\log N$. Sorprendentemente poco, se si considera che per la maggior parte dei numeri reali si ha una dipendenza dell'ordine N.

Questa discussione suggerisce allora, che un primo tentativo per introdurre un concetto di casualità per una stringa singola è legato alla ricerca di possibili strutture all'interno della stringa. Queste strutture ne favorirebbero la descrizione, quindi più strutture sono presenti meno casuale sarebbe la stringa.

Ripercorrendo storicamente i tentativi di definire un concetto di casualità per stringhe, notiamo come i primi tentativi presero spunto dalla teoria della probabilità, e riguardarono stringhe infinite. Abbiamo detto che lo scopo è quello di cercare strutture. La più semplice di queste strutture è la frequenza di ogni simbolo.

Definizione 3.1 (R. von Mises, 1919)**.** Una stringa infinita $\omega \in \{0,1\}^{\mathbb{N}}$ è *casuale* se soddisfa le seguenti condizioni:

1. sia $f_n = d(\{1 \leq i \leq n \ : \ \omega_i = 1\})$. Allora per un fissato $p \in [0,1]$ deve valere
$$\lim_{n\to\infty} \frac{f_n}{n} = p$$

2. sia $\phi : \{0,1\}^* \to \{0,1\}$ una "regola di selezione di posto", ossia una funzione parziale[1] scelta con lo scopo di selezionare gli indici n per cui $\phi(\omega_1^{n-1}) = 1$, dove $\omega_1^{n-1} := (\omega_1, \ldots, \omega_{n-1})$. Allora deve valere la proprietà 1, con lo stesso p, anche per ogni sottosuccessione $(\omega_{n_k})_k$ scelta con regole si selezione ammissibili [2].

Nella definizione di Von Mises, il punto 1 richiede che ci sia una convergenza verso una frequenza limite. Il punto 2 richiede invece che la stessa frequenza limite valga per ogni sottosuccessione. Questo garantisce che, se per esempio la stringa ω rappresenta una realizzazione di lanci di moneta, allora non esiste un metodo per decidere quando scommettere che garantisca una vincita.

[1] Vedi la Sezione 3.2 per la definizione

[2] Quali siano le regole ammissibili viene chiarito più avanti

La prima domanda è circa l'esistenza di stringhe Von Mises casuali. Il problema sta nella scelta della classe delle funzioni parziali ammissibili. Supponiamo che tutte le funzioni parziali siano ammissibili. Allora data una stringa ω casuale, scegliamo ϕ_0 e ϕ_1 nel modo seguente: poniamo $\phi_0(\omega_1^{i-1}) = 1$ se $\omega_i = 1$, e non definita altrimenti; poniamo poi $\phi_1(\omega_1^{i-1}) = 1 - \omega_i$. Supponendo che valga la proprietà 2 per entrambe le funzioni, si trova $p = 1$ per ϕ_0, e $p = 0$ per ϕ_1. Dunque ω non può essere casuale.

Questa contraddizione è risolta scegliendo come classe di funzioni parziali ammissibili, le *funzioni ricorsive parziali* (vedi Definizione 3.4). Nel 1940, A. Church ha dimostrato che, in questo caso e con $p = \frac{1}{2}$, le stringhe infinite Von Mises casuali formano un insieme di misura totale. Inoltre A. Wald (che aveva già risolto il problema dell'esistenza di stringhe casuali con un'altra scelta di funzioni amissibili) ha dimostrato che l'insieme delle stringhe Von Mises casuali, scelte con funzioni ricorsive, ha per ogni frequenza fissata la cardinalità del continuo.

Un esempio di una stringa infinita non Von Mises casuale è la *stringa di D.G. Champernowne*, data da

$$0\ 1\ 00\ 01\ 10\ 11\ 000\ 001\ \ldots$$

la concatenazione di tutte le stringhe finite di $\{0,1\}^*$, in ordine per lunghezza e poi lessicografico. Tuttavia la stringa di Champernowne soddisfa un'altra definizione di casualità dovuta a E. Borel, in cui si suppone che tutte le stringhe finite abbiano frequenza di apparizione dipendente solo dalla loro lunghezza. La stringa di Champernowne comunque è costruttiva, mentre ogni stringa che sia Von Mises casuale non può esserlo. Quindi la definizione di Von Mises comincia a creare la differenza tra pure proprietà statistiche e proprietà di "struttura".

La definizione di Von Mises non soddisfa ancora tutte le richieste che si possono fare a una buona definizione di casualità. Supponiamo infatti che nello spazio $\Omega = \{0,1\}^{\mathbb{N}}$ di misura finita, si dimostri che una certa proprietà vale per quasi ogni stringa. Allora, se il concetto di "casualità" di una stringa è collegato al concetto di "tipicità", ci aspettiamo che una stringa casuale soddisfi tutti i criteri che sappiamo valere per quasi ogni stringa. J. Ville dimostrò che esistono stringhe Von Mises casuali che non soddisfano due di queste proprietà: la *legge di ricorrenza infinita* e la *legge del logaritmo iterato*. La prima legge dice che per quasi ogni stringa vale $f_n = \frac{n}{2}$ infinite volte, mentre la seconda legge dice che per quasi ogni stringa $\limsup_n \frac{2f_n - n}{\sqrt{2n\log(\log n)}} = 1$.

Consideriamo allora un insieme $\mathcal{I}$ di test sulle stringhe infinite.

Definizione 3.2 (P. Martin-Löf, 1966)**.** Una stringa infinita $\omega \in \{0,1\}^{\mathbb{N}}$ è *casuale* se appartiene all'insieme $\cap_{i\in\mathcal{I}} P_i$, dove P_i è l'insieme di stringhe che soddisfa il test i-esimo, e dove si pone $\mathcal{I}$ dato dai test soddisfatti da quasi ogni stringa, e per i quali l'insieme delle stringhe che non soddisfa un test è ricorsivamente enumerabile (vedi Definizione 3.6).

A questo punto, prima di passare al successivo passo, l'introduzione di casualità per stringhe finite e senza dipendenza da una "scelta" di leggi o funzioni, esponiamo brevemente i concetti base della teoria della computabilità.

3.2 Nozioni base di teoria della computabilità

Uno dei problemi emersi nella discussione della definizione del concetto di casualità per una stringa simbolica, è legato alla scelta della classe di funzioni da considerare. In particolare, è possibile individuare nell'insieme delle funzioni $\varphi : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$ una classe di funzioni "effettivamente computabili"? La risposta è data dalla teoria della computabilità. In questo paragrafo, illustriamo le idee che stanno alla base di tale teoria, e che ci forniscono le nozioni di cui avremo bisogno nel seguito.

Nel nostro approccio, possiamo considerare come passo fondamentale i lavori di Alan Turing. Nel lavoro [Tu], Turing descrive una *macchina* che possa simulare ogni calcolo possibile. Alcuni brani liberamente tradotti dal suo articolo descrivono in maniera molto intuitiva la sua idea.

> Il calcolo si effettua normalmente scrivendo certi simboli su un foglio...Si può supporre di sostituire il foglio con un nastro diviso in quadrati, e si suppone che si possano scrivere un numero finito di simboli.
> Il comportamento della macchina calcolatrice è determinato in ogni istante dal simbolo che osserva e dal suo "stato mentale". Inoltre la macchina può osservare contemporaneamente un numero limitato di simboli, e ha un numero finito di stati mentali possibili.
> La macchina effettua semplici operazioni, cambiando non più di un simbolo.
> Lo stato del sistema è descritto dalla successione di simboli sul nastro, da quelli osservati dalla macchina e dallo "stato mentale".

> Le operazioni semplici che effettua la macchina sono: (i) cambiare un simbolo tra quelli osservati; (ii) cambiare quadrato oseervato spostandosi non più di un numero fissato di quadrati.

Seguendo le parole di Turing possiamo dare una definizione formale della sua macchina calcolatrice.

Definizione 3.3. Una *macchina di Turing* consiste di un *programma finito* che può manipolare una lista di cellette, il *nastro*, con un *pointer*. Il programma ha un numero finito di stati Q e ogni celletta contiene uno dei simboli dell'insieme $\mathcal{A} = \{\lambda, 0, 1\}$. Il tempo è discreto, e ad ogni istante di tempo il pointer è su una particolare celletta, che è quella su cui operare.

Descriviamo adesso quale tipo di operazioni effettua il pointer e in che modo. Al tempo $t = 0$ il pointer sia su una celletta fissata, la *celletta di partenza*, e il programma sia in uno stato $q_0 \in Q$, detto *stato iniziale*. Supponiamo inoltre che tutte le cellette contengano il simbolo λ, tranne al più un numero finito, tra di loro contigue. Tali cellette si estendano verso destra a partire dalla celletta iniziale. Tale successione di cellette è detta *input*. Le operazioni ammesse siano:

- scrivere $0, 1$ oppure λ sulla celletta su cui operare;
- spostare il pointer di una celletta a destra o a sinistra.

Alla fine di ogni operazione (che richiede un'unità di tempo) il programma sia in uno degli stati dell'insieme Q.

Indicando allora con $\mathcal{O} = \{\lambda, 0, 1, D, S\}$ l'insieme delle possibili operazioni, possiamo dire che il programma di una macchina di Turing obbedisce a un insieme di *regole*, funzioni $r : Q \times \mathcal{A} \to \mathcal{O} \times Q$. In generale l'unione dei domini delle regole è strettamente contenuto nell'insieme $Q \times \mathcal{A}$. Se a un certo istante il sistema è in uno stato al di fuori del dominio delle regole, la macchina si ferma. L'*output* di una macchina di Turing è la stringa binaria massimale circondata da cellette con il simbolo λ che si ottiene quando la macchina si ferma. A questo scopo si può inserire uno stato $q_1 \in Q$, detto *stato finale*, su cui non è definita alcuna regola, che fa quindi fermare la macchina.

La macchina di Turing permette di definire in maniera formale le funzioni computabili sugli interi, usando la rappresentazione degli interi tramite stringhe binarie finite. Per definire funzioni su n-uple di interi sarà necessario utilizzare rappresentazioni autodelimitanti (ossia tramite codici prefissi) degli interi.

Definizione 3.4. Una funzione *ricorsiva parziale* o *computabile* è un'operazione, associata a una macchina di Turing, che trasforma una n-upla di interi ($n \geq 1$), che rappresentano l'input, in un intero, l'output della macchina. Se la macchina di Turing associata a una funzione si ferma su ogni input, la funzione si dice *ricorsiva totale*.

Esempio 3.1. Date due stringhe $s, t \in \{0,1\}^*$, esempi di funzioni ricorsive sono:

- la funzione *complementare* $f : \{0,1\}^* \to \{0,1\}^*$ data da $f(s) = s^c$, dove indichiamo $0^c = 1$ e $1^c = 0$, e per le parole finite supponiamo che il complementare commuti con la concatenazione;
- $g : (\{0,1\}^*)^2 \to \{0,1\}^*$ data da $g(s,t) = s^c t$;
- $h : \{0,1\}^* \to \{0,1\}^*$ data da $h(\tilde{s}t) = t$, dove con $\tilde{s}$ indichiamo la codifica autodelimitante delle stringhe binarie $\tilde{s} = (s_1 s_1 s_2 s_2 \dots s_n s_n^c)$

Le funzioni f e g sono ricorsive totali, mentre la funzione h è ricorsiva parziale, non essendo definita per esempio sulla stringa binaria (1111). È istruttivo pensare a come definire possibili macchine di Turing per la rappresentazione delle funzioni f, g e h. $\triangle$

Nella classe di funzioni computabili rientrano naturalmente gli esempi classici di somma e differenza di numeri naturali, ma anche tutte le funzioni che ci aspettiamo di saper calcolare.

> **Tesi di Church:** La classe delle funzioni algoritmicamente computabili numericamente (in senso intuitivo) coincide con la classe delle funzioni ricorsive parziali

Studiamo ora qualche proprietà delle macchine di Turing. Innanzitutto osserviamo che è possibile enumerare le macchine di Turing, fornendo una bigezione dell'insieme delle macchine con i numeri naturali.

