Università degli Studi di Pisa

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea in Matematica

Tesi di Laurea Specialistica

30 novembre 2007

Entropia di Kolmogorov-Sinai: dinamica simbolica e sostituzioni di Grassberger

Candidato

Lucio M. Calcagnile

Relatori

Dr. Stefano Galatolo

Dr. Giulia Menconi

Università di Pisa

Controrelatore

Dr. Carlo Carminati

Università di Pisa

Anno Accademico 2006/2007

Indice

Ringraziamenti

Ringrazio tutti coloro che in un modo o nell'altro mi hanno aiutato mentre lavoravo alla tesi e quando ho dovuto risolvere problemi con il LaTeX.

Ringrazio la mia famiglia e i miei amici, in particolare Alessandra (φ), Giovanni, Margherita, Riccardo, per l'ospitalità a Pisa, l'aiuto per la programmazione in C e i problemi di stampa su file dei grafici, il supporto morale.

Infine, oltre che ai miei relatori Stefano e Giulia, sono grato ai professori Dario Bini del Dipartimento di Matematica dell'Università di Pisa ed Emanuele Caglioti del Dipartimento di Matematica dell'Università di Roma "La Sapienza" per la disponibilità e la cortesia con le quali mi hanno offerto il loro aiuto.

Introduzione

Questa tesi si propone di studiare da un punto di vista teorico e computazionale un metodo di stima dell'entropia di serie temporali simboliche ovvero di sistemi dinamici ergodici e di confrontarlo con altri metodi classici di stima.

Il metodo in questione è stato proposto originariamente da Jimenez-Montaño, Ebeling e altri [EJM], ma è stato formalizzato nel linguaggio della teoria della probabilità e studiato piú approfonditamente da Grassberger [Gr] prima (2002) e Benedetto, Caglioti, Gabrielli [BCG] poi (2006). In [Gr] l'autore propone alcune dimostrazioni di fatti teorici e i risultati di applicazioni del metodo alla stima dell'entropia della lingua inglese. In [BCG] gli autori dimostrano molte questioni lasciate in sospeso da Grassberger e ne lasciano aperte delle altre.

Nella tesi chiameremo il metodo con l'acronimo NSRPS (da Non-Sequential Recursive Pair Substitution) o metodo delle sostituzioni di Grassberger.

Per spiegare come funziona il metodo NSRPS, supponiamo di avere una sorgente stazionaria a stati finiti, ossia un dispositivo che emette stringhe di simboli $x_0x_1x_2\ldots$, dove ogni x_i è un elemento di un alfabeto finito $A = \{a_1, a_2, \ldots, a_n\}$, in modo tale che la probabilità di ricevere una data stringa finita non vari nel tempo. Data una successione estratta dalla sorgente, il metodo di Grassberger prescrive di individuare la coppia (o una delle coppie) di simboli di maggiore frequenza e di sostituire tutte le occorrenze non sovrapposte, procedendo da sinistra verso destra, con un nuovo simbolo $\alpha \notin A$. Esemplificando, data la stringa

$$s^0 = 0110101110110001110110100 11,$$

estratta da una sorgente stazionaria μ con alfabeto $A = \{0, 1\}$ per cui $\mu(01)$ è massima tra le probabilità delle coppie di simboli, sostituiamo ogni occorrenza della sottostringa 01 con il nuovo simbolo 2, ottenendo

$$s^1 = 2122112100211212021.$$

Nel caso in cui la coppia da sostituire sia costituita da due simboli uguali, non vengono sostituite *tutte* le occorrenze della coppia, ma solo quelle non sovrapposte, partendo da sinistra. Esemplificando, se la sostituzione da eseguire è $xx \to y$, con $y \notin A$, una sottostringa costituita da $2k$ simboli x consecutivi viene sostituita da k simboli y, mentre una sottostringa di $2k+1$ simboli x consecutivi viene sostituita dalla stringa formata da k simboli y e un simbolo x. Se $A = \{0, 1\}$ e la sostituzione è $00 \to 2$, si ha ad esempio

$$001101000010100010000001100001 \to 211012210120122011221.$$

Se si parte da una sorgente μ con alfabeto $A = \{0, 1\}$, dopo la prima sostituzione abbiamo una nuova sorgente con alfabeto $A_1 = \{0, 1, 2\}$ e una misura μ_1 sulle stringhe finite ereditata da μ. Possiamo poi ripetere il procedimento introducendo nuovi simboli 3, 4, ... (ottenendo nuove sorgenti μ_2, μ_3, ...), fino a quando per qualche motivo non sia ragionevole concludere che ulteriori sostituzioni non siano utili.

Per esempio, se si ha a che fare con stringhe finite e la trasmissione di esse, si potrebbe ritenere di arrestare il processo delle sostituzioni quando le coppie di simboli piú frequenti non contano piú di una occorrenza, giacché in tal caso non vi sarebbe guadagno nel proseguire con un'altra sostituzione $xy \to z$, dovendo poi trasmettere oltre alla stringa accorciata anche la regola di decodifica $z \to xy$. Se le sostituzioni di coppie si attuano per motivi diversi (vedi capitolo 5) si dovrà determinare caso per caso quale ne sia il numero piú opportuno.

Il teorema principale riguardo al metodo NSRPS (teorema 3.4.1) afferma che l'entropia h di una sorgente ergodica μ, definita mediante le misure di tutte le stringhe finite $\underline{x} \in \cup_{i \in \mathbb{N}} A^i$ da

$$h(\mu) = \lim_{k \to \infty} -\frac{1}{k} \sum_{\underline{x} \in A^k} \mu(\underline{x}) \log_2 \mu(\underline{x}),$$

si può calcolare, nel limite per il numero n di sostituzioni che tende a infinito, conoscendo solo le probabilità $\mu_n(a)$ e $\mu_n(ab)$, $a, b \in A_n$, dei singoli simboli e delle coppie di simboli. Equivalentemente, si può dire che il procedimento delle sostituzioni di coppie di simboli descritto poco sopra trasforma, nel limite per il numero di sostituzioni che tende a infinito, qualunque processo ergodico in un processo 1-markoviano (vedi sezione 3.4).

La condizione di sostituire ad ogni passo una delle coppie di simboli con probabilità massima è una condizione sufficiente ma non necessaria perché valga la tesi del teorema 3.4.1. Sarebbe interessante capire quali sono le condizioni che il processo e le sostituzioni debbono necessariamente soddisfare affinché valga la conclusione del teorema 3.4.1. Ciò potrebbe essere una direzione di indagine futura.

Dal punto di vista teorico abbiamo studiato approfonditamente proprio un esempio di processo in cui, pur non valendo l'ipotesi sufficiente del teorema 3.4.1, vale la conclusione del medesimo teorema, e anzi una condizione ancora piú forte. L'esempio è stato proposto da Benedetto, Caglioti, Gabrielli nell'articolo [BCG], dove gli autori elencano solo i risultati omettendo del tutto i calcoli espliciti. Esponiamo con dovizia di particolari i dettagli di tale processo nella sezione 3.5.

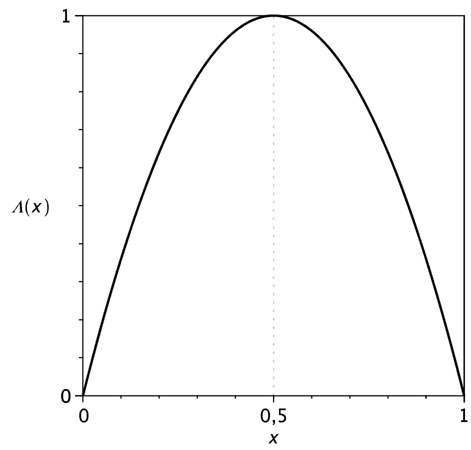
Per quanto riguarda il punto di vista computazionale, abbiamo applicato il metodo di Grassberger per ottenere una stima dell'entropia di trasformazioni ergodiche dell'intervallo $[0, 1]$ con la σ-algebra dei boreliani. Abbiamo a tal fine considerato quattro tipi di trasformazioni: la mappa logistica (figura 1(a)), una mappa espandente a pezzi con due rami suriettivi e derivata uniformemente maggiore di uno (figura 1(b)), una mappa di tipo Lorenz con due rami non suriettivi e derivata uniformemente maggiore di uno (figura 1(c)), una mappa di Manneville con due rami suriettivi e derivata maggiore di uno fuori da 0 e uguale a uno nell'origine (figura 1(d)).

Per poter giudicare la precisione e la velocità dell'algoritmo di stima suggerito dal metodo NSRPS, abbiamo applicato alle stesse trasformazioni altri tre metodi di stima dell'entropia: uno è il metodo delle frequenze empiriche dei blocchi simbolici, un altro è il metodo dei tempi di ritorno, il terzo è il calcolo dell'esponente di Lyapunov.

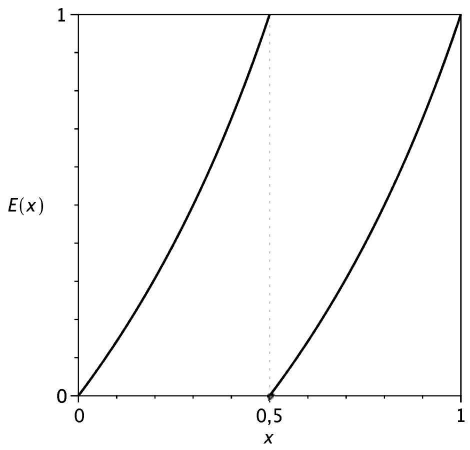
È importante osservare che vi è una sostanziale differenza fra il metodo di stima dell'entropia basato sul calcolo dell'esponente di Lyapunov e gli altri tre. Il primo infatti presuppone la conoscenza esplicita della mappa, mentre gli altri si applicano a successioni simboliche, le quali in generale possono essere ottenute senza che sia nota la definizione della mappa ovvero rilevate empiricamente. Da tale osservazione segue che il confronto tra la stima dell'entropia come esponente di Lyapunov e mediante l'algoritmo delle sostituzioni di Grassberger ha significato solo negli ambiti di applicazione comuni ad entrambi. Riprenderemo tale questione nelle *Conclusioni*, dove discuteremo il confronto fra i vari metodi essendo in possesso dei risultati delle simulazioni al calcolatore.

Tutti i programmi per le simulazioni sono stati realizzati nel linguaggio di programmazione *C*, utilizzando la libreria *GMP - GNU Multiple Precision* per la gestione di numeri in virgola mobile con precisione arbitraria. Si veda a tal proposito il sito Internet `http://gmplib.org/` dove si trova la documentazione completa.

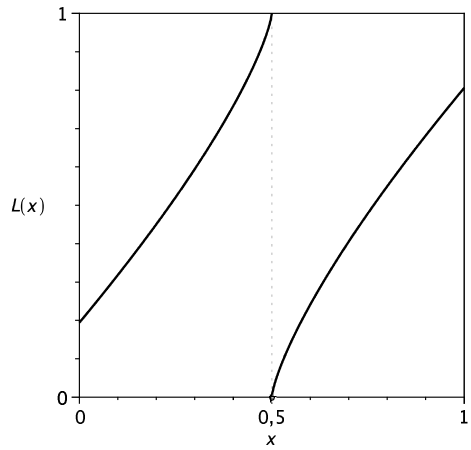
Nel *Capitolo 1* diamo una presentazione generale della teoria ergodica, introducendo gli endomorfismi di spazi di misura che la teoria ergodica considera, cioè le trasformazioni che conservano la misura (t.c.m.), il concetto di ergodicità, la dinamica simbolica con le trasformazioni di *shift*.