Dalla definizione, abbiamo visto che ciò che identifica in maniera univoca una macchina di Turing è l'insieme delle regole il programma esegue. Costruiamo un codice prefisso per le macchine di Turing. L'insieme $Q \cup \mathcal{A} \cup \mathcal{O}$ contiene $(d(Q)+5)$ elementi, quindi ogni suo elemento può essere descritto, in modo non prefisso, con $s \leq \lceil \log(d(Q)+5) \rceil$ bits, e indichiamo con ϵ la codifica fissata. Data una macchina di Turing T, siano k le sue regole, e siano date dalle quadruple (p_i, t_i, s_i, q_i), per $i = 1, \dots, k$. Sia E_1 la codifica per numeri interi introdotta nell'equazione (1.5). Il codice prefisso $E(T)$ per le macchine di Turing è definito da

$$E(T) := E_1(s)\ E_1(k)\ \epsilon(p_1)\epsilon(t_1)\epsilon(s_1)\epsilon(q_1)\dots\epsilon(p_k)\epsilon(t_k)\epsilon(s_k)\epsilon(q_k) \qquad (3.1)$$

Per questo codice vale $|E(T)| \leq 4sk + 2\log(s+1) + 2\log(k+1) + 2$. Inoltre ordinando in maniera lessicografica le stringhe binarie $E(T)$ è possibile stabilire un ordinamento delle rispettive macchine di Turing.

Definizione 3.5. Il *numero di Gödel* $n(T)$ di una macchina di Turing T è il numero d'ordine della stringa $E(T)$ nell'ordinamento delle stringhe codice che rappresentano le macchine di Turing.

Possiamo allora parlare della i-esima macchina di Turing T_i, riferendoci alla macchina con numero di Gödel uguale a i. Analogamente si può fare per le funzioni ricorsive parziali, che si possono numerare in modo che φ_i sia la funzione calcolata dalla i-esima macchina di Turing T_i.

La numerazione delle macchine di Turing permette la costruzione di macchine di Turing *universali*, nel senso che possono simulare il comportamento di una qualsiasi altra macchina di Turing. Vediamo come usare la codifica introdotta nell'equazione (3.1) per costruire una macchina universale. In generale le macchine di Turing universali sono una quantità numerabile.

Esempio 3.2 (Macchina di Turing universale). Sia U una macchina di Turing che accetta, come input, stringhe della forma $(1^i 0p)$, dove $p \in \{0,1\}^*$. La macchina U funzioni nel seguente modo: sul lato sinistro del nastro rispetto all'input, genera in ordine lessicografico, secondo la lunghezza, tutte le stringhe di $\{0,1\}^*$; per ogni stringa generata decide se è della forma $E(T)$ per qualche T; ogni volta che la stringa generata descrive una macchina di Turing, sostituisce uno dei simboli "1" dell'input con il simbolo λ e legge il simbolo a destra; quando legge il simbolo "0", esegue sulla stringa p le regole specificate dalla stringa binaria $E(T)$ che ha sul lato sinistro del nastro. Quindi, possiamo scrivere $U(1^i 0p) = T_i(p)$.

Definizione 3.6. Un insieme $A \subset \{0,1\}^*$ si dice *ricorsivamente enumerabile* se è vuoto o è l'immagine di una funzione ricorsiva parziale f. Si dice che f *enumera* A.

In maniera equivalente, un insieme A è ricorsivamente enumerabile se e solo se esiste una macchina di Turing T tale che $T(p)$ dà come output "sì" se $p \in A$, e altrimenti dà come output "no" o non si ferma.

Definizione 3.7. Un insieme $A \subset \{0,1\}^*$ si dice *ricorsivo* se la sua funzione caratteristica è una funzione ricorsiva totale.

Equivalentemente, un insieme A è ricorsivo se e solo se A e A^c sono ricorsivamente enumerabili, ossia se e solo se esiste una macchina di Turing T tale che $T(p)$ dà come output "sì" se $p \in A$ e "no" altrimenti.

Esempi di insieme ricorsivi sono: l'insieme dei numeri dispari; l'insieme dei numeri primi; ogni insieme finito; ogni insieme con complementare finito. Come esempio di insieme ricorsivamente enumerabile, e non ricorsivo, indichiamo l'insieme

$$\{n \in \mathbb{N} \ : \ 0^n \text{ appare tra le cifre decimali di } \pi\}$$

Proposizione 3.1. *Ogni insieme ricorsivamente enumerabile infinito contiene un insieme ricorsivo infinito.*

Dimostrazione. Sia A un insieme infinito e ricorsivamente enumerabile, e sia f la funzione ricorsiva che enumera A. Definiamo una nuova funzione ricorsiva g tramite $g(0) = f(0)$, e $g(n+1) = f(m)$ dove m è il più piccolo valore tale che $f(m) > g(n)$. Sia B l'immagine di g. Poiché A è infinito, anche B lo è. Inoltre g enumera B in ordine crescente, quindi data una stringa p è possibile decidere se $p \in B$ o no in tempo finito, semplicemente analizzando le stringhe di B che sono minori o uguali di p. □

La definizione di macchina di Turing permette quindi di stabilire quali calcoli devono considerarsi "effettivamente computabili" e quali no. Tuttavia resta aperta una questione importante legata alla possibilità di una macchina di Turing di non fermarsi su alcuni input. È quindi naturale chiedersi quali macchine di Turing si fermano e su quali input. Questo problema è noto con il nome di *halting problem*. La soluzione è contenuta nel seguente risultato.

Teorema 3.2 (Turing)**.** *Sia $(\varphi_i)_i$ la numerazione delle funzioni ricorsive parziali. Non esiste una funzione ricorsiva totale g definita su $(\{0,1\}^*)^2$ tale che per ogni coppia (s,t) valga $g(s,t) = 1$ se $\varphi_s(t)$ è definita (ossia T_s si ferma su t), e $g(s,t) = 0$ altrimenti.*

Dimostrazione. Supponiamo per assurdo che esista una tale funzione g ricorsiva totale. Definiamo allora una nuova funzione ricorsiva parziale ψ definita da $\psi(s) = 1$ se $g(s,s) = 0$, e $\psi(s)$ non sia definita altrimenti. Allora esiste un intero k tale che ψ è la k-esima funzione ricorsiva parziale φ_k nella numerazione introdotta sopra. Ma allora $\varphi_k(k)$ è definita se e solo se $g(k,k) = 0$, contrariamente a come è stata definita la funzione g. □

L'argomento della dimostrazione è noto con il nome di *diagonalizzazione*. Tale argomento è stato usato da Cantor per dimostrare che i numeri reali sono più che numerabili.

Corollario 3.3. *Definiamo gli insiemi* $K_0 := \{(s,t) \ : \ \varphi_s(t) < \infty\}$ *e* $K := \{s \ : \ \varphi_s(s) < \infty\}$. *Allora* K_0 *e* K *sono ricorsivamente enumerabili ma non ricorsivi.*

Il fatto che l'insieme K_0 non è ricorsivo è equivalente al Teorema di Incompletezza di Gödel. Infatti ogni espressione corretta della forma $n \in K_0$ appartiene alla teoria Aritmetica di Peano. Quindi, siccome K_0 non è ricorsivo, esiste un'espressione $n \notin K_0$ che non è dimostrabile.

Concludiamo questa breve introduzione alla teoria della computabilità estendendo la nozione di computabilità alle funzioni a valori reali.

Definizione 3.8. Una funzione $f : \{0,1\}^* \to \mathbb{R}$ è *ricorsiva* se esiste una funzione ricorsiva totale g di due variabili, non descrescente nella seconda variabile, tale che $|f(n) - g(n,k)| < \frac{1}{k}$.

È possibile usare il concetto di funzione reale ricorsiva per definire anche misure ricorsive. Consideriamo l'identificazione dell'intervallo unitario $[0,1]$ con l'insieme $\Omega = \{0,1\}^{\mathbb{N}}$, tramite la rappresentazione binaria. Indichiamo con Γ_s il cilindro di Ω dato da

$$\Gamma_s := \{(s\omega) \ : \ s \in \{0,1\}^n \quad \omega \in \Omega\}$$

Definizione 3.9. Sia μ misura sui boreliani di $[0,1]$. La misura μ è *ricorsiva* se la funzione reale $f(s) := \mu(\Gamma_s)$ è ricorsiva.

3.3 Complessità algoritmica e applicazioni

Nella Sezione 3.1 abbiamo messo in evidenza le differenze tra il concetto statistico di stringa casuale, legato all'entropia di Shannon, e quello che ci aspettiamo intuitivamente. L'aspetto principale di questa differenza era che l'approccio intuitivo si basa soprattutto sulla descrizione di una singola stringa, mentre l'approccio statistico studia le proprietà della sorgente di informazione. Per formalizzare l'approccio intuitivo risulta quindi fondamentale la macchina di Turing, come strumento per la computazione.

Il primo passo è definire un concetto di contenuto di informazione di una stringa, che sia indipendente dalla sorgente di informazione. Il concetto di contenuto di informazione algoritmico è stato introdotto separatamente da Chaitin ([Ch]) e Kolmogorov ([Ko]).

Definizione 3.10. Sia U una macchina di Turing universale. Date due stringhe $s, t \in \{0,1\}^*$, si chiama *Algorithmic Information Content* (o *complessità di Kolmogorov*) di s relativa a t il numero

$$AIC(s|t) := \begin{cases} \min\{|p| \ : \ p \in \{0,1\}^* \text{ t.c. } U(p,t) = s\} \\ +\infty \quad \text{se} \quad \nexists \text{ tale } p \in \{0,1\}^* \end{cases}$$

La stringa p è chiamata il *programma* per calcolare s data t.

Nel caso in cui t sia la stringa vuota λ, indichiamo $AIC(s|\lambda)$ semplicemente con $AIC(s)$.

La scelta di una macchina di Turing universale nella definizione del AIC, indica la possibilità di ottenere la migliore "descrizione" della stringa s. Infatti una macchina di Turing universale permette di scegliere la macchina di Turing più adatta a "descrivere" s. Quello che si paga per questa scelta è una costante, indipendente dalle stringhe s, t. Infatti, data una macchina di Turing T, esiste una costante $c(T,U)$, dipendente solo da T e da U, tale che

$$AIC(s|t) \leq AIC_T(s|t) + c(T,U) \tag{3.2}$$

dove AIC_T indica il contenuto di informazione calcolato usando la macchina di Turing T. La costante $c(T,U)$ corrisponde alla stringa che serve come input alla macchina U per simulare la macchina T.

Osserviamo anche che la definizione di AIC non dipende dalla scelta di una particolare macchina universale. Siano U e U' due macchine universali. Usando la disuguaglianza (3.2) per U e U', e scambiando i ruoli, si ottiene

$$|AIC_U(s|t) - AIC_{U'}(s|t)| \leq c(U,U') \qquad \forall \ s,t \tag{3.3}$$

Scegliere la macchina U o U' comporta quindi una differenza nella lunghezza del programma più corto, che non dipende dalle stringhe s, t. Sono in particolare uguali i comportamenti asintotici di AIC_U e $AIC_{U'}$ al crescere della lunghezza di s. Nel seguito scegliamo allora di usare la macchina universale U descritta nell'esempio 3.2.

Prima di studiare le proprietà del AIC, diamo un'interessante interpretazione della definizione data. Data una stringa s, il $AIC(s)$ può essere diviso in due parti: la prima parte è la stringa codice della macchina di Turing più adatta alla descrizione di s; la seconda parte è il contenuto di informazione di s. Intuitivamente, la prima parte contiene le regolarità di s, mentre la seconda contiene la parte casuale. Se, per esempio, la stringa s è una stringa periodica, allora la prima parte consiste nella descrizione della macchina di

Turing che ripete un numero dato di volte il suo input, mentre la seconda parte è la stringa da ripetere. La prima parte di $AIC(s)$ contiene allora il codice della macchina di Turing e la lunghezza di s, la seconda parte è invece una costante che non dipende dalla lunghezza di s.

La definizione di AIC permette di ottenere facilmente limiti superiori per $AIC(s)$. Basta infatti descrivere una macchina di Turing T che generi s e applicare la disuguaglianza (3.2).

Proposizione 3.4. *Date due stringhe $s, t \in \{0,1\}^*$ vale*

(i) $AIC(s) \leq |s| + cost$

(ii) $AIC(s|t) \leq AIC(s) + cost$

(iii) $AIC(ss) \leq AIC(s) + cost$

Dimostrazione. (i) Basta definire una macchina di Turing T che corrisponda alla funzione identità.

(ii) Costruiamo una macchina T tale che $T(t,p) = s$ se e solo se $U(p) = s$. La tesi segue da $AIC_T(s|t) = AIC(s)$ e dalla disuguaglianza (3.2).