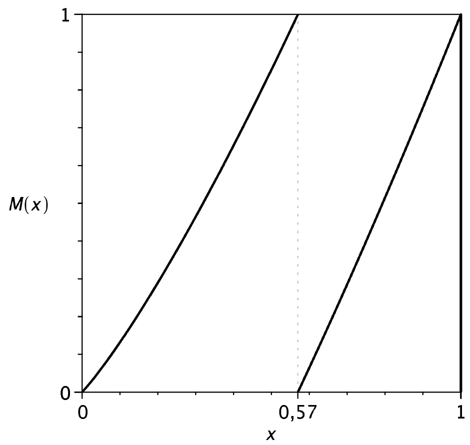
(a) La mappa logistica, definita da $\Lambda(x) = 4x(1-x)$.



(b) Una mappa espandente a pezzi, definita da $E(x) = \frac{4x}{3-2x}$ per $0 \le x \le 1/2$ e $E(x) = E(x - 1/2)$ per $1/2 < x \le 1$. Per essa vale $E' > 0$ uniformemente.



(c) Una mappa di tipo Lorenz, definita da $L(x) = \left(\frac{6x-3}{4}\right)^{3/4}$ per $1/2 \le x \le 1$ e $L(x) = 1 - L(1-x)$ per $0 \le x < 1/2$. La derivata è uniformemente maggiore di 1.



(d) La mappa di Manneville con esponente $\frac{3}{2}$. $M(x) = x + x^{\frac{3}{2}} \pmod 1$. La derivata è uguale a 1 in 0 e maggiore di 1 altrove.

Figura 1: I quattro tipi di trasformazioni dell'intervallo $[0, 1]$ su cui verranno applicati i vari metodi di stima dell'entropia.

Nel *Capitolo 2* presentiamo il concetto di entropia di una trasformazione che conserva la misura e il teorema piú importante riguardo ad essa, il teorema di Kolmogorov-Sinai.

Nel *Capitolo 3* illustriamo approfonditamente il metodo delle sostituzioni di Grassberger, discutendo e dimostrando il risultato fondamentale (teorema 3.4.1). Le dimostrazioni tecniche di lemmi e proposizioni sono rimandate all'appendice A per far sí che il lettore non perda di vista il filo conduttore. Quando nel corso del capitolo è stato fatto uso di equazioni ottenute nelle dimostrazioni tecniche e riportate in appendice, esse vengono preliminarmente richiamate. Nella sezione 3.5 esponiamo nel dettaglio un esempio di processo in cui vale la tesi del teorema centrale 3.4.1 pur non valendone l'ipotesi.

Nel *Capitolo 4* diamo una presentazione dei fatti teorici alla base degli altri tre metodi di stima dell'entropia con i quali confrontiamo il metodo di Grassberger.

Nel *Capitolo 5* presentiamo le trasformazioni dell'intervallo $[0, 1]$ considerate, illustriamo le simulazioni realizzate al calcolatore e ne riportiamo i risultati anche in forma di grafici.

Nelle *Conclusioni* infine operiamo una sintesi dei risultati ottenuti ed evidenziamo pregi e difetti dei vari metodi di stima considerati, con particolare attenzione al confronto fra il metodo delle sostituzioni di coppie e gli altri. Concludiamo indicando le direzioni di indagine piú interessanti riguardo al metodo NSRPS.

Capitolo 1

Generalità sulla teoria ergodica

1.1 Introduzione

Un sistema dinamico è definito da uno spazio X e dall'azione di un gruppo G su X. Lo spazio X è detto *spazio delle fasi* e rappresenta l'insieme degli stati possibili del sistema. L'evoluzione del sistema è determinata da delle trasformazioni $\{T_t : X \to X\}_{t\in G}$ tali che

(i) Se $e \in G$ è l'elemento neutro, allora $T_e = \mathrm{id}$

(ii) Per ogni $s, t \in G$, $\phi_{st} = \phi_s \circ \phi_t$.

Lo spazio X ha una struttura (tipicamente X è uno spazio di misura o uno spazio topologico o una varietà differenziabile) e le trasformazioni T_t preservano tale struttura (le T_t sono rispettivamente trasformazioni che conservano la misura o omeomorfismi o diffeomorfismi).

Il gruppo G è generalmente $\mathbb{R}$ (sistemi dinamici *continui*) o $\mathbb{Z}$ (sistemi dinamici *discreti*), ma spesso si definiscono sistemi dinamici dati dall'azione di un semigruppo, ad esempio $\mathbb{N}$.

Nel corso di questa tesi l'ambito sarà quello della teoria ergodica e l'oggetto che noi considereremo sarà uno spazio di probabilità $(X, \mathcal{B}, \mu)$ con una trasformazione T che conserva la misura.

La teoria ergodica è lo studio matematico del comportamento a lungo termine di certe quantità medie di dati sistemi dinamici. Il problema dell'esistenza delle medie attirò l'attenzione dei matematici in seguito all'interesse dei fisici riguardo alla cosiddetta *ipotesi ergodica*. Essa fu formulata da Boltzmann nell'ambito della meccanica statistica nel tentativo di concludere che la media temporale di una quantità osservabile f,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x),$$

e la media sullo spazio delle fasi,

$$\int_X f \, \mathrm{d}\mu,$$

coincidessero quasi ovunque. È senz'altro desiderabile che il valore atteso di una variabile fisica coincida con la sua media pesata su tutti i possibili stati del sistema, tuttavia l'ipotesi ergodica come espressa sopra è in generale falsa. La proprietà che il sistema deve possedere affinché la media sullo spazio e le medie temporali di una funzione a valori reali siano uguali è quella che oggi viene chiamata *ergodicità*.

La parola *ergodico* fu introdotta da Boltzmann e deriva dalla fusione delle parole greche $\varepsilon\rho\gamma o\nu$ (lavoro) e $o\delta o\varsigma$ (cammino).

1.2 Spazi di misura e trasformazioni che conservano la misura

Sia $\mathcal{X} = (X, \mathcal{B}, \mu)$ uno spazio di probabilità. Nel corso della tesi il nostro oggetto di indagine sarà sempre una trasformazione $T : \mathcal{X} \to \mathcal{X}$ misurabile che conserva la misura, cioè tale che

$$\forall B \in \mathcal{B} \quad T^{-1}(B) \in \mathcal{B} \text{ e } \mu(T^{-1}(B)) = \mu(B).$$

Equivalentemente, la condizione che T conserva la misura μ è espressa dicendo che μ è T-invariante.

Osserviamo che, poiché la composizione di trasformazioni misurabili che conservano la misura è misurabile e conserva la misura, ha senso considerare le trasformazioni iterate $T^n : \mathcal{X} \to \mathcal{X}$.

È difficile verificare in pratica se una data trasformazione conservi la misura oppure no usando la definizione appena data, poiché in genere non si conoscono esplicitamente tutti gli elementi di $\mathcal{B}$. Tuttavia è sufficiente, come in molti casi accade, avere esplicita conoscenza di una semi-algebra che genera $\mathcal{B}$. Una *semi-algebra* è una collezione $\mathcal{S}$ di sottoinsiemi di X per la quale valgono:

(i) $\varnothing \in \mathcal{S}$;

(ii) se $A, B \in \mathcal{S}$ allora $A \cap B \in \mathcal{S}$;

(iii) se $A \in \mathcal{S}$ allora $X \setminus A = \cup_{i=1}^n E_i$ con $E_i \in \mathcal{S}$ per ogni i e gli E_i a due a due disgiunti.

Ad esempio, nel caso in cui lo spazio è l'intervallo unitario $[0,1]$, la collezione dei sottointervalli è una semi-algebra. Vale infatti il seguente teorema (vedi Walters [Wa1], p. 20).

Teorema 1.2.1. *Siano* $\mathcal{X}_1 = (X_1, \mathcal{B}_1, \mu_1)$ *e* $\mathcal{X}_2 = (X_2, \mathcal{B}_2, \mu_2)$ *due spazi di probabilità e* $T : \mathcal{X}_1 \to \mathcal{X}_2$ *una trasformazione. Sia* $\mathcal{S}_2$ *una semi-algebra che genera* $\mathcal{B}_2$*. Se per ogni* $A_2 \in \mathcal{S}_2$ *si ha* $T^{-1}(A_2) \in \mathcal{B}_1$ *e* $\mu_1(T^{-1}(A_2)) = \mu_2(A_2)$*, allora* T *conserva la misura.*

Con un abuso di notazione, nel seguito useremo la scrittura $T : X \to X$ (e analoghe) al posto della piú corretta $T : \mathcal{X} \to \mathcal{X}$.

1.3 Esempi di trasformazioni che conservano la misura

1. La trasformazione identica su uno spazio $(X, \mathcal{B}, \mu)$ conserva ovviamente la misura.

2. Siano $S^1 = \{z \in \mathbb{C} : |z| = 1\}$, $\mathcal{B}$ la σ-algebra dei sottoinsiemi boreliani di S^1 e μ la misura di Haar. Sia inoltre $a \in S^1$ e si definisca $T : S^1 \to S^1$ come $T(z) = az$. Poiché μ è la misura di Haar, T conserva la misura. La trasformazione T è detta una *rotazione* di S^1.

3. La trasformazione $T(x) = ax$ definita su un qualunque gruppo compatto G (dove a è un dato elemento di G) conserva la misura di Haar. Tali trasformazioni sono dette *rotazioni* di G.

4. Ogni automorfismo continuo di un gruppo compatto conserva la misura.

   Dimostrazione. Sia $A : G \to G$ un automorfismo continuo e si denoti con m la misura di Haar su G. Si definisca una misura di probabilità μ sui sottoinsiemi boreliani di G tramite $\mu(E) = m(A^{-1}(E))$. La misura μ è regolare poiché m è regolare. Inoltre, $\mu(Ax \cdot E) = m(A^{-1}(Ax \cdot E)) = m(x \cdot A^{-1}E) = \mu(E)$ e quindi la misura μ è invariante per rotazioni. Da ciò segue che $\mu = m$, per l'unicità della misura di Haar. Ciò dimostra che A conserva la misura. $\square$

5. Per $n \in \mathbb{Z} \setminus \{0\}$ le trasformazioni $T(z) = z^n$ conservano la misura di Haar di S^1.

1.4 Isomorfismo di trasformazioni che conservano la misura

Nello studio delle trasformazioni che conservano la misura si rivela molto utile introdurre una nozione di isomorfismo tra di esse. Grazie a tale nozione è in molti casi possibile ricondurre lo studio di una trasformazione che conserva la misura di un dato spazio a una trasformazione di *shift* su spazi di successioni simboliche (vedi sezione 1.5).

Due trasformazioni isomorfe avranno o non avranno entrambe certe proprietà (ad esempio l'*ergodicità*, vedi sezione 1.6) e per esse si troveranno delle quantità invarianti (ad esempio l'*entropia*, vedi capitolo 2).

Definizione 1.4.1. *Siano* $(X_1, \mathcal{B}_1, \mu_1)$ *e* $(X_2, \mathcal{B}_2, \mu_2)$ *due spazi di probabilità con rispettive trasformazioni che conservano la misura*

$$T_1 : X_1 \to X_1, \quad T_2 : X_2 \to X_2.$$

Si dice che T_1 *e* T_2 *sono* isomorfe *(o, piú precisamente, che i sistemi dinamici* $(X_1, \mathcal{B}_1, \mu_1, T_1)$ *e* $(X_2, \mathcal{B}_2, \mu_2, T_2)$ *sono* isomorfi*), e si scrive* $T_1 \sim T_2$*, se esistono due insiemi* $M_1 \in \mathcal{B}_1$*,* $M_2 \in \mathcal{B}_2$ *con* $\mu_1(M_1) = 1$*,* $\mu_2(M_2) = 1$ *tali che*

(i) $T_1 M_1 \subseteq M_1$*,* $T_2 M_2 \subseteq M_2$*,*

(ii) esiste una trasformazione invertibile che conserva la misura

$$\phi : M_1 \to M_2 \quad \text{con } \phi T_1(x) = T_2 \phi(x) \ \forall x \in M_1,$$

tale cioè che sia commutativo il diagramma seguente.

$$\begin{tikzcd}
(M_1, \mu_1) \arrow[r, "T_1"] \arrow[d, "\phi"'] & (M_1, \mu_1) \arrow[d, "\phi"'] \\
(M_2, \mu_2) \arrow[r, "T_2"] & (M_2, \mu_2)
\end{tikzcd}$$

(Nella condizione (ii) l'insieme M_i*,* $i = 1, 2$*, si intende dotato della* σ*-algebra* $M_i \cap \mathcal{B}_i = \{M_i \cap B \mid B \in \mathcal{B}_i\}$ *e della restrizione della misura* μ_i *a tale* σ*-algebra.)*

Si dice che T_2 *è un* fattore *di* T_1 *se esistono* M_1 *e* M_2 *come sopra tali che vale (i) e*

(ii') esiste una trasformazione non necessariamente invertibile che conserva la misura

$$\phi : M_1 \to M_2 \quad \text{con } \phi T_1(x) = T_2 \phi(x) \ \forall x \in M_1.$$

In tal caso, T_1 *è detta un'*estensione *di* T_2*.*

Osserviamo che l'isomorfismo di trasformazioni che conservano la misura è una relazione di equivalenza. Inoltre, se $T_1 \sim T_2$, allora $T_1^n \sim T_2^n$ per ogni $n > 0$.

Esempio. Sia T la trasformazione $Tz = z^2$ sulla circonferenza unitaria S^1 con la σ-algebra di Borel e la misura di Haar, e sia U definita da $Ux = 2x \pmod 1$ su $[0, 1)$ con la σ-algebra di Borel e la misura di Lebesgue. Consideriamo l'applicazione $\phi : [0, 1) \to S^1$ definita da $\phi(x) = e^{2\pi i x}$. La mappa ϕ è biiettiva e conserva la misura (è sufficiente verificarlo per gli intervalli e usare il teorema 1.2.1). Inoltre vale $\phi U = T\phi$ e perciò $U \sim T$.

1.5 Dinamica simbolica

Nel descrivere il comportamento di un sistema dinamico si rivela molto utile avere opportuni modelli. Una buona fonte di modelli è fornita dagli spazi di successioni con le trasformazioni di *shift*. Presentiamo qui nel seguito tali spazi e trasformazioni di *shift* e come essi sono uno strumento importante per lo studio di un sistema dinamico.

1.5.1 Processi e spazi di successioni

Un processo stocastico a tempo discreto è una successione di variabili aleatorie $X_1, X_2, X_3, \ldots$ definite su uno spazio di probabilità (X, Σ, μ). In molte occasioni è utile considerare successioni bi-infinite $\ldots, X_{-1}, X_0, X_1, \ldots$, e quanto diremo per i processi indicizzati su $\mathbb{N}$ vale con le dirette generalizzazioni per i processi indicizzati su $\mathbb{Z}$. Per maggiori dettagli si veda Shields [Sh].

Diciamo che il processo ha *alfabeto* A se ogni X_i è a valori in A. A noi interesseranno solo processi ad alfabeto finito, e nel seguito intenderemo col termine 'processo' un processo stocastico a tempo discreto (indicizzato su $\mathbb{N}$ o su $\mathbb{Z}$) e ad alfabeto finito.

La successione finita $a_m, a_{m+1}, \ldots, a_n$, dove $a_i \in A$ per $m \leq i \leq n$, sarà indicata con a_m^n. L'insieme di tutte le stringhe a_m^n sarà indicato con A_m^n, tranne nel caso $m = 1$ in cui si userà A^n.

La *distribuzione congiunta di ordine* k del processo $\{X_k\}$ è la misura μ_k su A^k definita da

$$\mu_k(a_1^k) = \mathrm{Prob}(X_1^k = a_1^k), \quad \text{per ogni } a_1^k \in A^k.$$

Chiameremo i valori $\mu_k(a_1^k)$, $a_1^k \in A^k$, anche *valori* k*-marginali di* μ.

Chiamiamo *distribuzione* del processo l'insieme delle distribuzioni congiunte $\{\mu_k : k \geq 1\}$. La distribuzione di un processo può essere anche definita specificando la distribuzione iniziale μ_1 e per ogni $k \geq 2$ le distribuzioni condizionate

$$\mu(a_k | a_1^{k-1}) = \mathrm{Prob}(X_k = a_k | X_1^{k-1} = a_1^{k-1}) = \frac{\mu_k(a_1^k)}{\mu_{k-1}(a_1^{k-1})}.$$

La distribuzione di un processo è quindi una famiglia di distribuzioni di probabilità, una per ogni k. La successione non può però essere completamente arbitraria, poiché è implicito nella definizione di processo che valga la seguente *condizione di consistenza* per ogni $k \geq 1$,

$$\mu_k(a_1^k) = \sum_{a_{k+1} \in A} \mu_{k+1}(a_1^{k+1}), \quad \text{per ogni } a_1^k \in A^k. \tag{1.1}$$

Un processo è definito dalle sue distribuzioni congiunte e ciò significa in particolare che lo spazio su cui le funzioni misurabili X_n sono definite non è importante. Dato un processo, esiste un modo canonico di definire uno spazio di probabilità e delle variabili aleatorie su di esso che realizzano il processo dato. Tale costruzione è la cosiddetta *rappresentazione di Kolmogorov* ed è basata sugli spazi di successioni simboliche. La descriviamo brevemente qui di seguito, rimandando al testo di Shields [Sh] per una presentazione piú dettagliata e formale.

Sia A un insieme finito. Indichiamo con $A^{\mathbb{N}}$ l'insieme delle successioni infinite

$$x = (x_i), \quad x_i \in A, \ i \geq 1.$$

Il *cilindro* determinato da a_m^n, con $a_i \in A$ per $m \leq i \leq n$, che indichiamo con $[a_m^n]$, è il sottoinsieme di $A^{\mathbb{N}}$ definito da

$$[a_m^n] = \{x \in A^{\mathbb{N}} : x_i = a_i, \, m \leq i \leq n\}.$$

Indichiamo con Σ la σ-algebra generata dai cilindri. Gli elementi di Σ sono chiamati gli insiemi di Borel di $A^{\mathbb{N}}$. Per ogni $n \geq 1$ la funzione coordinata $\hat{X}_n : A^{\mathbb{N}} \to A$ è definita da $\hat{X}_n(x) = x_n$. Il prossimo teorema stabilisce che ogni processo con alfabeto A può essere pensato come il processo delle funzioni coordinate $\{\hat{X}_n\}$ sullo spazio $A^{\mathbb{N}}$ dotato della σ-algebra dei boreliani e di una misura di probabilità.

Teorema 1.5.1 (Teorema di rappresentazione di Kolmogorov)**.** *Se $\{\mu_k\}$ è una successione di misure per cui valgono le condizioni di consistenza (1.1), allora esiste un'unica misura di probabilità di Borel μ su $A^{\mathbb{N}}$ tale che $\mu([a_1^k]) =$*

$\mu_k(a_1^k)$, per ogni k e ogni a_1^k. In altre parole, se $\{X_n\}$ è un processo ad alfabeto finito A, esiste un'unica misura di Borel μ su $A^{\mathbb{N}}$ per la quale la successione delle funzioni coordinate $\{\hat{X}_n\}$ ha la stessa distribuzione di $\{X_n\}$.

Per la dimostrazione si veda Shields [Sh], pag. 2.

Un processo si dice *stazionario* se le distribuzioni congiunte non dipendono dalla scelta dell'origine dei tempi, cioè se

$$\mathrm{Prob}(X_i = a_i,\ m \leq i \leq n) = \mathrm{Prob}(X_{i+1} = a_i,\ m \leq i \leq n), \tag{1.2}$$

per ogni m, n e ogni a_m^n. L'idea intuitiva della stazionarietà è che il meccanismo che genera le variabili aleatorie non cambia nel tempo.

È spesso conveniente pensare che il processo vada avanti da molto tempo prima dell'osservazione, cosí che siano visibili solo gli effetti non transitori. Per tale interpretazione è piú funzionale lo spazio delle successioni bi-infinite $A^{\mathbb{Z}}$. La definizione di cilindro si estende immediatamente a questo caso e vale per lo spazio $A^{\mathbb{Z}}$ un analogo del teorema 1.5.1. Cioè, se $\{\mu_k\}$ è una famiglia di misure per cui valgono le condizioni di consistenza (1.1) e di stazionarietà (1.2), allora esiste un'unica misura di Borel su $A^{\mathbb{Z}}$ tale che $\mu([a_1^k]) = \mu_k(a_1^k)$, per ogni $k \geq 1$ e ogni a_1^k. La proiezione di μ su $A^{\mathbb{N}}$ è ovviamente la misura di Kolmogorov su $A^{\mathbb{N}}$ per il processo definito da $\{\mu_k\}$.

1.5.2 Trasformazioni di *shift*

Se le successioni simboliche costituiscono gli spazi di probabilità che in molti casi utilizzeremo, lo *shift* è la trasformazione che regola la dinamica di tali spazi.

Lo *shift* (verso sinistra) è l'applicazione $\sigma : A^{\mathbb{N}} \to A^{\mathbb{N}}$ definita da

$$(\sigma x)_n = x_{n+1}, \quad x \in A^{\mathbb{N}}, n \geq 1.$$

In sostanza, lo *shift* trasla tutti gli elementi della successione di un posto a sinistra e la sua definizione si può riscrivere in forma simbolica come

$$x = (x_1, x_2, x_3, \ldots) \Rightarrow \sigma x = (x_2, x_3, x_4, \ldots).$$

Si osservi che

$$\sigma^{-1}[a_m^n] = [b_{m+1}^{n+1}], \quad b_{i+1} = a_i, m \leq i \leq n,$$

quindi l'applicazione tra insiemi σ^{-1} manda insiemi cilindrici in insiemi cilindrici. Da ciò segue che per ogni insieme boreliano B la sua controimmagine $\sigma^{-1}B$ è anch'essa un insieme boreliano, e quindi lo *shift* σ è un'applicazione misurabile rispetto alla σ-algebra di Borel.

La condizione di stazionarietà del processo si traduce nelle uguaglianze $\mu([a_m^n]) = \mu([b_{m+1}^{n+1}])$ per ogni m, n e ogni a_m^n, cioè $\mu(B) = \mu(\sigma^{-1}B)$ per ogni cilindro B. Segue da ciò che vale l'uguaglianza $\mu(B) = \mu(\sigma^{-1}B)$ per ogni insieme boreliano B, cioè la trasformazione di *shift* σ conserva la misura μ o, equivalentemente, μ è σ-invariante.

Per processi stazionari esistono due rappresentazioni canoniche, una come misura di Kolmogorov su $A^{\mathbb{N}}$, l'altra come estensione σ-invariante allo spazio $A^{\mathbb{Z}}$. La seconda rappresentazione è detta *modello bi-infinito di Kolmogorov di un processo stazionario*. Si osservi che lo *shift* σ nel modello bi-infinito è invertibile.

1.5.3 Dalle trasformazioni che conservano la misura alla dinamica simbolica

Esponiamo in questa sezione un metodo molto utile per ricondurre lo studio di un sistema dinamico su un generico spazio allo studio di uno spazio di successioni con la trasformazione di *shift*.

Sia T una trasformazione invertibile che conserva la misura su uno spazio di probabilità $(X, \mathcal{B}, \mu)$ e sia $\alpha = \{A_1, \ldots, A_k\}$ una partizione misurabile finita di X. Indichiamo con Ω lo spazio prodotto $\{1, 2, \ldots, k\}^{\mathbb{Z}}$, cosí che un punto di Ω è una successione bi-infinita $\omega = (\omega_n)_{-\infty}^{\infty}$ dove $\omega_n \in \{1, 2, \ldots, k\}$ per ogni n. Su Ω è definita la trasformazione di *shift* σ come nella sezione 1.5.2.