(iii) Costruiamo una macchina T che raddoppi l'output di U. Ossia valga $T(p) = U(p)U(p)$ per ogni input p. Sia m il numero di Gödel della macchina T, allora $U(1^m 0p) = U(p)U(p)$ per ogni input p. Se in particolare p è il programma di s si ha $AIC(ss) \leq 1 + m + AIC(s)$. □

Viene naturale pensare che il AIC sia una funzione sub-additiva, ossia che valga $AIC(s,t) \leq AIC(s) + AIC(t) + cost$ per ogni coppia di stringhe s, t. Se infatti p e q sono programmi per s e t, rispettivamente, il programma per (s,t) sarà la concatenazione pq. Tuttavia, la macchina U deve sapere interpretare la stringa pq come due programmi diversi. Dobbiamo quindi scrivere uno dei due con un codice autodelimitante. Possibili programmi per (s,t) sono quindi $E_1(|p|)pq$ e $E_1(|q|)qp$, che implicano la disuguaglianza

$$AIC(s,t) \leq AIC(s) + AIC(t) + 2\log(\min\{AIC(s), AIC(t)\}) + cost \quad (3.4)$$

Lo stesso accade per $AIC(st)$, diversamente da $AIC(ss)$. La stima di $AIC(s,t)$ si può leggermente migliorare come nell'esercizio 3.2. Notiamo che un modo per avere la sub-additività è calcolare il AIC relativo, infatti

$$AIC(s,t|AIC(s)) \leq AIC(s) + AIC(t) + cost$$

Nella Sezione 1.2, abbiamo visto che il concetto di contenuto di informazione di una stringa era legato alla relazione che c'era tra la stringa data

e l'insieme di tutte le stringhe possibili. Il AIC è invece indipendente dalla conoscenza di questo insieme. Tuttavia la sua conoscenza è un'informazione che può essere utile nella descrizione di una stringa.

Teorema 3.5 (Kolmogorov)**.** *Sia $A \subset \mathbb{N} \times \mathbb{N}$ ricorsivamente enumerabile. Fissato un elemento $y \in \mathbb{N}$ definiamo $Y := \{x \in \mathbb{N} \ : \ (x, y) \in A\}$. Se Y è finito allora per ogni $x \in Y$ vale $AIC(x|y) \leq |d(Y)| + c(A)$, dove $c(A)$ è una costante che dipende solo dall'insieme A.*

Dimostrazione. Sia T la macchina di Turing che enumera l'insieme A senza ripetizioni. Fissato $Y \in \mathbb{N}$, indichiamo con $(x_{i_k}, y_{i_k})_k$ la sotto-successione di elementi di A tali che $x_{i_k} \in Y$, dove $1 \leq k \leq d(Y)$. Modifichiamo allora T, imponendo che sui programmi p che verificano $1 \leq p \leq d(Y)$ valga $T(p) = x_{i_p}$. Allora $AIC(x|y) \leq AIC_T(x) + c(A) \leq |d(Y)| + c(A)$ per ogni $x \in Y$. □

Corollario 3.6. *Sia $A \subset \mathbb{N}$ ricorsivamente enumerabile e indichiamo con $A^{\leq n} := \{x \in A \ : \ |x| \leq n\}$. Sia $d(A^{\leq n}) \leq p(n)$, dove $p(\cdot)$ è una funzione polinomiale. Allora per ogni $x \in A^{\leq n}$ vale $AIC(x) = O(\log n)$.*

Dimostrazione. Sia $A' := \{(x, |x|) \ : \ x \in A\}$. L'insieme $A' \subset \mathbb{N} \times \mathbb{N}$ è ricorsivamente enumerabile. Basta modificare la macchina T che enumera A in modo da far calcolare anche la lunghezza della stringa output. Possiamo allora applicare ad A' il Teorema 3.5, dove, fissato $n \in \mathbb{N}$, abbiamo $Y = \{x \in A \ : \ |x| = n\}$. Quindi $AIC(x|n) \leq |d(Y)| + cost \leq |p(n)| + cost$. Notiamo poi che per ogni stringa $s \in \{0,1\}^*$ vale

$$AIC(s) \leq AIC(s||s|) + 2|\bar{s}| + O(1) \tag{3.5}$$

da cui, per ogni $x \in A^{\leq n}$, otteniamo la tesi. □

Applichiamo il AIC alla stringa data dall'espansione decimale di π. Abbiamo osservato nella Sezione 3.1 che l'espansione decimale di π, nonostante sia intuitivamente casuale, può essere riprodotta usando algoritmi basati sulle uguaglianze tra π e alcune serie numeriche convergenti. Ne segue che, se π_1^n indica la stringa in $\{0, 1, 2, \ldots, 9\}^n$ data dai primi n elementi nell'espansione decimale di π, vale $AIC(\pi_1^n|n) \leq cost$ e $AIC(\pi_1^n) = O(\log n)$.

L'espansione decimale di π è quindi una stringa la cui rappresentazione, attraverso una macchina di Turing, è molto più corta della stringa stessa. Vediamo che è in generale semplice creare stringhe con la stessa proprietà.

Sia per esempio f la funzione ricorsiva definita da $f(1) = 2$ e $f(i) = 2^{f(i-1)}$. Si verifica facilmente che se $s = 1^{f(k)}$, allora $AIC(s) \leq AIC(k) +$

$cost \leq 2\log(k+1) + cost$, mentre $|s| = f(k) >> \log k$. In questo modo è possibile costruire stringhe s per cui il rapporto $\frac{AIC(s)}{|s|}$ è piccolo a piacere. Basta considerare funzioni ricorsive f che crescono velocemente e ripetere il ragionamento di prima. Le stringhe s con questa proprietà si dicono *comprimibili*.

Esistono stringhe non comprimibili? Consideriamo l'insieme $\{0,1\}^n$. Si ha $d(\{0,1\}^n) = 2^n$, mentre

$$d(\{0,1\}^{\leq(n-1)}) = \sum_{i=0}^{n-1} 2^i = 2^n - 1$$

Quindi, se interpretiamo l'insieme $\{0,1\}^{\leq(n-1)}$ come l'insieme dei possibili programmi per le stringhe in $\{0,1\}^n$, ne risulta che esiste almeno una stringa s con $n = |s| \leq AIC(s)$.

Definizione 3.11. Data una costante $c \in \mathbb{N}$, si dice che una stringa $s \in \{0,1\}^*$ è *c-random* se $AIC(s) \geq |s| - c$.

Le stringhe c-random sono quindi le stringhe che non hanno una "struttura", la cui descrizione non può essere molto più corta della stringa stessa. D'altra parte ricordiamo che in generale vale $AIC(s) \leq |s| + cost$. Vediamo adesso che le stringhe c-random sono molte di più delle stringhe comprimibili.

Teorema 3.7 (Incompressibility Theorem)**.** *Data una costante $c \in \mathbb{N}$, per ogni stringa t fissata, ogni insieme A con $d(A) = m < \infty$ è tale che la disuguaglianza*

$$AIC(s|t) \geq \log m - c$$

vale per almeno $m(1 - \frac{1}{2^c}) + 1$ suoi elementi s.

Dimostrazione. Basta contare i programmi di lunghezza minore di $\log m - c$. Si ha

$$\sum_{i=0}^{\log m - c - 1} 2^i = 2^{\log m - c} - 1$$

Quindi almeno $m - 2^{\log m - c} + 1$ elementi di A hanno un programma di lunghezza $\geq \log m - c$. □

Corollario 3.8. *Data una costante $c \in \mathbb{N}$, almeno $2^n - 2^{n-c} + 1$ stringhe di $\{0,1\}^n$ sono c-random.*

In particolare, la densità delle stringhe c-random, per una data costante $c \in \mathbb{N}$, è maggiore o uguale di

$$\lim_{n\to\infty} \frac{2^n - 2^{n-c} + 1}{2^n} = 1 - \frac{1}{2^c} \geq \frac{1}{2}$$

Continuiamo l'osservazione sulla sub-additività del AIC alla luce del Teorema 3.7. Consideriamo l'insieme $A := \{(s,t) : |s| + |t| = n\}$. Si ha $d(A) = (n+1)2^n$, quindi, scegliendo $c = 1$ nell'Incompressibility Theorem, esiste una coppia (s,t) di stringhe tali che $AIC(s,t) \geq n + \log n - 1$. Usando poi la Proposizione 3.4,(i) si ottiene che $AIC(s) + AIC(t) \leq |s| + |t| + cost$, dove la costante è indipendente da s, t. Ne segue che esiste almeno una coppia di stringhe per cui $AIC(s,t) \geq AIC(s) + AIC(t) + log(|s|+|t|) - cost$, con costante indipendente da s, t. Quindi la sub-additività non si può ottenere, e la stima (3.4) con il logaritmo è "abbastanza buona".

Concludiamo la discussione sul concetto di comprimibilità con alcune importanti osservazioni.

Osservazione 3.9. Sia s una stringa c-random, cosa possiamo concludere sulle sottostringhe di s? La prima tentazione sarebbe di pensare che le sottostringhe di s siano c'-random, per una costante c' diversa. Sia $|s| = n$, e scriviamo $s = uvw$, la concatenazione di tre sottostringhe. Allora, se $U(p) = v$, possiamo costruire una macchina di Turing T capace di calcolare s a partire da un input della forma $E_1(|p|)pE_1(|u|)uw$. Quindi esiste una costante $c(T)$ tale che

$$AIC(s) \leq |p| + 2(\log|p| + \log|u| + 1) + n - |v| + c(T)$$

Ricordando che $|p| = AIC(v)$ e s è c-random, possiamo quindi scrivere

$$n - c \leq AIC(s) \leq AIC(v) + n - |v| + 4\log n + O(1)$$

da cui si deduce che $AIC(v) \geq |v| - O(\log n)$. La tentazione iniziale corrisponderebbe a ipotizzare $AIC(v) \geq |v| - O(1)$ per ogni sottostringa v. Ma questo significherebbe escludere tutte le sottostringhe comprimibili per cui $AIC(v) \sim \log|v|$, come ad esempio la stringa $v = (0^k)$. L'assurdo si ottiene ragionando come per il Corollario 3.6. Infatti, escludendo tutte le sottostringhe abbastanza regolari, s diventerebbe elemento di un insieme di stringhe con cardinalità bassa. Quindi il $AIC(s)$ sarebbe piccolo e s non potrebbe essere c-random. Osserviamo che la stessa conclusione si ottiene nella teoria della probabilità: ogni stringa casuale (generata da un processo stocastico) contiene sottostringhe consecutive periodiche di qualsiasi lunghezza. $\Diamond$

Osservazione 3.10. Sia $s \in \{0,1\}^*$ e sia p il suo programma, $U(p) = s$. Possiamo chiederci se p non sia a sua volta comprimibile, rivelando quindi ulteriori strutture in s. Supponiamo che per ogni costante $c \in \mathbb{N}$ esista una stringa s, con programma p tale che $AIC(p) < |p| - c$. Sia q il programma per tale p, ossia $U(q) = p$. Possiamo allora definire la macchina di Turing $V = U \circ U$, ossia $V(q) = U(U(q)) = U(p) = s$. Sia i il numero di Gödel della macchina V, allora $U(1^i 0 q) = s$, quindi

$$AIC(s) \leq |q| + i + 1 \leq |p| - c + i + 1$$

Ripetiamo adesso il ragionamento, scegliendo una costante $c \geq i + 1$. Allora esiste una stringa s tale che $AIC(s) \leq AIC(s) - c'$ con c' costante positiva. Questa contraddizione ci dice quindi che esiste una costante $c \in \mathbb{N}$ tale che per ogni stringa s, se p è il suo programma, vale $AIC(p) \geq |p| - c$. $\Diamond$

Osservazione 3.11. Per studiare il comportamento del AIC sulle stringhe binarie, bisogna analizzare come si comporta sui prefissi. Ossia, se la stringa $s = tu$, allora è vero che necessariamente $AIC(t) \leq AIC(s)$? Che questa conclusione sia falsa si ottiene con semplici esempi. Consideriamo i numeri naturali $n = 2^{100}$ e m numero di 20 cifre tale che $AIC(m) \geq |m| \sim 20$. Allora certamente $n >> m$, ma $AIC(n) \approx \log 2 + \log 100 << AIC(m)$. Ne segue che, date stringhe $s = (1^n)$ e $t = (1^m)$, si ha $s = tu$ per una qualche u, ma $AIC(s) < AIC(t)$.

Un primo tentativo per risolvere questo problema sarebbe considerare $AIC(s||s|)$, ossia eliminare la dipendenza del AIC di s dalla lunghezza della stringa. Tuttavia il problema rimane, come mostra chiaramente il prossimo esempio. Sia $s = (n0^{n-|n|})$, dove usiamo la stringa binaria $\bar{n}$ per rappresentare n. Allora esiste una costante indipendente da s tale che $AIC(s||s|) \leq cost.$ Se però $t = (n)$ e si sceglie n 0-random, si ha $s = tu$ e $AIC(t) \geq |n|$, da cui $AIC(t||t|) \geq AIC(t) - AIC(|n|) \geq |n| - 2\log(\log n)$ (vedi la disuguaglianza (3.5)). $\Diamond$

Osservazione 3.12 (Chaitin). Il Teorema 3.7 indica anche la strada per una dimostrazione alternativa della non-finitezza dei numeri primi. L'Incompressibility Theorem implica infatti che molti numeri naturali sono poco comprimibili, ossia $AIC(n) \approx |n| \sim \log n$. Se i numeri primi fossero finiti con $d = d(\{\text{numeri primi}\})$, usando il teorema fondamentale dell'aritmetica, si potrebbero comprimere tutti i numeri naturali molto di più, ottenendo stime dell'ordine $\log(\log n)$. Il programma di un numero naturale n sarebbe semplicemente la stringa, di lunghezza finita d, che contiene tutte le potenze

$(e_1, \ldots, e_d)$ con cui i primi compaiono in n. Quindi avremmo

$$AIC(n) \leq AIC(e_1, \ldots, e_d) + cost \leq \sum_{i=1}^{d} |E(e_i)| + cost$$

dove $E(e_i)$ è un qualsiasi codice prefisso sui numeri naturali. L'assurdo si ottiene osservando che, per ogni i, vale $|e_i| + cost \sim \log(\log n)$. $\Diamond$

In tutto il capitolo abbiamo usato, anche in maniera tacita, l'equivalenza tra le stringhe di $\{0,1\}^*$ e i numeri naturali, per esempio attraverso il codice D definito nell'equazione (1.3). Guardiamo quindi ora al AIC come una funzione definita su $\mathbb{N}$. Innanzitutto, usando la Proposizione 3.4 e il Corollario 3.8, otteniamo che per ogni $n \in \mathbb{N}$, vale $AIC(n) \leq |n| + cost$, per una costante indipendente da n, e almeno

$$r(n) \approx \sum_{i=1}^{\lfloor \log(n+1) \rfloor} \left(2^i - 2^{i-1} + 1\right)$$

dei primi n numeri naturali sono 1-random.

Teorema 3.13. *Per la funzione $AIC : \mathbb{N} \to \mathbb{N}$, valgono le proprietà:*

(i) $AIC(n)$ è illimitata;

(ii) la funzione $b(n) := \min\{AIC(k) \ : \ k \geq n\}$ è illimitata;

(iii) per ogni funzione ricorsiva parziale $\phi(n)$ che, da un certo n_0 in poi, è monotona e converge a $+\infty$, vale $b(n) < \phi(n)$ tranne che per un numero finito di numeri naturali.

Dimostrazione. (ii) Per ogni $j \in \mathbb{N}$ esiste un naturale n_j tale che, per ogni $n > n_j$, il programma p di n verifica $|p| \geq j$. Questo fatto segue facilmente dalla finitezza dei programmi di lunghezza minore di j. La successione $(n_j)_j$ è non descrescente e $b(n) = j$ se $n \in (n_j, n_{j+1})$.

(i) Segue da (ii), osservando che $AIC(n) \geq b(n)$ per ogni n.

(iii) Ragioniamo per assurdo. Sia $\phi(n)$ funzione ricorsiva parziale non descrescente, illimitata, e tale che $\phi(n) \leq b(n)$ per infiniti n. Sia $A \subset \mathbb{N}$ il dominio di ϕ. Allora A è un insieme infinito e ricorsivamente enumerabile. Quindi, per la Proposizione 3.1, esiste un insieme $B \subset A$ infinito e ricorsivo. Definiamo allora la funzione

$$\psi(n) := \begin{cases} \phi(n) & n \in B \\ \phi(k) & n \notin B, \text{ dove } k = \max\{m \ : \ m \in B, \ m < n\} \end{cases}$$

La funzione ψ è ricorsiva totale, è non decrescente e illimitata, e $\psi(n) \leq b(n)$ infinite volte.

Definiamo poi $M(a) := \max\{n \ : \ AIC(n) \leq a\}$. Allora $M(a) + 1 = \min\{n \ : \ b(n) > a\}$. Quindi per infiniti a vale

$$F(a) := \max\{n \ : \ \psi(n) \leq a + 1\} \geq \min\{n \ : \ b(n) > a\} > M(a)$$

e la funzione $F(a)$ è ricorsiva totale. Quindi $AIC(F(a)) > a$ per infiniti a. Ma

$$AIC(F(a)) \leq AIC_F(F(a)) + cost \leq |a| + cost$$

Quindi esiste una costante c tale che $|a| + c \geq a$ per infiniti a, il che è assurdo. □

La proprietà (i) vale anche per la funzione $AIC(n||n|)$, mentre la stringa s dell'osservazione 3.11 è un contro-esempio per le proprietà (ii) e (iii). In particolare $AIC(n||n|)$ segue lo stesso andamento di $AIC(n)$ tranne che per infiniti naturali, per i quali è limitata da una costante.

Un ragionamento analogo a quello usato per la dimostrazione del punto (iii) permette di dimostrare l'importante teorema seguente.

Teorema 3.14 (Teorema di Non-Computabilità)**.** *La funzione $AIC(n)$ non è ricorsiva parziale. Inoltre, nessuna funzione ricorsiva parziale, definita su un insieme infinito, può coincidere con AIC su tutto il suo dominio.*

Dimostrazione. Sia ϕ una funzione ricorsiva parziale con dominio A infinito, che coincida con AIC su A. Sia $B \subset A$ un insieme infinito ricorsivo. La funzione $\psi(n) := \min\{k \in B \ : \ AIC(k) \geq n\}$ è ricorsiva totale (poiché $AIC(k) = \phi(k)$ su B) ed è illimitata. Inoltre, per definizione, vale $AIC(\psi(n)) \geq n$. D'altra parte

$$AIC(\psi(n)) \leq AIC_\psi(\psi(n)) + cost \leq |n| + cost$$

quindi $n \leq \log n + cost$, il che è assurdo. □

Questo teorema, di fondamentale importanza, implica quindi che il AIC non può essere calcolato tramite un algoritmo su un insieme infinito di stringhe. Si può provare a calcolare il AIC per qualche stringa particolare. Ma per determinare numericamente se una generica stringa sia o non sia c-random bisogna cercare di costruire algoritmi che approssimino il AIC. Un primo risultato in questa direzione è il seguente teorema.

Teorema 3.15. *Esiste una funzione ricorsiva totale $\phi(n,t)$, non-crescente in t, tale che $\lim_{t\to\infty} \phi(n,t) = AIC(n)$ per ogni $n \in \mathbb{N}$.*

Dimostrazione. Definiamo la funzione $\phi(n,t)$ nel seguente modo: dato n, sappiamo che esiste una costante c per cui $AIC(n) \leq |n|+c$; facciamo quindi eseguire il calcolo $U(p)$ per t passi, al variare di p tra tutti i programmi di lunghezza al più $|n| + c$, usando la macchina di Turing universale definita nell'esempio 3.2. Se un qualche programma p dà n come output, poniamo $\phi(n,t)$ uguale alla lunghezza del più breve programma con tale proprietà; altrimenti si pone $\phi(n,t) = |n| + c$. Per definizione $\phi(n,t)$ è ricorsiva totale e non-crescente in t. Inoltre $\lim_{t\to\infty} \phi(n,t)$ esiste ed è uguale a $AIC(n)$, poiché per ogni n esiste un programma p tale che $U(p) = n$, e $|p| = AIC(n)$. □

Nel Capitolo **??** vedremo altri esempi di algoritmi costruiti per approssimare AIC, gli algoritmi di compressione.

Concludiamo questo capitolo con un teorema di confronto tra AIC ed entropia di Shannon. Nel prossimo capitolo questo confronto verrà esteso e generalizzato per i sistemi dinamici.

Teorema 3.16. *Sia $s = t_1 t_2 \dots t_m$ la concatenazione di m stringhe binarie, tutte di lunghezza n, generate in maniera indipendente da una sorgente di informazione. Siano $(p_k)_k$, con $1 \leq k \leq 2^n$, le frequenze di apparizione delle stringhe di $\{0,1\}^n$ ordinate lessicograficamente. Allora vale*

$$AIC(s) \prec m\left(H + 2^{n+1}\frac{|m|}{m}\right) + cost$$

dove $H = -\sum_k p_k \log p_k$ è la funzione entropia di Shannon (vedi equazione (1.1)), e "$\prec$" indica una disuguaglianza asintotica per $m \to \infty$. Quindi, per n fissato, si ha

$$\limsup_{|s|\to\infty} \frac{AIC(s)}{|s|} \leq \frac{H}{n}$$

ossia il AIC definisce un contenuto di informazione medio per ogni simbolo di s minore o uguale di quello definito usando la funzione entropia di Shannon.

Dimostrazione. Usando l'idea del "frequency coding" (vedi Esempio 1.3), se $r_k := mp_k$ indica quante volte la k-esima stringa di $\{0,1\}^n$ appare in s, al variare di k da 1 a 2^n, abbiamo

$$AIC(s) \leq 2\left(\sum_{i=1}^{2^n} |r_i|\right) + \left|\begin{pmatrix} m \\ r_1 r_2 \dots r_{2^n} \end{pmatrix}\right| + cost$$

La tesi si ottiene usando la formula di Stirling per approssimare il coefficiente multinomiale, e usando $r_i \leq m$ per ogni i. □

Nello spirito della teoria dei codici, mostriamo adesso un esempio in cui è possibile ottenere anche la relazione opposta. Il caso generale è affrontato nel Teorema di Brudno (Teorema 4.1).

Esempio 3.3. Consideriamo l'insieme $\{0,1\}^n$ con la distribuzione di probabilità uniforme $P(s) = 2^{-n}$. In questo caso la funzione entropia $H(P) = n$. Interpretiamo i programmi delle stringhe di $\{0,1\}^n$ come stringhe codice. Allora la lunghezza media delle stringhe codice, $L_{AIC} = \sum_{i=1}^{2^n} 2^{-n} AIC(s_i) \leq n + cost$, rappresenta anche la "complessità media" delle stringhe binarie. Notiamo che il codice così descritto non è necessariamente prefisso. Tuttavia, usando il Corollario 3.8, abbiamo che, per ogni costante $c \in \mathbb{N}$, vale

$$\frac{n}{n + cost} \leq \frac{H(P)}{L_{AIC}} \leq \frac{n}{(n-c)(1-2^{-c})}$$

da cui, prendendo $c = \log n$, si ottiene

$$\lim_{n\to\infty} \frac{H(P)}{L_{AIC}} = 1 \qquad \triangle$$

Esercizi

3.1. Dimostrare che data una stringa $s \in \{0,1\}^*$, se s^R indica la stringa inversa (ossia la stringa s letta in ordine inverso), vale $|AIC(s) - AIC(s^R)| \leq cost$.

3.2. Dimostrare che date due stringhe $s, t \in \{0,1\}^*$ vale

$$AIC(s,t) \leq AIC(s) + 2|\overline{AIC(s)}| + AIC(t|s) + O(1)$$

3.3. Dimostrare la disuguaglianza (3.5).

3.4. Usare il Corollario 3.6 per dimostrare che il AIC di una stringa periodica s è $O(\log |s|)$, e che $AIC(s||s|) \leq cost$.

Capitolo 4

La complessità nei sistemi dinamici

Il concetto di Algorithmic Information Content (Definizione 3.10) introduce una nozione di contenuto di informazione per una stringa infinita che non dipende dal contesto in cui troviamo la stringa. Usando il AIC abbiamo poi introdotto la nozione di casualità per una stringa finita (Definizione 3.11). L'estensione di questi concetti alla dinamica simbolica, e quindi ai sistemi dinamici, viene trattata in questo capitolo. La prima sezione è in larga parte tratta da [Br].

4.1 Il Teorema di Brudno

Nella Sezione 3.3, abbiamo rivolto la nostra attenzione allo studio del AIC di una stringa finita come funzione della lunghezza della stringa. In particolare, nel Teorema 3.16, abbiamo studiato il comportamento asintotico di $\frac{AIC(s)}{|s|}$ rispetto alla funzione entropia di Shannon. In questa sezione, estendiamo questo risultato ai sistemi dinamici simbolici.