Mostriamo ora che esistono una misura di probabilità σ-invariante ν sulla σ-algebra di Borel $\mathcal{B}(\Omega)$ di Ω e un'applicazione che conserva la misura $\phi_\alpha : (X, \mathcal{B}, \mu) \to (\Omega, \mathcal{B}(\Omega), \nu)$ tale che $\phi_\alpha T = \sigma\phi_\alpha$. Per $x \in X$, si definisce

$$\phi_\alpha x = (\omega_n)_{-\infty}^{\infty} \quad \text{se } T^n x \in A_{\omega_n}.$$

La n-esima coordinata di $\phi_\alpha x$ è cioè il numero dell'elemento di α cui appartiene $T^n x$.

Cosí risulta definita un'applicazione $\phi_\alpha : X \to \Omega$ e vale ovviamente $\sigma(\phi_\alpha x) = \phi_\alpha(Tx)$, per ogni $x \in X$. Per vedere che ϕ è misurabile si osservi che per ogni insieme cilindrico $[\omega_m^n]$, $m < n$, vale

$$\phi_\alpha^{-1}([\omega_m^n]) = \bigcap_{j=m}^{n} T^{-j} A_{\omega_j}.$$

Quindi, per ogni cilindro C vale $\phi_\alpha^{-1}(C) \in \mathcal{B}$, e poiché i cilindri generano $\mathcal{B}(\Omega)$ si ha $\phi_\alpha^{-1}(\mathcal{B}(\Omega)) \subseteq \mathcal{B}$.

La mappa ϕ_α trasporta in modo naturale la misura μ su $(\Omega, \mathcal{B}(\Omega))$:

$$\nu(E) \stackrel{\text{def}}{=} \mu(\phi_\alpha^{-1} E) \quad \text{per ogni insieme misurabile } E \subseteq \Omega.$$

Chiaramente $\phi_\alpha : X \to \Omega$ conserva la misura e dall'uguaglianza $\sigma\phi_\alpha = \phi_\alpha T$ discende che σ conserva ν.

Nel caso in cui la mappa T non sia invertibile, un'analoga costruzione si ha prendendo in considerazione lo spazio delle successioni simboliche $\Omega^+ = \{1, 2, \ldots, k\}^{\mathbb{N}}$.

La costruzione ora esposta è molto importante e su di essa baseremo le simulazioni al calcolatore che tratteremo nel capitolo 5. Stabiliamo ora della terminologia.

Definizione 1.5.2. *La successione $\phi_\alpha(x)$ è detta l'(α, T)*-nome di *x. Un punto $x \in X$ è detto α*-tipico per *T (o semplicemente* tipico*) se ogni blocco $\omega_1\omega_2 \ldots \omega_n$ appare nell'(α, T)-nome di x con frequenza $\mu(A_{\omega_1} \cap T^{-1}A_{\omega_2} \cap \ldots \cap T^{-n+1}A_{\omega_n})$.*

Osserviamo che, per il teorema ergodico di Birkhoff 1.6.3, quasi ogni punto di X è α-tipico per T.

L'applicazione $\phi_\alpha : X \to \Omega$ non è in generale invertibile, per cui non fornisce sempre un isomorfismo tra X e Ω. In generale, $(\Omega, \mathcal{B}(\Omega), \nu, \sigma)$ sarà solo un fattore di $(X, \mathcal{B}, \mu, T)$ (vedi definizione 1.4.1). Tuttavia, se la partizione α è generatrice (vedi definizione 1.5.3 qui sotto), allora si ha un isomorfismo, come stabilito dalla proposizione 1.5.4, per la cui dimostrazione si veda Petersen [Pe], pag. 274.

Definizione 1.5.3. *Sia T una trasformazione dello spazio di probabilità $(X, \mathcal{B}, \mu)$ che conserva la misura μ. Una partizione finita $\alpha = \{A_1, \ldots, A_n\}$ di X è detta* generatrice *se gli insiemi della forma*

$$A_{i_1} \cap T^{-1}A_{i_2} \cap \ldots \cap T^{-(k-1)}A_{i_k}$$

generano la σ-algebra $\mathcal{B}$.

Proposizione 1.5.4. *La mappa ϕ_α è un isomorfismo se e soltanto se α è una partizione generatrice per T.*

1.6 Ergodicità

Sia $(X, \mathcal{B}, \mu)$ uno spazio di probabilità e sia $T : X \to X$ una trasformazione che conserva la misura. Se $T^{-1}B = B$ per qualche $B \in \mathcal{B}$ (in questo caso B si dice *completamente invariante*), allora anche $T^{-1}(X \setminus B) = X \setminus B$, e potremmo studiare T considerando le due trasformazioni $T|_B$ e $T|_{X\setminus B}$. Se $0 < \mu(B) < 1$ ciò ci semplifica lo studio di T. Se invece $\mu(B) = 0$ o $\mu(X \setminus B) = 0$, considerare le restrizioni di T a B e al suo complementare non porta a una semplificazione significativa. Da tali considerazioni emerge l'importanza di studiare quelle trasformazioni che non possono essere decomposte come sopra. Le trasformazioni indecomponibili sono dette *ergodiche*:

Definizione 1.6.1. *Sia $(X, \mathcal{B}, \mu)$ uno spazio di probabilità. Una trasformazione T su $(X, \mathcal{B}, \mu)$ che conserva la misura si dice* ergodica *se i soli insiemi misurabili completamente invarianti hanno misura nulla o piena, cioè se*

$$T^{-1}B = B,\ B \in \mathcal{B} \quad \Rightarrow \quad \mu(B) = 0 \textit{ oppure } \mu(B) = 1.$$

Esistono molti altri modi equivalenti di definire la proprietà di ergodicità. Ne elenchiamo alcuni nel prossimo teorema, per la cui dimostrazione si veda ad esempio Walters [Wa1] o un qualunque testo di teoria ergodica.

Teorema 1.6.2. *Se $T : X \to X$ è una trasformazione che conserva la misura dello spazio di probabilità $(X, \mathcal{B}, \mu)$, allora le seguenti affermazioni sono equivalenti:*

(i) T *è ergodica.*

(ii) *Per ogni $A \in \mathcal{B}$ con $\mu(A) > 0$ si ha $\mu(\cup_{n=1}^{\infty} T^{-n}A) = 1$.*

(iii) *Per ogni A, $B \in \mathcal{B}$ con $\mu(A) > 0$, $\mu(B) > 0$ esiste $n > 0$ tale che $\mu(T^{-n}A \cap B) > 0$.*

(iv) *Se $f \in L^1(X, \mathcal{B}, \mu)$ è T-invariante (cioè $f \circ T = f$ q. o.) allora f è costante quasi ovunque.*

(v) *Per ogni A, $B \in \mathcal{B}$ si ha $\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) = \mu(A)\mu(B)$.*

Uno dei piú importanti risultati in teoria ergodica fu dimostrato da G. D. Birkhoff nel 1931 [Bir]. Noi lo enunciamo nel caso degli spazi di probabilità, ma esso vale piú in generale nel caso di spazi di misura σ-finita.

Teorema 1.6.3 (Teorema ergodico, Birkhoff 1931)**.** *Siano $(X, \mathcal{B}, \mu)$ uno spazio di probabilità e $T : X \to X$ una trasformazione misurabile su di esso che conserva la misura. Data $f \in L^1(X, \mathcal{B}, \mu)$, esiste $f^* \in L^1(X, \mathcal{B}, \mu)$ tale che*

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \quad \textit{q. o.}$$

Inoltre f^* è T-invariante (cioè $f^* \circ T = f^*$ q. o.) e $\int_X f^* \mathrm{d}\mu = \int_X f \mathrm{d}\mu$.

Osserviamo che, se T è ergodica, dall'affermazione (iv) del teorema 1.6.2 e dal teorema ergodico di Birkhoff discende che $f^*(x) = \int_X f^* \mathrm{d}\mu$ q. o. Da tale osservazione segue il seguente corollario.

Corollario 1.6.4. *Siano* $(X, \mathcal{B}, \mu)$ *uno spazio di probabilità e* $T : X \to X$ *una trasformazione che conserva la misura. Sia inoltre* $E \in \mathcal{B}$. *Se* T *è ergodica allora*

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_E\big(T^i(x)\big) \to \mu(E) \quad q.\ o.$$

In altre parole, se T *è ergodica allora l'orbita di quasi ogni punto di* X *entra nell'insieme* E *con frequenza asintotica uguale a* $\mu(E)$.

Il teorema seguente asserisce che l'ergodicità è una proprietà invariante per isomorfismo (per la dimostrazione si veda per esempio Walters [Wa1]).

Teorema 1.6.5. *Siano* $(X_1, \mathcal{B}_1, \mu_1)$ *e* $(X_2, \mathcal{B}_2, \mu_2)$ *due spazi probabilità con trasformazioni misurabili che conservano la misura* $T_i : X_i \to X_i$, $(i = 1, 2)$. *Se* T_1 *è isomorfa a* T_2 *e* T_2 *è ergodica, allora anche* T_1 *è ergodica.*

Capitolo 2

Entropia

2.1 Introduzione

Il concetto di entropia fu introdotto per la prima volta nella termodinamica da Clausius nel 1854; Shannon lo trasportò in teoria dell'informazione nel 1948 e Kolmogorov in teoria ergodica nel 1958. In tutti i casi l'entropia è una misura della casualità o del disordine.

Quando Claude Shannon iniziò uno studio matematico della trasmissione dell'informazione, si basò sulle idee fondamentali che (1) la quantità di informazione trasmessa è misurata dalla quantità di incertezza rimossa e (2) l'incertezza di una trasmissione è proporzionale al valore atteso del logaritmo della probabilità del simbolo ricevuto.

Per essere piú specifici, si consideri una *sorgente*, ovvero un oggetto che emette stringhe di simboli $\dots x_0 x_1 x_2 \dots$, dove gli a_i sono elementi di un alfabeto finito $A = \{a_1, a_2, \dots, a_n\}$. Si supponga di ricevere il simbolo a_i con probabilità p_i, per ogni $i = 1, \dots, n$. Se ogni simbolo viene trasmesso indipendentemente da ciò che è stato trasmesso in precedenza, si può prendere come *quantità di informazione trasmessa mediamente per simbolo* la quantità

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{2.1}$$

(Si usa 2 come base del logaritmo perché normalmente l'informazione è misurata in termini binari, contando il numero delle domande a risposta affermativa/negativa cui è necessario rispondere per comunicare l'informazione stessa.)

Si noti che se qualche $p_{i_0} = 1$ (e tutti gli altri $p_i = 0$), allora non c'è incertezza alcuna su ciò che farà la sorgente. In questo caso, usando la convenzione $0 \log 0 = 0$, si ha $H = 0$. H è massima invece quando tutti i p_i sono uguali a $1/n$. Per vedere questo, si consideri la funzione f definita in $[0,1]$

$$f(t) = \begin{cases} 0 & \text{se } t = 0 \\ -t \log_2 t & \text{se } 0 < t \leq 1. \end{cases} \tag{2.2}$$

La funzione f è continua, non negativa e convessa. Dalla disuguaglianza di Jensen segue che comunque si scelgano positivi $\lambda_1, \lambda_2, \ldots, \lambda_n \in [0,1]$,

$$f\left(\frac{1}{n}\sum_{i=1}^{n} \lambda_i\right) \geq \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i). \tag{2.3}$$

Quindi, per ogni vettore di probabilità $(\lambda_1, \lambda_2, \ldots, \lambda_n)$,

$$\frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \leq f\left(\frac{1}{n}\sum_{i=1}^{n} \lambda_i\right) = f\left(\frac{1}{n}\right) = -\frac{1}{n}\log_2 \frac{1}{n} = \frac{1}{n}\log_2 n, \tag{2.4}$$

per cui $H(\lambda) \leq \log_2 n$. E chiaramente $H(1/n, 1/n, \ldots, 1/n) = \log_2 n$. Questo caso in cui tutti i simboli sono equiprobabili è quello in cui è massima l'incertezza su quale sarà il simbolo emesso dalla sorgente e perciò quello in cui è massima la quantità di informazione comunicata.