Sia $\mathcal{A} = \{1, \ldots, N\}$ un alfabeto finito, e $\Omega = \mathcal{A}^{\mathbb{N}_0}$ sia l'insieme delle stringhe $\omega = (\omega_i)_{i \geq 0}$ infinite con simboli dall'alfabeto $\mathcal{A}$. Indichiamo con ω^n la sotto-stringa finita $(\omega_0 \ldots \omega_{n-1})$.

Definizione 4.1. Data una stringa $\omega \in \Omega$, chiamiamo *complessità* di ω il limite

$$K(\omega) := \limsup_{n \to \infty} \frac{AIC(\omega^n)}{n}$$

La complessità di una stringa infinita si può quindi interpretare come il contenuto di informazione medio di ogni simbolo della stringa, indipen-

dentemente dal contesto. Nel teorema seguente, vedremo che coincide con l'idea di contenuto di informazione introdotto dall'entropia di Shannon.

Sia (Ω, τ) un sistema dinamico simbolico, dove τ è lo shift definito nell'equazione (2.7). Per una misura di probabilità ν su Ω, che sia τ-invariante, indichiamo con $h_\nu(\tau)$ l'entropia metrica (vedi Definizione 2.17).

Teorema 4.1 (Brudno)**.** *Se ν è misura di probabilità su Ω, τ-invariante ed ergodica, allora $K(\omega) = h_\nu(\tau)$ per ν-q.o. $\omega \in \Omega$.*

Dimostrazione. Per calcolare $h_\nu(\tau)$ usiamo il Teorema 2.15 applicato alla partizione generante $Z = \{C(j, 0, 1) \ : \ j = 1, \dots, N\}$, dove ricordiamo che $C(j, 0, 1) = \{\omega \ : \ \omega_0 = j\}$.

Osserviamo innanzitutto che la complessità di una stringa finita è invariante per l'azione dello shift, ossia per ogni $\omega \in \Omega$

$$K(\tau^k(\omega)) = K(\omega) \qquad \forall\ k \in \mathbb{N} \tag{4.1}$$

Fissato $k \in \mathbb{N}$, indichiamo con $\tilde{\omega} = \tau^k(\omega)$. Sia U la macchina di Turing universale scelta per calcolare il AIC. Costruiamo una macchina di Turing T che accetta input della forma $p = \bar{p}_1\bar{p}_2p_3$, dove $p_i \in \{0,1\}^*$ e $\bar{p}_i$ indica una rappresentazione auto-delimitante di p_i (ad esempio possiamo scegliere $\bar{p}_i = E_1(p_i)$). Il funzionamento della macchina T avvenga secondo queste regole: prima di tutto viene calcolata $U(p_3)$; vengono poi cancellati p_2 simboli di $U(p_3)$; in testa a quello che resta, viene scritta la stringa $U(p_1)$. Così se p_3 è un programma per ω^{n+k}, e se $p_1 = 0$ e $p_2 = k$, si ha $T(\bar{p}_1\bar{p}_2p_3) = \tilde{\omega}^n$. Applicando l'equazione (3.2), si ottiene

$$AIC(\tilde{\omega}^n) \leq AIC(\omega^{n+k}) + 2\log k + cost$$

quindi, per la complessità, si ottiene

$$K(\tilde{\omega}) \leq \limsup_{n\to\infty} \ \frac{\big[AIC(\omega^{n+k}) + 2\log k + cost\big]}{n+k} \ \frac{n+k}{n} = K(\omega)$$

In maniera del tutto analoga, se p_3 è un programma per $\tilde{\omega}^{n-k}$, se $p_2 = 0$ e se $U(p_1) = \omega^k$, si ottiene $T(\bar{p}_1\bar{p}_2p_3) = \omega^n$. Ragionando come prima si ottiene $K(\omega) \leq K(\tilde{\omega})$. Abbiamo quindi dimostrato l'equazione (4.1).

Definiamo poi per la complessità gli insiemi di livello, e di sopra e sotto-livello. Poniamo $A_t := \{\omega \ : \ K(\omega) = t\}$, $A_t^+ := \{\omega \ : \ K(\omega) > t\}$ e $A_t^- := \{\omega \ : \ K(\omega) < t\}$. L'equazione (4.1) implica che questi insiemi sono τ-invarianti. Dimostriamo che sono anche insiemi misurabili. Osserviamo

che si può scrivere

$$A_t^- = \bigcup_{k\geq 1} \bigcup_{N\geq 1} \bigcap_{n>N} \left\{\omega \ : \ AIC(\omega^n) < \left(t - \frac{1}{k}\right) n\right\}$$

La misurabilità di A_t^- segue allora dalla misurabilità degli insiemi

$$\left\{AIC(\omega^n) < \left(t - \frac{1}{k}\right) n\right\}$$

che sono unione finita di cilindri. Un ragionamento analogo vale per A_t^+, e la misurabilità di A_t segue da quella di A_t^- e A_t^+.

Lemma 4.2. *Se ν è misura di probabilità, τ-invariante ed ergodica, allora $K(\omega) \geq h_\nu(\tau)$ per ν-q.o. $\omega \in \Omega$.*

Dimostrazione. Se $h_\nu(\tau) = 0$ la tesi è banalmente vera. Supponiamo allora $h_\nu(\tau) > 0$ e ragioniamo per assurdo. Indichiamo con Q l'insieme $Q := \{\omega \ : \ K(\omega) < h_\nu(\tau)\}$, e supponiamo $\nu(Q) > 0$. Per l'equazione (4.1) Q è invariante, ed è misurabile perché è della forma A_t^- con $t = h_\nu(\tau)$. Allora, per l'ergodicità di ν, si ha $\nu(Q) = 1$ (vedi Definizione 2.22). Ponendo $Q_r := \left\{\omega \ : \ K(\omega) < h_\nu(\tau) - \frac{1}{r}\right\}$, si ha $Q = \cup_{r\in\mathbb{N}} Q_r$, con Q_r insiemi invarianti e misurabili (per gli stessi motivi di Q), e $Q_1 \subseteq Q_2 \subseteq \cdots \subseteq Q$. Nuovamente per l'ergodicità di ν, esiste un $R \in \mathbb{N}$ tale che $\nu(Q_R) = 1$. Poniamo poi $Q_R = \cup_{k\geq 1} Q_{R,k}$ dove

$$Q_{R,k} := \left\{\omega \ : \ AIC(\omega^l) < \left(h_\nu(\tau) - \frac{1}{R}\right) l \qquad \forall\ l \geq k\right\}$$

e vale $Q_{R,1} \subseteq Q_{R,2} \subseteq \cdots \subseteq Q_R$. Quindi $\nu(Q_{R,k}) \nearrow 1$, e per ogni $\delta > 0$ esiste un $\bar{k}$ tale che $\nu(Q_{R,k}) > 1 - \delta$ per ogni $k > \bar{k}$.

Applichiamo alla partizione Z dei cilindri di lunghezza uno la Proprietà di Equipartizione (Corollario 2.22). Fissato un $\varepsilon < \min\left\{\frac{1}{R}, 1 - \delta\right\}$, esiste $n_0(\varepsilon)$ tale che $Z_n = T_n \cup N_n$, con $\nu\left(\cup_{C\in N_n} C\right) < \varepsilon$, e se ω^n è tale che $C(\omega^n, 0, n) \in T_n$ vale

$$2^{-n(h_\nu(\tau)+\varepsilon)} < \nu(C(\omega^n, 0, n)) < 2^{-n(h_\nu(\tau)-\varepsilon)} \qquad \forall\ n \geq n_0(\varepsilon) \tag{4.2}$$

Poniamo $Q_{R,k}^T := Q_{R,k} \cap T_k$ e $Q_{R,k}^N := Q_{R,k} \cap N_k$. Allora per ogni $k > \max\left\{\bar{k}, n_0(\varepsilon)\right\}$ si ha $\nu(Q_{R,k}^N) \leq \nu(N_k) < \varepsilon$, e $\nu(Q_{R,k}^T) > 1 - \delta - \varepsilon > 0$.

D'altra parte, possiamo stimare $\nu(Q_{R,k}^T)$ con

$$\nu(Q_{R,k}^T) \leq d\left(\left\{\omega^k \ : \ \omega \in Q_{R,k}^T\right\}\right)\ \nu(C(\omega^k, 0, k)) \tag{4.3}$$

Inoltre, se $\omega \in Q^T_{R,k}$, si ha $AIC(\omega^k) < k\left(h_\nu(\tau) - \frac{1}{R}\right)$, quindi

$$d\left(\left\{\omega^k \ : \ \omega \in Q^T_{R,k}\right\}\right) \leq 2^{k\left(h_\nu(\tau)-\frac{1}{R}\right)} \tag{4.4}$$

Mettendo insieme le disuguaglianze (4.2), (4.3) e (4.4) otteniamo

$$\nu(Q^T_{R,k}) \leq 2^{k\left(h_\nu(\tau)-\frac{1}{R}\right)}\ 2^{-k(h_\nu(\tau)-\varepsilon)} = 2^{k\left(\varepsilon-\frac{1}{R}\right)}$$

che implica $\lim_{k\to\infty} \nu(Q^T_{R,k}) = 0$, in contraddizione al fatto che $\nu(Q^T_{R,k}) > 1-\delta-\varepsilon > 0$ per ogni $k > \max\left\{\bar{k}, n_0(\varepsilon)\right\}$. □

Lemma 4.3. *Sia $P = \{P_1, \ldots, P_M\}$ partizione finita misurabile di Ω. Fissati $r, k \in \mathbb{N}$ con $r < k$, poniamo per $m = 1, \ldots, M$*

$$p^{r,k}_m(\omega, n) := \frac{1}{\left\lfloor \frac{n}{k} \right\rfloor} \sum_{i=0}^{\left\lfloor \frac{n-r}{k} \right\rfloor} \chi_{P_m}(\tau^{ik+r}(\omega))$$

dove χ_{P_m} è la funzione indicatrice dell'insieme P_m. Se ν è misura di probabilità τ-invariante ed ergodica, allora per ν-q.o. $\omega \in \Omega$ valgono le relazioni:

$$(a) \qquad \exists\ p^{r,k}_m(\omega) := \lim_{n\to\infty} p^{r,k}_m(\omega, n) \qquad \forall\ m = 1, \ldots, M$$

$$(b) \qquad \exists\ r < k \quad t.c. \quad -\sum_{m=1}^{M} p^{r,k}_m(\omega) \log p^{r,k}_m(\omega) \leq H_\nu(P)$$

dove $H_\nu(P)$ è l'entropia metrica della partizione P (vedi Definizione 2.14).

Dimostrazione. Innanzitutto osserviamo che, essendo ν misura di probabilità τ-invariante ed ergodica, ν è invariante anche per τ^k, per ogni $k > 1$. Quindi possiamo applicare il Teorema di Birkhoff 2.20 alle funzioni indicatrici degli insiemi della partizione P, ottenendo (a).

Inoltre, per ν-q.o. ω, vale $p^{0,1}_m(\omega) = \nu(P_m)$, per ogni $m = 1, \ldots, M$, grazie all'ergodicità di ν rispetto a τ, e si verifica che

$$p^{0,1}_m(\omega) = \frac{1}{k} \sum_{r=0}^{k-1} p^{r,k}_m(\omega)$$

Basta verificarlo per $p^{r,k}_m(\omega, n)$ e $p^{0,1}_m(\omega, n)$, e passare al limite. Usando allora la concavità della funzione $f(x) = -x \log x$, si ottiene

$$H_\nu(P) \geq \frac{1}{k} \sum_{r=0}^{k-1} \left(-\sum_{m=1}^{M} p^{r,k}_m(\omega) \log p^{r,k}_m(\omega) \right)$$

Quindi la tesi (b) si ottiene per assurdo. □

Lemma 4.4. *Se ν è misura di probabilità, τ-invariante ed ergodica, allora $K(\omega) \leq h_\nu(\tau)$ per ν-q.o. $\omega \in \Omega$.*

Dimostrazione. Applichiamo l'idea legata al frequency coding dell'Esempio 1.3. Fissato un intero $k > 0$, per ogni stringa $\omega \in \Omega$ e per ogni $n \in \mathbb{N}$, possiamo scrivere

$$\omega^n = \omega_0^r \ \omega_{i_1}^k \ \omega_{i_2}^k \dots \ \omega_{i_m}^k$$

dove $n = mk + r$ con $r < k$, e $\omega_{i_j}^k$ sono parole in $\mathcal{A}^k$. Supponiamo che $\mathcal{A}^k$ sia ordinato lessicograficamente, e sia $M = d(\mathcal{A}^k)$. Se indichiamo con σ_h, con $h = 1, \dots, M$ le parole di $\mathcal{A}^k$, possiamo contare le apparizioni in ω^n di ciascuna parola di $\mathcal{A}^k$, definendo per ogni $h = 1, \dots, M$

$$s_h(\omega^n) = d\left(\left\{j = 1, \dots, m \ : \ \omega_{i_j}^k = \sigma_h\right\}\right)$$

Le stringhe che hanno le stesse frequenze di apparizione delle parole dell'insieme $\mathcal{A}^k$ sono

$$D_{m,s_1,\dots,s_M} := \left(\begin{array}{c} m \\ s_1 \ s_2 \dots s_M \end{array} \right) = \frac{m!}{s_1!\ s_2! \dots s_M!}$$

e sia $N(\omega^n)$ il numero d'ordine di ω^n in questo insieme. Quindi una possibile codifica di ω^n è la stringa

$$S := \left(k,\ m,\ s_1,\ s_2, \dots,\ s_M,\ r,\ \omega_0^r,\ N(\omega^n)\right)$$

trasmessa usando il codice prefisso E_1 (vedi equazione (1.5)) per codificare $k, m, s_1, \dots, s_M, r$, concatenati con $E_1(|q_0^r|)q_0^r$, dove q_0^r è il programma di ω_0^r sulla macchina di Turing universale U, e aggiungendo infine $\overline{N(\omega^n)}$ (vedi equazione (1.3)). Allora $AIC(\omega^n) \leq |E(S)|$, dove E indica il codice prefisso appena descritto per S. Ne segue che

$$AIC(\omega^n) \leq 2 \log\left(kms_1 \dots s_M r AIC(\omega_0^r)\right) + AIC(\omega_0^r) + |D_{m,s_1,\dots,s_M}| + O(1)$$

Osserviamo che $m \leq n$, e quindi $\log s_h \leq \log n$ per ogni $h = 1, \dots, M$. Inoltre esiste una costante R, che dipende solo da k, tale che $AIC(\omega_0^r) \leq R$ per ogni $r < k$. Quindi

$$AIC(\omega^n) \leq |D_{m,s_1,\dots,s_M}| + O(\log n)$$

e, come abbiamo visto nell'Esempio 1.3, vale

$$|D_{m,s_1,\dots,s_M}| \sim m \left(-\sum_{h=1}^{M} p_h^{r,k}(\omega, n) \log p_h^{r,k}(\omega, n) \right)$$

dove $p_h^{r,k}(\omega, n) := \frac{s_h(\omega^n)}{m}$, e la notazione è quella del Lemma 4.3 rispetto alla partizione $P = Z_k$. Il punto (b) del Lemma 4.3 implica che, scegliendo opportunamente r, si ottiene

$$AIC(\omega^n) \leq mH_\nu(Z_k) + o(n)$$

per ν-quasi ogni $\omega \in \Omega$. Quindi

$$K(\omega) = \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n} \leq \limsup_{n\to\infty} \frac{mH_\nu(Z_k)}{n} = \frac{H_\nu(Z_k)}{k} \tag{4.5}$$

lungo la sotto-successione $n = km + r$ con r fissato. Quindi la prima uguaglianza non coincide con la definizione di complessità di una stringa, ma ne discende facilmente per la limitatezza di $r < k$.

La tesi segue allora dalla Definizione 2.16 di entropia metrica, applicando la disuguaglianza (4.5) al variare di $k \to \infty$. □

Mettendo insieme il Lemma 4.2 e il Lemma 4.4, otteniamo la tesi del Teorema. □

Affrontiamo adesso il problema per un sistema dinamico (X, T, μ) generale. Innanzitutto vediamo come definire la complessità dei punti $x \in X$. Da quanto visto nel Capitolo 2, lo strumento per passare da un sistema dinamico alla dinamica simbolica è una rappresentazione simbolica.

Definizione 4.2. Dato un sistema dinamico (X, T), sia Z una partizione finita e misurabile di X, e sia φ_Z la rappresentazione simbolica associata. Poniamo $AIC(x, n, Z) := AIC(\varphi_Z(x)^n)$. La *complessità* $K(x, T, Z)$ di un punto $x \in X$, relativa alla partizione Z, è definita tramite

$$K(x, T, Z) := K(\varphi_Z(x)) = \limsup_{n\to\infty} \frac{AIC(x, n, Z)}{n}$$

Teorema 4.5. *Sia (X, T) sistema dinamico e μ misura di probabilità T-invariante ed ergodica. Data una partizione Z finita e misurabile, per μ-quasi ogni $x \in X$ vale $K(x, T, Z) = h_\mu(T, Z)$.*

Dimostrazione. Sia ν_Z la misura $\varphi_Z^*\mu$ indotta su $\Omega_Z := \varphi_Z(X) \subset \mathcal{A}^{\mathbb{N}_0}$, dove $\mathcal{A}$ è l'alfabeto indotto da Z. Allora ν_Z è misura di probabilità, invariante per il sistema (Ω_Z, τ) ed ergodica. Inoltre $h_{\nu_Z}(\tau) = h_\mu(T, Z)$. Applicando il Teorema di Brudno 4.1 al sistema (Ω_Z, τ, ν_Z), si ottiene che, se $W := \{\omega \in \Omega_Z \ : \ K(\omega) = h_{\nu_Z}(\tau)\}$, si ha $\nu_Z(W) = 1$. Quindi, poiché $Y := \{x \in X \ : \ K(x, T, Z) = h_\mu(T, Z)\} = \varphi_Z^{-1}(W)$, per la definizione di ν_Z, si ottiene $\mu(Y) = 1$. □

Come già osservato nel Capitolo 2, l'utilizzo della rappresentazione simbolica mette in evidenza le proprietà statistiche di un sistema dinamico, relativamente alla partizione scelta. Per avere informazioni su proprietà assolute del sistema, bisognerebbe considerare l'estremo superiore al variare di tutte le partizioni finite e misurabili. Nel Capitolo 2, abbiamo visto che questo ragionamento funziona bene per l'entropia metrica (vedi Teorema 2.15), mentre presenta dei problemi per l'entropia topologica. Mostriamo adesso che per la complessità $K(x, T, Z)$ non si può considerare l'estremo superiore al variare di Z.

Proposizione 4.6 (Brudno [Br])**.** *Sia $x \in X$ punto non periodico per T, allora per ogni $N \in \mathbb{N}$ esiste una partizione P finita e misurabile, tale che $K(x, T, P) = \log N$.*

Dimostrazione. Scelto $N \in \mathbb{N}$, sia $\mathcal{A} = \{1, \dots, N\}$, e $\bar{\omega} \in \mathcal{A}^{\mathbb{N}_0}$ tale che $K(\bar{\omega}) = \log N$. Una tale scelta è sempre possibile, grazie al Teorema 4.1 con la misura ν definita in (2.13) e simboli di $\mathcal{A}$ equiprobabili. Fissato $x \in X$, poniamo per $i = 1, \dots, N$

$$P_i := \left\{ y \in X \ : \ \exists\, k \in \mathbb{N} \ \text{ t.c. } T^k(x) = y \ \text{ e } \ \bar{\omega}_k = i \right\}$$

e $P_0 = X \setminus \cup_{i=1}^{N} P_i$. La partizione $P := \{P_0, P_1, \dots, P_N\}$ è finita e misurabile, e $\varphi_P(x) = \bar{\omega}$. Quindi $K(x, T, P) = \log N$. □

Come già nel caso dell'entropia topologica, si potrebbe aggirare il problema considerando i ricoprimenti aperti di X, e definire un concetto di complessità di un punto adatto (vedi [Br]). Tuttavia, noi ci restringiamo a considerare il calcolo della complessità dei punti relativamente a una partizione fissata.

Nel caso di una misura di probabilità ν su Ω, che sia τ-invariante e non ergodica, ritroviamo l'entropia metrica $h_\nu(\tau)$ facendo la media della complessità delle stringhe.

Teorema 4.7. *Data la dinamica simbolica (Ω, τ, ν), con ν di probabilità e τ-invariante, vale*

$$\int_\Omega K(\omega)\ d\nu = h_\nu(\tau)$$

Dimostrazione. La misurabilità della funzione $\omega \to K(\omega)$ segue dalla misurabilità degli insiemi A_t definiti nella dimostrazione del Teorema 4.1. La tesi segue poi dall'applicazione di un metodo di teoria ergodica noto come *decomposizione ergodica*. Si ottiene l'esistenza di una famiglia $(\Omega_j, \nu_j)_{j \in J}$ di

sottospazi di Ω, invarianti per l'azione dello shift τ, e tali che le misure ν_j, con supporto in Ω_j, siano misure di probabilità τ-invarianti ed ergodiche. Inoltre, lo spazio J è uno spazio di Lebesgue con misura di probabilità P. Possiamo quindi scrivere

$$\int_\Omega K(\omega)\ d\nu = \int_J \left(\int_{\Omega_j} K(\omega)\ d\nu_j \right)\ dP = \int_J h_{\nu_j}(\tau)\ dP = h_\nu(\tau)$$

dove la prima e l'ultima uguaglianza derivano dalla decomposizione ergodica, mentre la seconda segue dal Teorema 4.1. □

Considerando una partizione Z per un sistema dinamico (X, T, μ), con μ misura di probabilità T-invariante, si ottiene, ragionando come prima,

$$\int_X K(x, T, Z)\ d\mu = h_\mu(T, Z)$$

L'applicazione del Teorema 4.5 è di immediata interpretazione per i sistemi caotici (vedi Definizione 2.21). Per un sistema dinamico (X, T), date μ misura fisica con entropia metrica $h_\mu(T)$ positiva, e Z partizione generante, o per cui valga $h_\mu(T, Z) > 0$, il Teorema 4.5 implica che per μ-quasi ogni $x \in X$ vale

$$AIC(x, n, Z) \sim n\ h_\mu(T, Z) \tag{4.6}$$

Questa conclusione vale quindi per la mappa del panettiere, per la mappa T_4 della famiglia quadratica, e per le mappe della famiglia di Pomeau-Manneville per $1 < z < 2$, con le partizioni e le misure indicate negli Esempi 2.16, 2.17 e 2.18. Notiamo che per tutti questi sistemi, l'equazione (4.6) è verificata per quasi ogni punto rispetto alla misura di Lebesgue sull'intervallo $[0, 1]$.

4.2 Il caos debole

Nella Definizione 2.21, abbiamo definito debolmente caotici i sistemi dinamici che hanno entropia metrica nulla rispetto a ogni misura fisica. Esempi di questi sistemi sono le rotazioni del cerchio, la mappa T_{λ_∞} della famiglia quadratica e le mappe della famiglia di Pomeau-Manneville per $z \geq 2$. L'applicazione del Teorema 4.5 a un sistema (X, T) debolmente caotico implica che, data una misura fisica μ ergodica, vale

$$AIC(x, n, Z) = o(n) \tag{4.7}$$

per μ-q.o. $x \in X$ e per ogni partizione Z finita e misurabile. Questo risultato non può essere soddisfacente per due ragioni. Innanzitutto, osserviamo che non consente di classificare i sistemi debolmente caotici, fornendoci solo una stima molto rozza del comportamento asintotico di $AIC(x, n, Z)$. Questo problema si verifica anche per l'applicazione del Teorema di Shannon-McMillan-Breiman (Teorema 2.21), che nel caso debolmente caotico implica solo che $\log \frac{1}{\mu(Z_n(x))} = o(n)$, per μ-q.o. $x \in X$ e per ogni partizione Z.

Ma un secondo problema riguarda il supporto delle misure fisiche. Per le mappe T_z di Pomeau-Manneville con $z \geq 2$, per esempio, l'unica misura fisica è singolare rispetto alla misura di Lebesgue. Quindi il Teorema 4.5 è, a priori, applicabile solo a un insieme di punti di misura di Lebesgue nulla. Per ottenere un risultato che valga per un insieme di punti di misura di Lebesgue positiva, dobbiamo allora considerare il caso delle misure invarianti infinite.