H è la quantità media di informazione trasmessa con ogni simbolo, se si prende come misura di informazione di un evento E la quantità $-\log_2 P(E)$. Il fatto che questa scelta sia ragionevole e utile è motivato dal fatto che l'unica misura di informazione della forma $f(P(E))$, dove $f : \mathbb{R} \to [0,\infty)$ è continua, additiva su eventi indipendenti ($f(P(E_1 \cap E_2)) = f(P(E_1)P(E_2)) = f(P(E_1)) + f(P(E_2))$) è $-\log_r P(E)$ per qualche $r > 0$. Per un'altra giustificazione della definizione di H si veda Khintchine [Kh].

Per una sorgente generica, la probabilità di ricevere un dato simbolo può dipendere da ciò che è stato ricevuto precedentemente. Per esempio, se la sorgente emette testo di prosa in italiano, la probabilità di ricevere una 'u' salirà enormemente ogni qual volta venga emssa una 'q'. Tali effetti di dipendenza possono essere tenuti in conto nel calcolo dell'informazione media per simbolo considerando blocchi di simboli.

Per ogni $k = 1, 2, \ldots$ sia $\mathcal{C}_k$ l'insieme di tutti i blocchi di lunghezza k formati da simboli dell'alfabeto A. Ogni blocco C ha una certa probabilità $P(C)$ (eventualmente 0) di essere ricevuto. Pertanto, l'informazione media *per simbolo* in una trasmissione di lunghezza k è

$$H_k = -\frac{1}{k}\sum_{C \in \mathcal{C}_k} P(C)\log_2 P(C). \tag{2.5}$$

Si definisce l'*entropia della sorgente* come

$$h = \lim_{k\to\infty} -\frac{1}{k}\sum_{C \in \mathcal{C}_k} P(C)\log_2 P(C). \tag{2.6}$$

Il limite esiste (vedi proposizione 2.3.3), ed esso rappresenta la quantità media di informazione che la sorgente comunica con ogni simbolo emesso. Equivalentemente, h misura il grado medio di incertezza che un lettore della sorgente ha circa il simbolo successivo che verrà emesso, qualunque cosa abbia ricevuto prima.

2.2 Teoria ergodica

La sorgente discussa nella sezione precedente non era altro che un processo stazionario a stati finiti.

Gli *output* della sorgente sono stringhe infinite $\dots x_0 x_1 x_2 \dots$ di simboli presi da un alfabeto $A = \{a_1, a_2, \dots, a_n\}$, ovvero elementi dello spazio di successioni $A^{\mathbb{Z}}$. Assumendo che la sorgente sia stazionaria, tale cioè che la probabilità di ricevere una data stringa non cambi nel tempo, esiste allora una misura su $A^{\mathbb{Z}}$ conservata dalla trasformazione di *shift* $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ (vedi sezione 1.5.1). Se μ è la misura prodotto generata dai pesi p_i sugli a_i, si è in presenza dello *shift* di Bernoulli; esso è il caso in cui i simboli sono trasmessi indipendentemente da ciò che è stato trasmesso in precedenza. Calcolando l'entropia della sorgente secondo la formula (2.6), si è trovata l'entropia di un sistema dinamico $(A, \mathcal{B}, \mu, \sigma)$.

Ricordiamo che, dato un sistema dinamico $(X, \mathcal{B}, \mu, T)$ che conserva la misura, si può associare ad esso un processo stazionario a stati finiti considerando una partizione misurabile finita $\alpha = \{A_1, A_2, \dots, A_n\}$ di X. Ogni punto $x \in X$ determina una stringa simbolica infinita che al tempo j ha il simbolo A_i se $T^j x \in A_i$ (vedi sezione 1.5.3).

Definizione 2.2.1. *L'entropia della partizione $\alpha = \{A_1, A_2, \dots, A_n\}$ è la quantità*

$$H(\alpha) = -\sum_{i=1}^{n} \mu(A_i) \log_2 \mu(A_i). \tag{2.7}$$

La comparsa di un blocco di simboli $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ a partire dal posto di indice j corrisponde all'entrata del punto $T^j x$ nell'insieme $A_{i_1} \cap T^{-1} A_{i_2} \cap \dots \cap T^{-k+1} A_{i_k}$. Tale insieme è un elemento della partizione $\alpha_0^{k-1} = \alpha \vee T^{-1}\alpha \vee \dots \vee T^{-k+1}\alpha$, che è il minimo comune raffinamento delle partizioni $\alpha, T^{-1}\alpha, \dots, T^{-k+1}\alpha$.

Definizione 2.2.2. *Se $\alpha = \{A_1, \dots, A_n\}$ e $\beta = \{B_1, \dots, B_m\}$ sono due partizioni, allora*

- *$T^{-1}\alpha$ è la partizione $\{T^{-1}(A_1), \dots, T^{-1}(A_n)\}$,*

- $\alpha \vee \beta$ *è la partizione* $\{A_i \cap B_j : i = 1, \ldots, n; j = 1, \ldots, m\}$,
- β *è un* raffinamento *di* α *(si scrive* $\beta \geq \alpha$*), se ogni* B_j *è, a meno di un insieme di misura nulla, un sottoinsieme di qualche* A_i.

Alla luce di queste definizioni, la quantità H_k in (2.5) è

$$\frac{1}{k} H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-k+1}\alpha),$$

e l'entropia della sorgente definita dalla (2.6) è

$$h(\alpha, T) = h_\mu(\alpha, T) = \lim_{k \to \infty} \frac{1}{k} H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-k+1}\alpha).$$

Definizione 2.2.3. *La quantità* $h(\alpha, T)$ *è detta* entropia della trasformazione T rispetto alla partizione α*.*

Il limite nella definizione di $h(\alpha, T)$ esiste, come dimostreremo piú avanti nella proposizione 2.3.3.

Il numero $h(\alpha, T)$ è una misura dell'incertezza media per unità di tempo che si ha riguardo a quale sia l'elemento della partizione α in cui entrerà un dato punto x (mosso dalla trasformazione T), dato il suo passato. Ovviamente, se α è scelta in modo banale potrebbe non esserci molta incertezza su quali saranno gli elementi della partizione stessa in cui cadranno i $T^j x$, e dire quali siano effettivamente potrebbe non comunicare molta piú informazione di quanto non fosse possibile indovinare *a priori*. Pertanto, l'*entropia della trasformazione* T è definita come la massima incertezza su tutti i possibili processi a stati finiti associati a T:

$$h(T) = h_\mu(T) = \sup_\alpha h(\alpha, T).$$

Il numero $h(T)$ è una misura della nostra incertezza media riguardo a dove T muove i punti di X. La grandezza di $h(T)$ riflette il grado con cui la trasformazione T 'mischia' lo spazio.

Chiaramente, $h(T)$ è un invariante nella classe di isomorfismo di T. Piú in generale, se S è un fattore di T allora $h(S) \leq h(T)$ (perché ogni partizione relativa al sistema dinamico S determina una partizione di uguale entropia relativa a T, ma non necessariamente viceversa).

Si può inoltre pensare ad una partizione α come la collezione dei possibili risultati di un esperimento (per esempio la lettura del risultato di un lancio di dado). Il numero $H(\alpha)$ è una misura dell'incertezza media che si ha circa il risultato dell'esperimento, ovvero, equivalentemente, della quantità di informazione che si guadagna realizzando l'esperimento. Il raffinamento comune (o *join*) $\alpha \vee \beta$ rappresenta l'esperimento composto, formato dalla realizzazione simultanea degli esperimenti α e β. Pertanto,

$$\frac{1}{k} H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-k+1}\alpha).$$

è la quantità media di informazione per ripetizione ottenuta realizzando k ripetizioni dell'esperimento α, $h(\alpha, T)$ è la media temporale del contenuto di informazione dell'esperimento α, e $h(T)$ è la massima informazione per ripetizione che si può ottenere da un qualsivoglia esperimento, fintanto che si usi T per l'evoluzione del sistema.

È naturale chiedersi se sia possibile ideare un esperimento che estragga dal sistema tutta l'informazione che *a priori* si può avere, cioè tale che

$$h(\alpha, T) = h(T).$$

È ragionevole credere che ciò possa accadere se le partizioni $\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-k+1}\alpha$ generano l'intera σ-algebra, giacché in tal caso i possibili risultati di un qualunque esperimento possono essere descritti in termini di stringhe finite nell'alfabeto dato da α con un grado di accuratezza qualunque.

Nel 1958 Kolmogorov (in [Ko]) definí l'entropia di T come $h(\alpha, T)$ nel caso in cui T abbia una partizione generatrice α e ∞ altrimenti; la definizione data in precedenza è dovuta a Sinai (vedi [Si], 1959) e l'equivalenza delle due definizioni è data da quello che solitamente si chiama teorema di Kolmogorov-Sinai (teorema 2.4.1). Esso è lo strumento principale per il calcolo dell'entropia negli esempi concreti.

Prima della dimostrazione del teorema di Kolmogorov-Sinai enunciamo alcune proprietà dell'entropia e dell'entropia condizionata e dimostriamo l'esistenza del limite

$$h(\alpha, T) = \lim_{k \to \infty} \frac{1}{k} H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-k+1}\alpha).$$

2.3 Informazione e condizionamento

In questa sezione raccogliamo i risultati tecnici per la dimostrazione del teorema di Kolmogorov-Sinai (teorema 2.4.1). Per le dimostrazioni si veda ad esempio Petersen [Pe].

È utile considerare i concetti di informazione ed entropia condizionate, i quali tengono conto dell'informazione di cui si dovesse essere già in possesso.

Sia $\mathcal{F} \subset \mathcal{B}$ una sotto-σ-algebra della famiglia dei sottoinsiemi misurabili di X. Si ricordi che per $\phi \in L^1(X)$ la *speranza condizionata* $E(\phi|\mathcal{F})$ di ϕ

data $\mathcal{F}$ è una funzione $\mathcal{F}$-misurabile su X che soddisfa

$$\int_F E(\phi|\mathcal{F})\mathrm{d}\mu = \int_F \phi \mathrm{d}\mu$$

per ogni $F \in \mathcal{F}$. $E(\phi|\mathcal{F})(x)$ rappresenta il valore atteso di ϕ *data la conoscenza di* $\mathcal{F}$, cioè sapendo a quali insiemi di $\mathcal{F}$ appartiene il punto x. La *probabilità condizionata* di un insieme $A \in \mathcal{B}$ data $\mathcal{F}$ è

$$\mu(A|\mathcal{F}) = E(\chi_A|\mathcal{F}).$$

Essa rappresenta la misura della probabilità dell'occorrenza di A, rivalutata una volta che si conosca l'informazione contenuta in $\mathcal{F}$. Si definisce la *funzione di informazione condizionata* di una partizione α, data una σ-algebra $\mathcal{F} \subset \mathcal{B}$, come

$$I_{\alpha|\mathcal{F}}(x) = -\sum_{A\in\alpha} \log_2 \mu(\alpha|\mathcal{F})\chi_A(x).$$

L'*entropia condizionata* di α, data $\mathcal{F}$, è definita da

$$H(\alpha|\mathcal{F}) = \int_X I_{\alpha|\mathcal{F}}(x)\mathrm{d}\mu(x).$$

Proposizione 2.3.1. *$H(\alpha|\mathcal{F}) = 0$ se e solo se $\alpha \subseteq \mathcal{F}$ a meno di insiemi di misura nulla.*

Ricordiamo che $\alpha \subseteq \mathcal{F}$ significa che per ogni $A \in \alpha$ esiste $F \in \mathcal{F}$ tale che l'insieme $(A \setminus F) \cup (F \setminus A)$ ha misura nulla.

Lemma 2.3.2. *Se α e β sono due partizioni misurabili numerabili, allora*

(i) $H(T^{-1}\alpha) = H(\alpha)$.

(ii) Se $\alpha \leq \beta$ allora $H(\alpha) \leq H(\beta)$.

(iii) $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.

Proposizione 2.3.3. *Per ogni partizione misurabile numerabile α di X esiste il limite (eventualmente uguale a $+\infty$)*

$$h(\alpha, T) = \lim_{n\to\infty} \frac{H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha)}{n}.$$

Dimostrazione. Si supponga $H(\alpha) < \infty$, poiché ovviamente in caso contrario si ha $h(\alpha, T) = \infty$. Per ogni $n = 1, 2, \ldots$ sia $H_N = H(\alpha \vee T^{-1}\alpha \vee \ldots \vee$

$T^{-n+1}\alpha$). Dal lemma segue che la successione $\{H_n\}$ è crescente. Essa è anche *subadditiva*:

$$\begin{aligned} & H_{m+n} \\ = \quad & H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-m+1}\alpha \vee T^{-m}(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha)) \\ \leq \quad & H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-m+1}\alpha) + H(\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-n+1}\alpha) \\ = \quad & H_m + H_n. \end{aligned}$$

Chiaramente, $H_{jn} \leq jH_n$ e $H_j \leq jH_1$, e quindi la successione $\{H_j/j\}_{j\geq 1}$ è limitata. Pertanto,

$$L = \liminf \frac{H_j}{j} < \infty.$$

Dato $\varepsilon > 0$, si scelga j tale che $H_j/j < L + \varepsilon$. Per ogni $n > j$ si scelga $i(n)$ tale che

$$[i(n) - 1]j < n \leq i(n)j.$$

Allora $H_n \leq H_{i(n)j}$, e pertanto

$$\frac{H_n}{n} \leq \frac{Hi(n)j}{n} \leq \frac{H_{i(n)j}}{[i(n)-1]j} \leq \frac{i(n)H_j}{[i(n)-1]j} < \frac{i(n)}{i(n)-1}(L+\varepsilon).$$

Per n abbastanza grande si ha allora

$$L - \varepsilon < \frac{H_n}{n} < L + 2\varepsilon.$$

□

Proposizione 2.3.4. *Se $\mathcal{B}_1 \subset \mathcal{B}_\in \subset \ldots$ sono sotto-σ-algebre di $\mathcal{B}$ e α è una partizione misurabile* finita*, allora*

$$\lim_{n\to\infty} H(\alpha|\mathcal{B}_n) = H(\alpha|\bigvee_{n=1}^{\infty} \mathcal{B}_n).$$

Proposizione 2.3.5. *Comunque si prendano due partizioni misurabili numerabili α e β, vale*

$$h(\beta, T) = h(\alpha, T) + H(\beta|\mathcal{B}(\alpha)).$$

($\mathcal{B}(\alpha)$ è la σ-algebra generata da α.)

Proposizione 2.3.6. *Per ogni $n \in \mathbb{N}$ vale*

$$h(\bigvee_{i=-n}^{n} T^{-i}\alpha, T) = h(\alpha, T).$$

2.4 Il teorema di Kolmogorov-Sinai

Il calcolo di $h(T)$ dalla sua definizione come $\sup_\alpha h(\alpha, T)$ è raramente possibile. Ma se esistesse una partizione α per la quale $h(\alpha, T) = h(T)$, il calcolo di $h(T)$ sarebbe molto piú semplice. Tale è il caso in cui α è una partizione che genera l'intera σ-algebra $\mathcal{B}$.

Per ogni partizione α verranno usate le notazioni $\alpha_m^n = \bigvee_{k=m}^n T^{-k}\alpha$; nei casi in cui $m = -\infty$ o $n = \infty$, si intenderanno con tale notazione le σ-algebre generate dalle partizioni:

$$\begin{aligned}
\alpha_1^\infty &= \textstyle\bigvee_{k=1}^\infty \mathcal{B}(T^{-k}\alpha), \\
\alpha_{-\infty}^{-1} &= \textstyle\bigvee_{k=1}^\infty \mathcal{B}(T^{k}\alpha), \\
\alpha_{-\infty}^{\infty} &= \textstyle\bigvee_{k=-\infty}^\infty \mathcal{B}(T^{k}\alpha).
\end{aligned}$$

Una partizione finita α è generatrice per T se $\alpha_{-\infty}^\infty = \mathcal{B}$ a meno di insiemi di misura nulla (vedi definizione 1.5.3).

Teorema 2.4.1 (Kolmogorov-Sinai)**.** *Se α è una partizione generatrice per T, allora*

$$h(T) = h(\alpha, T).$$

Dimostrazione. È sufficiente dimostrare che per ogni partizione finita β si ha $h(\beta, T) \leq h(\alpha, T)$. Per le proposizioni 2.3.5 e 2.3.6 si ha, per ogni $n = 1, 2, \ldots$,

$$h(\beta, T) \leq h(\alpha_{-n}^n, T) + H(\beta|\mathcal{B}(\alpha_{-n}^n)) = h(\alpha, T) + H(\beta|\mathcal{B}(\alpha_{-n}^n)).$$

Ma poiché $\beta \subseteq \alpha_{-\infty}^\infty$ a meno di un insieme di misura nulla, per le proposizioni 2.3.1 e 2.3.4 si ha

$$0 = H(\beta|\alpha_{-\infty}^\infty) = \lim_{n\to\infty} H(\beta|\mathcal{B}(\alpha_{-n}^n)).$$

La tesi segue mandando n all'infinito nella prima formula. □

Nel 1970 Krieger dimostrò in [Kr] il seguente teorema riguardo all'esistenza di partizioni generatrici.

Teorema 2.4.2 (Krieger)**.** *Se un sistema dinamico $(X, \mathcal{B}, \mu, T)$ è ergodico e di entropia finita, allora esiste una partizione generatrice finita per T.*

Capitolo 3

Il metodo NSRPS

3.1 Introduzione

Verrà ora illustrato il metodo delle sostituzioni di coppie di simboli studiato da Grassberger in [Gr]. Chiameremo esso piú brevemente metodo NSRPS (da Non-Sequential Recursive Pair Substitution).

Si consideri una stringa finita s^0 di simboli presi da un alfabeto finito $A = \{a_0, \ldots, a_{m-1}\}$. Per ogni coppia di indici i, j sia n_{ij} il numero delle occorrenze non sovrapposte della stringa $a_i a_j$ in s^0 e sia i_0, j_0 la coppia (o una delle coppie) per cui n_{ij} è massimo. Si definisce una nuova successione s^1 ottenuta da s^0 sostituendo ogni occorrenza della coppia $a_{i_0} a_{j_0}$ con un nuovo simbolo a_m. La nuova stringa è piú corta della vecchia e il suo alfabeto contiene un simbolo in piú. Partendo da s^1 è possibile definire un'altra stringa s^2 seguendo la stessa procedura, e cosí via.

Esemplificando, si consideri dapprima la stringa

$$s^0 = 01101011101100011101101001 1,$$

nella quale si sostituisce la coppia 01 con il nuovo simbolo 2. Si ottiene

$$s^1 = 2122112100211212021.$$

Indicando con $|s|$ la lunghezza di una stringa s, si ha

$$|s^1| = |s^0| - \#\{01 \subseteq s^0\},$$

dove con $\#\{01 \subseteq s^0\}$ si è indicato il numero di occorrenze di 01 nella stringa s^0. Dividendo per $|s^0|$,

$$\frac{|s^1|}{|s^0|} = 1 - \frac{\#\{01 \subseteq s^0\}}{|s^0|}.$$

Nel seguito si avrà a che fare sempre con successioni estratte da una sorgente ergodica μ. Prendendo allora il limite per $|s^0| \to \infty$ si ottiene che, per quasi ogni successione s^0,

$$\frac{1}{Z} = \lim_{|s^0| \to \infty} \frac{|s^1|}{|s^0|} = 1 - \mu(01). \tag{3.1}$$

Si noti che la trasformazione è invertibile, per cui l'informazione contenuta nelle due successioni è la stessa. Pertanto, se $h(s)$ è l'entropia per simbolo di s,

$$h(s^0) = \frac{h(s^1)}{Z}.$$

Si consideri ora un esempio in cui la coppia da sostituire è formata da due simboli uguali. Nella stringa

$$0011010000101000100000110000 1$$

si esegua la sostituzione di 00 col simbolo 2. Si ottiene

$$2110122101201220112 21.$$

La differenza principale col caso precedente è che ora non vengono sostituite col nuovo simbolo *tutte* le coppie di zeri consecutivi. Per esempio $1001 \to 121$, ma $10001 \to 1201$. Si deduce facilmente che in questo caso la (3.1) diventa

$$\frac{1}{Z} = 1 - \mu(00) + \mu(000) - \mu(0000) + \mu(00000) - \ldots. \tag{3.2}$$

Questo esempio mostra che, tramite una sostituzione NSRPS, le probabilità delle stringhe possono comportarsi in maniera complicata. Tuttavia, le sostituzioni NSRPS trasformano una successione markoviana in una successione markoviana, come dimostrato da Grassberger in [Gr]. In generale, se la sorgente di partenza non è markoviana, il processo non diventa markoviano dopo un numero finito di trasformazioni. Tuttavia è ragionevole pensare che le successioni tendano a diventare markoviane man mano che il numero delle trasformazioni tende a infinito. Nelle prossime sezioni si procederà alla formalizzazione di quanto detto, all'enunciazione e dimostrazione dei risultati.

3.2 La trasformazione di stringhe e misure

3.2.1 Stringhe

Dato un alfabeto A, si denota con $A^* = \cup_{k=1}^{\infty} A^k$ l'insieme delle parole finite nell'alfabeto A. Gli elementi di A^* saranno indicati con lettere latine sottolineate $\underline{w}$, $\underline{z}$, etc. La medesima notazione verrà usata anche per parole infinite (elementi di $A^{\mathbb{N}}$ e $A^{\mathbb{Z}}$). Un elemento $\underline{w}$ ha lunghezza $|\underline{w}|$ e, se $|\underline{w}| = k$, esso sarà altresí indicato con $w_1^k = w_1 \dots w_k = (w_1, \dots, w_k)$.

Si considerino $x, y \in A$, $\alpha \notin A$ e $A' = A \cup \{\alpha\}$. Una *sostituzione di coppia* è una funzione $G = G_{xy}^{\alpha} : A^* \to A'^*$ che sostituisce in ordine le occorrenze di xy con α. Piú precisamente, $G\underline{w}$ è definito sostituendo in $\underline{w}$ la prima occorrenza da sinistra di xy con α, e ripetendo tale procedura fino al termine della stringa. Si definisce anche la funzione $S = S_{\alpha}^{xy} : A'^* \to A^*$, che agisce sulle parole $\underline{z} \in A'^*$ sostituendo ogni occorrenza del simbolo α con la coppia xy.

Si noti che la funzione G è iniettiva e non suriettiva mentre S è suriettiva e non iniettiva. Si noti inoltre che $S|_{G(A^*)} = G^{-1}$, cioè

$$S(G(\underline{w})) = \underline{w} \quad \text{per ogni } \underline{w} \in A^*. \tag{3.3}$$

Ci si rende conto facilmente che l'insieme delle parole ammissibili $G(A^*)$ è un sottoinsieme di A'^* che può essere descritto da vincoli su simboli consecutivi: nel caso $xy \to \alpha$, con $x \neq y$, $G(A^*)$ consiste delle stringhe di A'^* in cui non compare la coppia xy; nel caso $xx \to \alpha$, $G(A^*)$ consiste delle stringhe di A'^* in cui non compaiono le coppie xx e $x\alpha$. Un fatto importante è che dopo l'applicazione di piú sostituzioni, l'insieme delle parole ammissibili è ancora descrivibile mediante vincoli su simboli consecutivi. È il contenuto del prossimo teorema.