Teorema 4.8. *Dato un sistema dinamico (X, T) e una misura μ su X, che sia σ-finita, T-invariante ed ergodica*[1]*, per ogni partizione Z finita e misurabile vale $K(x, T, Z) = 0$ per μ-q.o. $x \in X$.*

Dimostrazione. Basta adattare i Lemmi 4.3 e 4.4 della dimostrazione del Teorema di Brudno, al caso di un sistema dinamico con misura infinita. In particolare, bisogna usare la versione del Teorema di Birkhoff (Teorema 2.20) per le misure infinite. □

Anche per i sistemi dinamici con misura infinita vale dunque l'equazione (4.7). Per ottenere quindi una classificazione dei sistemi debolmente caotici o con misura infinita, sempre rispetto al comportamento del AIC delle loro orbite simboliche, bisogna misurare i comportamenti asintotici sub-lineari. Iniziamo trattando il caso della dinamica simbolica.

Sia $\mathcal{A}$ alfabeto finito, e $\Omega = \mathcal{A}^{\mathbb{N}_0}$ lo spazio delle stringhe infinite con simboli di $\mathcal{A}$. Sia ν una misura di probabilità su Ω, non necessariamente invariante rispetto all'azione dello shift τ. Definiamo gli *indici di caos globali* e *locali*.

Definizione 4.3. L'*indice globale superiore di caos* $\overline{q}(\Omega, \nu)$, relativo alla misura ν, è dato da

$$\overline{q}(\Omega, \nu) = \inf \left\{ q \in (0, 1) \ : \ \limsup_{n \to \infty} \int_\Omega \frac{AIC(\omega^n)}{n^q} \, d\nu = 0 \right\}$$

In maniera analoga è definito l'*indice globale inferiore di caos* $\underline{q}(\Omega, \nu)$, sostituendo il limsup con il liminf.

[1] Sia sempre sottintesa l'ipotesi di conservatività (vedi Sezione 2.5)

Definizione 4.4. L'*indice locale superiore di caos* $\overline{q}(\omega)$ è dato da

$$\overline{q}(\omega) = \inf \left\{ q \in (0,1) \ : \ \limsup_{n\to\infty} \ \frac{AIC(\omega^n)}{n^q} = 0 \right\}$$

In maniera analoga è definito l'*indice locale inferiore di caos* $\underline{q}(\omega)$, sostituendo il limsup con il liminf.

Studiamo ora alcune proprietà degli indici di caos introdotti.

Proposizione 4.9. *Gli indici locali di caos sono una funzione misurabile su* Ω*.*

Dimostrazione. Dimostriamo la tesi per l'indice locale superiore. In maniera del tutto analoga si ragiona per l'indice inferiore.

La tesi si ottiene mostrando che gli insiemi $Q_\gamma = \{\omega \ : \ \overline{q}(\omega) < \gamma\}$ sono misurabili per ogni $\gamma \in (0,1)$. Fissato γ, scriviamo $Q_\gamma = \cup_{\bar{k}\in\mathbb{N}} \ A_{\bar{k},\gamma}$, dove

$$A_{\bar{k},\gamma} = \left\{ \omega \ : \ \forall \ k > \bar{k} \ \ \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^{\gamma - \frac{1}{k}}} = 0 \right\}$$

Basta quindi mostrare che sono misurabili gli insiemi $A_{\bar{k},\gamma}$. Scrivendo $A_{\bar{k},\gamma} = \cap_{k>\bar{k}} \ C_k$, dove

$$C_k = \left\{ \omega \ : \ \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^{\gamma - \frac{1}{k}}} = 0 \right\}$$

la tesi si ottiene dalla relazione

$$C_k = \bigcap_{h\in\mathbb{N}} \ \bigcup_{N\in\mathbb{N}} \ \bigcap_{n>N} \ \left\{ \omega \ : \ AIC(\omega^n) < n^{\gamma - \frac{1}{k}} \ \frac{1}{h} \right\}$$

che mostra la misurabilità degli insiemi C_k, e quindi degli insiemi $A_{\bar{k},\gamma}$. □

Proposizione 4.10. *Gli indici locali di caos sono invarianti per l'azione dello shift su* Ω*.*

Dimostrazione. Dalla definizione di Algorithmic Information Content (vedi Definizione 3.10), per ogni $\omega \in \Omega$ vale

$$AIC(\omega^n) \leq AIC((\tau\omega)^n) + c(\mathcal{A}) + cost \tag{4.8}$$

dove la costante $c(\mathcal{A})$ dipende solo dall'alfabeto. Infatti, dalla stringa $(\tau\omega)^n$, possiamo ricostruire la stringa ω^n aggiungendo l'informazione sul simbolo ω^0, e usando una macchina di Turing apposita (che comporta l'ulteriore costante, indipendente dalla stringa e dall'alfabeto). Seguendo un ragionamento analogo, si dimostra che

$$AIC((\tau\omega)^n) \leq AIC(\omega^n) + c(\mathcal{A}) + cost \tag{4.9}$$

Mettendo insieme le disuguaglianze (4.8) e (4.9), otteniamo che per ogni $q \in (0,1)$ vale

$$\frac{AIC(\omega^n)}{n^q} \leq \frac{AIC((\tau\omega)^n) + c(\mathcal{A}) + cost}{n^q} \leq \frac{AIC(\omega^n) + 2c(\mathcal{A}) + cost}{n^q} \tag{4.10}$$

e la tesi segue dalla definizione degli indici locali di caos. □

Come conseguenza delle Proposizioni 4.9 e 4.10, segue

Teorema 4.11. *Sia μ misura sullo spazio Ω, invariante per lo shift ed ergodica. Allora esistono $\overline{q}, \underline{q} \in (0,1)$ tali che per μ-q.o. $\omega \in \Omega$ si ha $\overline{q}(\omega) = \overline{q}$ e $\underline{q}(\omega) = \underline{q}$.*

Teorema 4.12. *Data la dinamica simbolica (Ω, τ, μ), con μ τ-invariante ed ergodica, e data una misura di probabilità $\nu \sim \mu$, vale*

$$\underline{q} \leq \underline{q}(\Omega, \nu) \leq \overline{q}(\Omega, \nu)$$

dove $\underline{q}$ è l'indice locale inferiore costante.

Dimostrazione. La tesi segue dall'applicazione del Lemma di Fatou, da cui otteniamo

$$\int_X \liminf_{n\to\infty} \frac{AIC(\omega^n)}{n^q}\, d\nu \leq \liminf_{n\to\infty} \int_X \frac{AIC(\omega^n)}{n^q}\, d\nu$$

e dall'ovvia disuguaglianza

$$\liminf_{n\to\infty} \int_X \frac{AIC(\omega^n)}{n^q}\, d\nu \leq \limsup_{n\to\infty} \int_X \frac{AIC(\omega^n)}{n^q}\, d\nu \qquad \square$$

Per definire gli indici di caos per un sistema dinamico generico, usiamo di nuovo la rappresentazione simbolica. Dato il sistema (X,T) con $X \subset \mathbb{R}^k$, per un certo $k \geq 1$, sia Z una partizione finita e misurabile di X e μ una misura di probabilità su X equivalente alla misura di Lebesgue m_k, non necessariamente T-invariante. La rappresentazione simbolica φ_Z associa, al sistema (X,T) con misura μ, un sottospazio $\Omega_Z \subset \mathcal{A}^{\mathbb{N}_0}$ (con $\mathcal{A}$ alfabeto associato a Z) e una misura di probabilità ν_Z.

Definizione 4.5. Gli *indici globali di caos* $\overline{q}(T, \mu, Z)$ e $\underline{q}(T, \mu, Z)$, relativi alla partizione Z e alla misura μ, coincidono con gli indici $\overline{q}(\Omega, \nu_Z)$ e $\underline{q}(\Omega, \nu_Z)$, rispettivamente.

Analogamente gli *indici locali* $\overline{q}(x, T, Z)$ e $\underline{q}(x, T, Z)$, relativi alla partizione Z, coincidono con gli indici $\overline{q}(\varphi_Z(x))$ e $\underline{q}(\varphi_Z(x))$, rispettivamente.

Valgono per gli indici globali e locali del sistema (X, T) gli analoghi dei Teoremi 4.11 e 4.12. Osserviamo che esistono esempi di sistemi dinamici, per cui si possono ottenere indici globali di caos diversi cambiando la misura di probabilità μ scelta.

Esempio 4.1. Trattiamo innanzitutto il caso delle rotazioni irrazionali del cerchio. Il nostro sistema dinamico (X, T_α, m) è dato da $X = [0,1]/(0 \sim 1)$, $T_\alpha : x \mapsto x + \alpha (\mathrm{mod}\ 1)$ con $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, e m misura di Lebesgue, che risulta T_α-invariante, ergodica e con entropia metrica nulla. La misura m è anche l'unica misura di probabilità invariante. Quindi le rotazioni irrazionali sono un esempio di sistema dinamico debolmente caotico.

Consideriamo la partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$, e calcoliamo gli indici locali di caos relativi alla partizione Z e alla misura m.

Teorema 4.13. *Per m-q.o $x \in X$ vale $\overline{q}(x, T_\alpha, Z) = \underline{q}(x, T_\alpha, Z) = 0$.*

Dimostrazione. La dimostrazione si basa sullo sviluppo in frazioni continue dell'angolo α di rotazione, e sulla sua approssimazione tramite razionali (vedi [RS]).

Dato un numero irrazionale $\alpha \in (0,1)$, la sua espansione in frazione continue è data da una successione di interi $(a_k)_{k \geq 1}$ detti *quozienti parziali*, tali che

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} \equiv [a_1, a_2, a_3, \dots] \tag{4.11}$$

Osserviamo che i numeri razionali hanno un'espansione finita, e indichiamo con $\frac{p_k}{q_k} = [a_1, a_2, \dots, a_k]$ la successione di approssimanti razionali del numero α. Questa approssimazione è la "migliore" possibile tramite numeri razionali, e in particolare vale la stima

$$\frac{1}{q_k^2 \cdot (a_{k+1} + 2)} < \left| \alpha - \frac{p_k}{q_k} \right| < \frac{1}{q_k^2 \cdot a_{k+1}} \tag{4.12}$$

Sia σ la stringa infinita in $\mathbb{N}^{\mathbb{N}}$ associata ad α, i cui simboli sono i quozienti parziali $(a_k)_k$. Analogamente, dato un punto $x \in (0,1)$, indichiamo con $[\xi_1, \xi_2, \dots, \xi_n, \dots]$ la sua espansione in frazioni continue, e sia s la stringa dei suoi quozienti parziali. Indichiamo inoltre con $\frac{r_h}{t_h}$ gli approssimanti razionali di x, ossia $\frac{r_h}{t_h} = [\xi_1, \xi_2, \dots, \xi_h]$.

Dalla stima (4.12) segue che la conoscenza dei primi $k+1$ quozienti parziali di α (che servono a specificare i primi k approssimanti razionali),

permette di concludere che, per ogni $n \in \mathbb{N}$, l'iterata n-esima $T_\alpha^n(x)$ si trova nell'intervallo

$$I_n^k := \left[(x + n\frac{p_k}{q_k} - \frac{n}{q_k^2 a_{k+1}})(\text{mod } 1), (x + n\frac{p_k}{q_k} + \frac{n}{q_k^2 a_{k+1}})(\text{mod } 1) \right]$$

D'altra parte, se l'intervallo I_n^k non contiene i punti 0 e $\frac{1}{2}$, la conoscenza degli estremi di questo intervallo permette di determinare il simbolo $\omega_n \in \{0, 1\}$ dell'orbita simbolica di x. Sia T_k la rotazione del cerchio di angolo razionale $\frac{p_k}{q_k}$, e sia ω^k l'orbita simbolica di x rispetto all'azione della mappa T_k. Allora, se l'intervallo I_n^k non contiene i punti 0 e $\frac{1}{2}$, si ha $\omega_n^k = \omega_n$. Poniamo

$$A(k, n) := \left\{ x \in X \ : \ \omega_n \neq \omega_n^k \right\}$$

Si ottiene allora che

$$m(A(k, n)) \leq \ \frac{4n}{q_k^2 a_{k+1}}$$

per ogni k e n. Quindi, la misura dei punti $x \in (0, 1)$ per cui $(\omega_1 \dots \omega_n) \neq (\omega_1^k \dots \omega_n^k)$ è data da

$$m\left(\bigcup_{i=1}^{n} A(k, i) \right) \leq \sum_{i=1}^{n} m(A(k, i)) \leq \ \sum_{i=1}^{n} \frac{4i}{q_k^2 a_{k+1}} = \frac{2n(n+1)}{q_k^2 a_{k+1}}$$

Sia $k(n)$ una successione di interi tale che

$$\sum_{n=1}^{\infty} \frac{4n}{q_{k(n)}^2 a_{k(n)+1}} \ < \infty \tag{4.13}$$

Per il Lemma di Borel-Cantelli, si conclude che per m-q.o. $x \in X$ esiste un intero $n_0 > 0$ tale che $x \in X \setminus A(k(n), n)$ per ogni $n \geq n_0$, e quindi

$$AIC(x, n, Z) \leq AIC(\omega_0^{k(n)} \omega_1^{k(n)} \dots \omega_{n-1}^{k(n)}) + cost$$

Basta infatti conoscere l'orbita simbolica di x per la mappa $T_{k(n)}$, e cambiare poi i primi n_0 simboli per ottenere l'orbita simbolica di x per la mappa T_α. La costante aggiunta dipende quindi dal punto x, ma non da n.