Teorema 3.2.1. *Sia $\{V_{a,b}\}_{a,b \in A}$ una matrice binaria (la matrice dei vincoli), e sia A_V^* il sottoinsieme di A^* i cui elementi $\underline{w}$ verificano*

$$\prod_{i=1}^{|\underline{w}|-1} V_{w_i, W_{i+1}} = 1$$

(A_V^* è l'insieme delle stringhe ammissibili rispetto ai vincoli su simboli consecutivi dati da V). Esiste una matrice di vincoli V' con indici in A' tale che

$$G(A_V^*) = A'^*_{V'}.$$

La dimostrazione segue da un controllo diretto. Qui di seguito verrà solo scritta V' in funzione di V.

Siano $z, w \in A \backslash \{x, y\}$. I valori degli elementi di V' sono dati dalle seguenti tabelle:

se $x \neq y$	x	y	w	α
x	$V_{x,x}$	0	$V_{x,w}$	$V_{x,x}$
y	$V_{y,x}$	$V_{y,y}$	$V_{y,w}$	$V_{y,x}$
z	$V_{z,x}$	$V_{z,y}$	$V_{z,w}$	$V_{z,x}$
α	$V_{y,x}$	$V_{y,y}$	$V_{y,w}$	$V_{y,x}$

se $x = y$	x	w	α
x	0	$V_{x,w}$	0
z	$V_{z,x}$	$V_{z,w}$	$V_{z,x}$
α	1	$V_{x,w}$	1

Si osservi che tali espressioni valgono se $V_{x,y} = 1$ e $V_{x,x} = 1$, rispettivamente; il caso contrario non è interessante e dà $G(A_V^*) = A_V^*$.

3.2.2 Misure

Qui e nel seguito si indicherà con $\mathcal{E}(A)$ l'insieme delle misure ergodiche stazionarie su $A^{\mathbb{Z}}$, le sole che verranno prese in considerazione. Se $\mu \in \mathcal{E}(A)$ verrà usata la notazione $\mu(\underline{w})$ per indicare il valore dei $|\underline{w}|$-marginali di μ sulla stringa $\underline{w}$. Ricordiamo che i k-marginali della misura μ di un processo stazionario costituito dalla successione di variabili aleatorie $\{X_i\}_{i\in\mathbb{N}}$ sono (vedi sezione 1.5.1) le misure secondo μ degli insiemi cilindrici della forma

$$[a_s, a_{s+1}, \ldots, a_{s+k-1}] = \bigcap_{i=0}^{k-1} \{X_{s+i} = a_{s+i}\}.$$

Il teorema e le proposizioni in questa sezione provengono dall'articolo [BCG]. Ne riportiamo le relative dimostrazioni nell'appendice A.

Le mappe G_{xy}^{α} e S_{α}^{xy} inducono in modo naturale le mappe $\mathcal{G} = \mathcal{G}_{xy}^{\alpha} : \mathcal{E}(A) \to \mathcal{E}(A')$ e $\mathcal{S} = \mathcal{S}_{\alpha}^{xy} : \mathcal{E}(A') \to \mathcal{E}(A)$. Siano $\mu \in \mathcal{E}(A)$ e $\underline{w} \in A^{\mathbb{N}}$ una successione tipica rispetto a μ; siano inoltre $\nu \in \mathcal{E}(A')$ e $\underline{z} \in A'^{\mathbb{N}}$ una successione tipica rispetto a ν. La successione $G\underline{w}$ è tipica per una misura ergodica che verrà chiamata $\mathcal{G}\mu$ e $S\underline{z}$ è tipica per una misura ergodica che verrà chiamata $\mathcal{S}\nu$.

Piú precisamente, indicando il numero di occorrenze di una sottoparola $\underline{s}$ in $\underline{r}$ con $\#\{\underline{s} \subseteq \underline{r}\} = \sum_{i=1}^{|\underline{r}|-|\underline{s}|+1} \mathbb{1}(r_i^{i+|\underline{s}|-1} = \underline{s})$, dove $\mathbb{1}$ è la funzione caratteristica, vale il seguente teorema.

Teorema 3.2.2. *Se $\underline{s} \in A'^*$, allora il limite*

$$\mathcal{G}\mu(\underline{s}) = \lim_{n\to\infty} \frac{\#\{\underline{s} \subseteq G(w_1^n)\}}{|G(w_1^n)|} \tag{3.4}$$

esiste ed è costante μ quasi ovunque in $\underline{w}$; inoltre $\{\mathcal{G}\mu(\underline{s})\}_{\underline{s}\in A'^*}$ sono i marginali di una misura ergodica su $A'^{\mathbb{Z}}$. Analogamente, se $\underline{r} \in A^*$, allora il limite

$$\mathcal{S}\nu(\underline{r}) = \lim_{n\to\infty} \frac{\#\{\underline{r} \subseteq S(z_1^n)\}}{|S(z_1^n)|} \tag{3.5}$$

esiste ed è costante ν quasi ovunque in $\underline{z}$; inoltre $\{\mathcal{S}\nu(\underline{r})\}_{\underline{r}\in A^*}$ sono i marginali di una misura ergodica su $A^{\mathbb{Z}}$. Vale

$$\mathcal{S}_{\alpha}^{xy}\mathcal{G}_{xy}^{\alpha}\mu = \mu. \tag{3.6}$$

Sia Z^{μ}_{xy} l'inverso dell'accorciamento medio rispetto a μ di una stringa sotto l'azione di G^{α}_{xy} e sia W^{ν}_{α} l'allungamento medio rispetto a ν di una stringa sotto l'azione di S^{xy}_{α}. Per queste quantità vale la seguente proposizione.

Proposizione 3.2.3. *Se $x \neq y$ allora*

$$Z^{\mu}_{xy} \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{n}{|G(w_1^n)|} = \frac{1}{1-\mu(xy)} \quad (\mu \text{ q. o. in } \underline{w}). \tag{3.7}$$

Se invece $x = y$ allora

$$Z^{\mu}_{xx} \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{n}{|G(w_1^n)|} = \frac{1}{1-\sum_{k=2}^{\infty}(-1)^k \mu(\underline{x}^k)} \quad (\mu \text{ q. o. in } \underline{w}), \tag{3.8}$$

dove $\underline{w}^k$ è la stringa formata da k simboli x. Per W^{ν}_{α} vale

$$W^{\nu}_{\alpha} \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{|S(z_1^n)|}{n} = 1+\nu(\alpha) \quad (\nu \text{ q. o. in } \underline{z}). \tag{3.9}$$

Inoltre,

$$W^{\mathcal{G}^{\alpha}_{xy}\mu}_{\alpha} = Z^{\mu}_{xy}. \tag{3.10}$$

Proposizione 3.2.4. *Se $\underline{r} \in A^*$, il valore di $\mathcal{S}\nu(\underline{r})$ dipende solo dai valori di $\nu(\underline{s})$ con $|\underline{s}| \leq |\underline{r}|$.*

Si osservi che la precedente affermazione è falsa per $\mathcal{G}\mu$, e, nel caso in cui $x = y$, $\mathcal{G}\mu(\underline{s})$ può coinvolgere la probabilità di infinite stringhe di lunghezza crescente.

Proposizione 3.2.5. *Se $\nu \in \mathcal{E}(A')$ rispetta i vincoli di coppia dati da $\mathcal{G}$, cioè per $\underline{z} \in A'^*$*

$$\nu(\underline{z}) = 0 \quad \text{se } \underline{z} \notin G(A^*),$$

allora

$$\nu = \mathcal{G}\mathcal{S}\nu.$$

3.3 La trasformazione delle entropie per simbolo

Dati $\mu \in \mathcal{E}(A)$ e $n \geq 1$, indicando con log il logaritmo in base 2,

$H_n(\mu) = -\sum_{\lvert\underline{z}\rvert=n} \mu(\underline{z}) \log \mu(\underline{z})$	è l'entropia di ordine n
$h_n(\mu) = H_{n+1}(\mu) - H_n(\mu)$	è l'entropia condizionata di ordine n
$h(\mu) = \lim_{n\to\infty} \frac{H_n(\mu)}{n} = \lim_{n\to\infty} h_n(\mu)$	è l'entropia di μ.

Ricordiamo che la misura μ si dice k-markoviana se, indicando le probabilità condizionate con $\mu(\underline{z}|\underline{w}) = \mu(\underline{wz})/\mu(\underline{w})$, si ha che per ogni $n > k$, $\underline{w} \in A^n$ e $a \in A$ vale $\mu(a|w_1^n) = \mu(a|w_{n-k+1}^n)$. In questo caso $h(\mu) = h_j(\mu)$ per ogni $j \geq k$. Osserviamo anche che $h(\mu) = h_k(\mu)$ implica che μ è una misura k-markoviana.

Diamo qui di seguito alcuni enunciati (vedi [BCG]) su come le entropie di trasformano sotto l'azione di $\mathcal{G}$. Le relative dimostrazioni si trovano nell'appendice A.

Per alleggerire un poco le notazioni, verranno usate le abbreviazioni $Z = Z^\mu_{xy}$ e talvolta $Z^\mu = Z^\mu_{xy}$ quando sarà necessario specificare la misura.

Teorema 3.3.1.

$$h(\mathcal{G}\mu) = Zh(\mu). \tag{3.11}$$

Il contenuto del precedente teorema non è una sorpresa, giacché la quantità di informazione della stringa $\underline{w}$ è la stessa della stringa $G(\underline{w})$.

Teorema 3.3.2.

$$h_1(\mathcal{G}\mu) \leq Zh_1(\mu). \tag{3.12}$$

Inoltre, se μ è una misura 1-markoviana allora anche $\mathcal{G}(\mu)$ è 1-markoviana.

Si noti come la seconda affermazione sia una conseguenza della prima: se μ è una misura 1-markoviana, allora

$$h(\mathcal{G}\mu) \leq h_1(\mathcal{G}\mu) \leq Zh_1(\mu) = Zh(\mu) = h(\mathcal{G}\mu). \tag{3.13}$$

Pertanto $h_1(\mathcal{G}\mu) = h(\mathcal{G}\mu)$ e ciò implica che $\mathcal{G}\mu$ è 1-markoviana.

Il prossimo teorema è una generalizzazione del teorema 3.3.2.

Teorema 3.3.3.

$$h_k(\mathcal{G}\mu) \leq Zh_k(\mu), \tag{3.14}$$

e $\mathcal{G}$ manda misure k-markoviane in misure k-markoviane.

3.4 Il teorema principale

Il teorema 3.3.2 dice che l'informazione contenuta in $G(\underline{w})$, che è pari a quella di $\underline{w}$ per il teorema 3.3.1, è maggiormente concentrata sulle coppie di simboli, rispetto alla stringa di partenza $\underline{w}$. Tale fatto suggerisce che una successione di sostituzioni di coppie possa trasferire tutta l'informazione nella distribuzione delle coppie di simboli. Per formalizzare questa asserzione definiamo ricorsivamente:

gli alfabeti $A_N = A_{N-1} \cup \{\alpha_N\}$ dove $\alpha_N \notin A_{N-1}$, con $A_0 = A$;

le mappe $G_N = G^{\alpha_N}_{x_N y_N} : A^*_{N-1} \to A^*_N$, dove $x_N, y_N \in A_{N-1}$;

le mappe corrispondenti $\mathcal{G}_N = \mathcal{G}^{\alpha_N}_{x_N y_N}$, $S_N = S^{x_N y_N}_{\alpha_N}$, $\mathcal{S}_N = \mathcal{S}^{x_N y_N}_{\alpha_N}$;

le misure $\mu_N = \mathcal{G}_N \mu_{N-1}$, con $\mu_0 = \mu$;

i fattori di normalizzazione $Z_N = Z^{\mu_{N-1}}_{x_N y_N}$;

le mappe composte

$$\begin{aligned} \overline{G}_N &= G_N \circ \ldots \circ G_1, & \overline{\mathcal{G}}_N &= \mathcal{G}_N \circ \ldots \circ \mathcal{G}_1, \\ \overline{S}_N &= S_1 \circ \ldots \circ S_N, & \overline{\mathcal{S}}_N &= \mathcal{S}_1 \circ \ldots \circ \mathcal{S}_N; \end{aligned}$$

le corrispondenti normalizzazioni $\overline{Z}_N = Z_N \ldots Z_1$ (si userà il simbolo $\overline{Z}^\mu_N$ quando sarà necessario specificare la misura iniziale).