Bisogna quindi stimare il contenuto di informazione dei primi n simboli della stringa ω^k. Innanzitutto bisogna conoscere gli approssimanti razionali $\frac{p_k}{q_k}$, quindi la stringa σ^{k+1}. Serve poi conoscere il punto x. Data la stringa s

dei suoi quozienti parziali, conoscendo s^{h+1} è possibile approssimare x con i razionali $\frac{r_h}{t_h}$, e quindi $x + n\frac{p_k}{q_k} (\text{mod } 1)$ è contenuto nell'intervallo

$$J_n^h := \left[(\frac{r_h}{t_h} + n\frac{p_k}{q_k} - \frac{1}{t_h^2\, \xi_{h+1}}) (\text{mod} 1),\ (\frac{r_h}{t_h} + \frac{p_k}{q_k} + \frac{1}{t_h^2\, \xi_{h+1}}) (\text{mod } 1) \right]$$

Ripetiamo adesso lo stesso ragionamento di prima, e scegliamo una successione $h(n)$ tale che

$$\sum_{n=1}^{\infty} \frac{4}{t_{h(n)}^2\, \xi_{h(n)+1}} < \infty \tag{4.14}$$

Allora per m-q.o. $x \in X$, esiste un $n_1 > 0$ tale che, per ogni $n \geq n_1$, è possibile determinare ω_n^k dalla conoscenza di $[\xi_1, \xi_2, \ldots, \xi_{h(n)+1}]$.

Mettendo insieme quello che abbiamo visto finora, concludiamo che per m-q.o. $x \in X$ esistono funzioni a valori interi $k(n)$ e $h(n)$, tali che per n grande abbastanza

$$\begin{aligned} AIC(x, n, Z) &\leq AIC(\omega_0^{k(n)} \omega_1^{k(n)} \ldots \omega_{n-1}^{k(n)}) + cost \leq \\ &\leq AIC(\sigma^{k(n)+1}) + AIC(s^{h(n)+1}) + \log_2 n + cost \end{aligned} \tag{4.15}$$

dove il termine $\log_2 n$ è l'informazione necessaria a ricostruire l'orbita simbolica lunga $k(n)$ del punto razionale $\frac{r_{h(n)}}{t_{h(n)}}$ per la mappa $T_{k(n)}$. Notiamo infine che la costante è indipendente dalla lunghezza dell'orbita, ma dipende dal punto x.

Per concludere la dimostrazione, resta il problema di stimare $AIC(\sigma^k)$, per una generica stringa $\sigma = (a_1 a_2 \ldots)$ di quozienti parziali. Una prima stima è

$$AIC(\sigma^k) \leq k\ \log_2(\max\{a_i\ :\ i = 1, \ldots, k\}) + cost \tag{4.16}$$

dove abbiamo semplicemente applicato la Proposizione 3.4 estesa a un generico alfabeto di cardinalità $\max\{a_i\ :\ i = 1, \ldots, k\}$.

Lemma 4.14. *Poniamo* $f(k) := \max_{i \leq k} a_i$ *e per una successione* $k(n)$ *imponiamo che valga*

$$\sum_{n=1}^{\infty} \frac{4n}{q_{k(n)}^2 a_{k(n)+1}} < \infty \tag{4.17}$$

Allora è possibile scegliere $k(n)$ *tale che* $k(n) \log(f(k(n))) = o(n^\epsilon)$ *per ogni* $\epsilon > 0$.

L'applicazione di questo lemma ai termini $AIC(\sigma^{k(n)+1})$ e $AIC(s^{h(n)+1})$, usando la disuguaglianza (4.15), implica la tesi. □

Dimostrazione del Lemma 4.14. Dato un qualsiasi $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, la migliore stima sui denominatori q_k dei suoi approssimanti razionali è

$$\frac{1}{2\prod_{i=0}^{k-1} G^i(\alpha)} < q_k < \frac{1}{\prod_{i=0}^{k-1} G^i(\alpha)} \tag{4.18}$$

dove $G : [0,1] \to [0,1]$ è la *mappa di Gauss*, definita da $G(x) = \{1/x\}$ per $x \in (0,1]$ and $G(0) = 0$. Otteniamo in particolare che, per ogni k, $q_k \geq \frac{1}{2} f(k)$. Un'altra stima dal basso alla crescita dei denominatori q_k è fornita dai denominatori del numero aureo $\bar{\alpha} = \frac{\sqrt{5}-1}{2}$, i cui quozienti parziali verificano $a_i = 1$ per ogni $i \geq 1$. Dunque i denominatori $\bar{q}_k$ del numero aureo sono i più piccoli possibili, e sono uguali ai numeri di Fibonacci F_k. Quindi, in generale $q_k \geq F_k$ per ogni k.

Le due stime che abbiamo ottenuto per i denominatori q_k hanno lo stesso ordine di grandezza se la funzione $f(k)$ cresce esponenzialmente.

Se, più in generale, $f(k) = \mathcal{O}(\exp(k))$, usiamo la stima $q_k \geq F_k$. Otteniamo

$$\sum_{n=1}^{\infty} \frac{4n}{q_{k(n)}^2 a_{k(n)+1}} \leq \sum_{n=1}^{\infty} \frac{4n}{F_{k(n)}^2} \sim \sum_{n=1}^{\infty} \frac{n}{\left(\frac{1}{\bar{\alpha}}\right)^{2k(n)}}$$

Scegliendo quindi $k(n) \sim \log n^{\frac{3}{2}}$ otteniamo la condizione (4.17). Inoltre risulta $k \log(f(k)) = \mathcal{O}(k^2)$, da cui segue la tesi.

Se invece $f(k)$ cresce più velocemente di un'esponenziale, risulta più utile la stima $q_k \geq f(k)$. Infatti, possiamo scrivere in questo caso

$$\sum_{n=1}^{\infty} \frac{4n}{q_{k(n)}^2 a_{k(n)+1}} \leq \sum_{n=1}^{\infty} \frac{4n}{f^2(k(n))}$$

e la condizione (4.17) si ottiene scegliendo $f(k(n)) \sim n^{\frac{3}{2}}$. Se ne deduce che si può scegliere $k(n) = o(n^\epsilon)$ per ogni $\epsilon > 0$, quindi $k(n) \log(f(k(n))) \sim k(n) \log n = o(n^\epsilon)$ per ogni $\epsilon > 0$. □

Per concludere la trattazione delle rotazioni, osserviamo che abbiamo ottenuto, nel corso della dimostrazione, un interessante collegamento tra il AIC delle orbite simboliche e le proprietà aritmetiche dell'angolo di rotazione e della condizione iniziale. In particolare, più sono comprimibili le stringhe dei loro quozienti parziali, più è comprimibile l'orbita simbolica. △

Per le rotazioni del cerchio, abbiamo ottenuto gli indici locali di caos usando solo strumenti di teoria dei numeri, e stime generiche per il AIC. I risultati per la mappa T_{λ_∞} della famiglia logistica, e per le mappe T_z con $z \geq 2$ della famiglia di Pomeau-Manneville, richiedono invece l'utilizzo di strumenti più raffinati della teoria dei sistemi dinamici, in particolare la rinormalizzazione e lo studio dei tempi di ricorrenza. Ci limitiamo allora ad enunciare i risultati.

Teorema 4.15. *Data una partizione Z di $[0,1]$, fatta di un numero finito di intervalli, per ogni $x \in [0,1]$ vale $\overline{q}(x, T_{\lambda_\infty}, Z) = \underline{q}(x, T_{\lambda_\infty}, Z) = 0$.*

Teorema 4.16. *Per ogni misura di probabilità μ su $X = [0,1]/(0 \sim 1)$ assolutamente continua rispetto alla misura m di Lebesgue, e per ogni partizione Z di X in un numero finito di intervalli, vale*

$$\overline{q}(T_z, \mu, Z) = \underline{q}(T_z, \mu, Z) = \begin{cases} 1 & \text{per } z = 2 \\ \frac{1}{z-1} & \text{per } z > 2 \end{cases}$$

Inoltre per la partizione $Z = \{[0, x_z), [x_z, 1)\}$, dove x_z è il numero reale in $(0,1)$ soluzione di $x_z + x_z^z = 1$, otteniamo che gli indici locali di caos sono costanti m-q.o., e i loro valori $\overline{q}(T_z, Z)$ e $\underline{q}(T_z, Z)$ verificano, per ogni $z \geq 2$,

$$\underline{q}(T_z, Z) \leq \underline{q}(T_z, \mu, Z) = \overline{q}(T_z, \mu, Z) \leq \overline{q}(T_z, Z)$$

dove μ è come sopra.

Osserviamo che la conclusione del Teorema 4.16 relativa agli indici locali è più forte del Teorema 4.12, in quanto riguarda anche l'indice locale superiore, per il quale il Lemma di Fatou non ci fornisce risultati nel caso generale.

I risultati dei Teoremi 4.15 e 4.16 indicano che gli indici di caos possono essere buoni strumenti per la classificazione dei sistemi dinamici debolmente caotici. Strumenti il cui legame con altri indicatori classici della teoria statistica dei sistemi dinamici va studiato e approfondito. Da una parte per ampliare la trattazione teorica dei sistemi dinamici, ma anche per gli aspetti numerici del metodo legato alla teoria dell'informazione. Infatti, nonostante il AIC di una stringa sia una una funzione non computabile, è molto sviluppata, nell'informatica teorica, la ricerca di algoritmi in grado di approssimare bene il AIC di una stringa, in particolare i cosiddetti *algoritmi di compressione*, che sono alla base dei moderni mezzi di compressione dati, come zip, gzip, ecc.

Bibliografia

[AKM] R.L.Adler, A.G.Konheim, M.H.McAndrew, *Topological entropy*, Trans. Amer. Math. Soc. **114** (1965), 309–319

[Br] A.A.Brudno, *Entropy and the complexity of the trajectories of a dynamical system*, Trans. Moscow Math. Soc. **2** (1983), 127–151

[Ch] G.J.Chaitin, "Information, Randomness and Incompleteness. Papers on Algorithmic Information Theory", World Scientific, 1987

[CFS] I.P.Cornfeld, S.V.Fomin, Ya.G.Sinai, "Ergodic Theory", Springer-Verlag, 1982

[De] R.L.Devaney, "An Introduction to Chaotic Dynamical Systems", Addison-Wesley, 1989

[Fe] M.J.Feingenbaum, *Quantitative universality for a class of nonlinear transformation*, J. Stat. Phys. **19** (1978), 25–52

[Ga] G.Gallavotti, "Aspetti della Teoria Ergodica, Qualitativa e Statistica del moto", Quaderni UMI **21**, Pitagora editrice, 1982

[KH] A.Katok, B.Hasselblatt, "Introduction to the Modern Theory of Dynamical Systems", Cambridge University Press, 1995

[Ko] A.N.Kolmogorov, *Three approaches to the quantitative definition of information*, Problems of Information Transmission **1** (1965), 1–7

[LV] M.Li, P.Vitányi, "An Introduction to Kolmogorov Complexity and Its Applications", second edition, GTCS, Springer-Verlag, 1997

[Pe] K.Petersen, "Ergodic Theory", Cambridge University Press, 1983

[PY] M.Pollicott, M.Yuri, "Dynamical Systems and Ergodic Theory", Cambridge University Press, 1998

[PM] Y.Pomeau, P.Manneville, *Intermittent transition to turbulence in dissipative dynamical systems*, Comm. Math. Phys. **74** (1980), 189–197

[RS] A.M.Rockett, P.Szüsz, "Continued Fractions", World Scientific, 1992

[Sh] C.E.Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656

[Tu] A.M.Turing, *On computable numbers with an application to the Entscheidungsproblem*, Proc. London Math. Soc. (Ser. 2) **42** (1936), 230–265

[Wa] P.Walters, "An Introduction to Ergodic Theory", GTM **89**, Springer-Verlag, 1982