Con le notazioni introdotte possiamo ora enunciare e dimostrare il teorema centrale (vedi Benedetto-Caglioti-Gabrielli [BCG]).

Teorema 3.4.1. *Se*

$$\lim_{N\to\infty} \overline{Z}_N = +\infty \tag{3.15}$$

allora

$$h(\mu) = \lim_{N\to\infty} \frac{h_1(\mu_N)}{\overline{Z}_N}. \tag{3.16}$$

Grassberger, in [Gr], sceglie ad ogni passo la coppia (o una delle coppie) di simboli con la massima frequenza di occorrenze non sovrapposte. Questo fatto assicura la divergenza di $\overline{Z}_N$, come si dimostrerà ora usando il teorema 3.3.2.

Teorema 3.4.2. *Se ad ogni passo N la coppia $x_N y_N$ è una coppia con la massima frequenza di occorrenze non sovrapposte tra tutte le coppie di simboli di A_{N-1}, allora*

$$\lim_{N\to\infty} \overline{Z}_N = +\infty. \tag{3.17}$$

Dimostrazione. Sia p_N il massimo della probabilità μ_{N-1} sulle coppie di simboli di A_{N-1}. Dalla definizione di Z_N segue che

$$\overline{Z}_N \geq \overline{Z}_{N-1} \left(1 + \frac{p_N}{2}\right),$$

(il fattore 2 compare per il caso di sostituzione di due simboli uguali). Si può stimare p_N con

$$p_N \geq 2^{-H_2(\mu_{N-1})},$$

dove $H_2(\mu_{N-1}) = -\sum_{a,b\in A_{N-1}} \mu_{N-1}(ab)\log\mu_{N-1}(ab)$ è l'entropia di ordine 2. Usando il teorema 3.3.2 e il fatto che $H_1(\mu_{N-1}) \le \log(N-1+|A|)$, con $|A|$ la cardinalità di A, si ha

$$H_2(\mu_{N-1}) = h_1(\mu_{N-1}) + H_1(\mu_{N-1}) \le \overline{Z}_{N-1}h_1(\mu) + \log(N-1+|A|).$$

Allora

$$\frac{\overline{Z}_N}{\overline{Z}_{N-1}} \ge 1 + \frac{2^{-\overline{Z}_{N-1}h_1(\mu)}}{2(N-1+|A|)}.$$

La successione $\overline{Z}_N$ è crescente: se per assurdo $\overline{Z}_N$ tendesse a una costante, dalla precedente equazione si avrebbe $\overline{Z}_N/\overline{Z}_{N-1} \ge 1 + c/(N-1)$, ma ciò implicherebbe $\overline{Z}_N \to \infty$. □

Dimostriamo ora il teorema centrale.
Dimostrazione del teorema 3.4.1. Per la composizione $\overline{S}_N$ vale

$$\overline{S}_N(s_1^n) = \overline{S}_N(s_1)\dots\overline{S}_N(s_n),$$

dove le $\overline{S}_N(s_i)$ sono parole nell'alfabeto originario A. Si consideri $\underline{r} \in A^*$, $|\underline{r}| = k$ e $\underline{s}$ una stringa tipica per μ_N.

$$\mu(\underline{r}) = \lim_{n\to\infty} \frac{\#\{\underline{r} \subseteq \overline{S}_N(s_1^n)\}}{|\overline{S}_N(s_1^n)|} = \lim_{n\to\infty} \frac{\#\{\underline{r} \subseteq \overline{S}_N(s_1)\dots\overline{S}_N(s_n)\}}{|\overline{S}_N(s_1^n)|}.$$

Si noti che

$$\begin{aligned}\#\{\underline{r} \subseteq \overline{S}_N(s_1)\dots\overline{S}_N(s_n)\} &= \sum_{g\in A_N} \#\{\underline{r} \subseteq \overline{S}_N(g)\}\#\{g \subseteq s_1^n\} \\ &+ \sum_{p=2}^{k} \sum_{g_1,\dots,g_p\in A_N} \#\{\underline{r} \frown \overline{S}_N(g_1)\dots\overline{S}_N(g_p)\}\#\{g_1\dots g_p \subseteq s_1^n\}\end{aligned}$$

dove $\#\{\underline{r} \frown \overline{S}_N(g_1)\dots\overline{S}_N(g_p)\}$ è il numero delle occorrenze di $\overline{r}$ nella stringa $\overline{S}_N(g_1)\dots\overline{S}_N(g_p)$ che inizia in $\overline{S}_N(g_1)$ e termina in $\overline{S}_N(g_p)$. Si ottiene

$$\begin{aligned}\mu(\underline{r}) = \lim_{n\to\infty} \frac{n}{|\overline{S}_N(s_1^n)|} &\Bigg(\sum_{g\in A_N} \#\{\underline{r} \subseteq \overline{S}_N(g)\} \frac{\#\{g \subseteq s_1^n\}}{n} \\ &+ \sum_{p=2}^{k} \sum_{g_1,\dots,g_p\in A_N} \#\{\underline{r} \frown \overline{S}_N(g_1)\dots\overline{S}_N(g_p)\} \frac{\#\{g_1\dots g_p \subseteq s_1^n\}}{n} \Bigg) \\ &= \frac{1}{\overline{Z}_N} \Bigg(\sum_{g\in A_N} \#\{\underline{r} \subseteq \overline{S}_N(g)\}\mu_N(g) \\ &+ \sum_{p=2}^{k} \sum_{g_1,\dots,g_p\in A_N} \#\{\underline{r} \frown \overline{S}_N(g_1)\dots\overline{S}_N(g_p)\}\mu_N(g_1\dots g_p) \Bigg).\end{aligned}$$

Sia $\mathcal{P}$ l'operatore di proiezione che mappa una misura μ nella sua approssimazione 1-markoviana $\mathcal{P}\mu$ e si definisca $\pi_N^j = \mathcal{S}_{j+1} \dots \mathcal{S}_N \mathcal{P}\mu_N$. In particolare si ha $\pi_N^0 = \overline{\mathcal{S}}_N \mathcal{P}\mu_N$ e $\pi_N^N = \mathcal{P}\mu_N$. Vale

$$\pi_N^N = \overline{\mathcal{G}}_N \pi_N^0. \tag{3.18}$$

Infatti le misure π_N^N e μ_N coincidono sulle coppie di simboli, e allora si ha $\pi_N^N(\underline{w}) = 0$ se $\underline{w} \notin \overline{G}_N(A^*)$, come segue dal teorema 3.2.1. Poiché $\overline{G}_N(A^*) \subseteq G_N(A^*_{N-1})$, si può applicare la proposizione 3.2.5 e ottenere

$$\pi_N^N = \mathcal{G}_N \pi_N^{N-1}.$$

Ora, anche π_N^{N-1} e μ_{N-1} coincidono sulle coppie di simboli (vedi proposizione 3.2.4) e allora si può iterare la procedura fino ad ottenere la (3.18).

Si noti che

$$\overline{Z}_N^{\pi_N^0} = \prod_{j=1}^{N} (1 + \pi_N^j(\alpha_j)) = \prod_{j=1}^{N} (1 + \mu_j(\alpha_j)) = \overline{Z}_N^{\mu}; \tag{3.19}$$

infatti π_N^j e μ_j coincidono sulle coppie di simboli di A_j. Pertanto, per ogni k e ogni $\underline{r}$ di lunghezza k, si ha

$$\begin{aligned} |\pi_N^0(\underline{r}) - \mu(\underline{r})| &\leq \\ &\leq \frac{1}{\overline{Z}_N} \sum_{p=3}^{k} \sum_{g_1,\dots,g_p \in A_N} \big(\mu_N + \pi_N\big)(g_1 \dots g_p) \# \{\underline{r} \frown \overline{S}_N(g_1) \dots \overline{S}_N(g_p)\} \\ &\leq 2 \frac{k^2}{\overline{Z}_N}, \end{aligned}$$

che tende a 0 per $N \to \infty$. Ciò implica che

$$\lim_{N \to \infty} h_k(\pi_N^0) = h_k(\mu).$$

In conclusione, per ogni k vale

$$h(\mu) = \frac{h(\mu_N)}{\overline{Z}_N} \leq \frac{h_1(\mu_N)}{\overline{Z}_N} = \frac{h(\pi_N^N)}{\overline{Z}_N} = h(\pi_N^0) \leq h_k(\pi_N^0).$$

Il terzo passo nella precedente catena di uguaglianze e disuguaglianze segue dalla definizione $\pi_N^N = \mathcal{P}\mu_N$, mentre il quarto passo segue dalle (3.18) e (3.19). Infine, prendendo i limiti per $N \to \infty$ e $k \to \infty$, si ha

$$h(\mu) = \lim_{N \to \infty} \frac{h_1(\mu_N)}{\overline{Z}_N},$$

cioè la tesi. $\square$

3.5 Un esempio interessante

La condizione (3.15), ipotesi del teorema centrale 3.4.1, è sufficiente ma non necessaria. In questa sezione analizzeremo nel dettaglio un esempio proposto da Benedetto, Caglioti, Gabrielli nell'articolo [BCG] (in cui gli autori elencano i risultati omettendo i calcoli). In tale esempio la suddetta condizione di divergenza del fattore di normalizzazione non è soddisfatta e tuttavia vale ancora la conclusione del teorema 3.4.1.

Nel corso di tutta questa sezione, indicheremo sempre con log il logaritmo in base 2.

L'alfabeto iniziale è $A = \{0, 1\}$. La prima sostituzione di coppia è $10 \to 2$, la seconda $20 \to 3$, e in generale l'N-esima sostituzione è $N0 \to N+1$. Si noti che la composizione infinita di tali sostituzioni corrisponde alla procedura di codifica che sostituisce blocchi massimali di k zeri consecutivi, e l'uno che li precede, con il nuovo simbolo $k+1$.

Se la misura iniziale assegna probabilità positiva alla coppia 11, allora il fattore di normalizzazione $\overline{Z}_N$ non può divergere. Infatti esso rappresenta l'inverso dell'accorciamento della lunghezza delle stringhe, ma per una tipica stringa iniziale di lunghezza L dopo le trasformazioni restano circa $\mu(11)L$ simboli. Si ha pertanto che per N grande

$$\overline{Z}_N \approx \frac{L}{\mu(11)L} = \frac{1}{\mu(11)} < \infty.$$

Si noti che solo la prima sostituzione coinvolge il simbolo 1. Pertanto,

$$\mu_N(1|1) = \mu_1(1|1) = \frac{\mu(11) - \mu(110)}{\mu(1) - \mu(10)} = \mu(1|11),$$
$$\mu_N(1|11) = \mu_1(1|11) = \frac{\mu(111) - \mu(1110)}{\mu(11) - \mu(110)} = \mu(1|111).$$

Se per la misura iniziale $\mu(1|111) \neq \mu(1|11)$, allora $\mu_N(1|11) \neq \mu_N(1|1)$ per ogni N e $h_1(\mu_N)/\overline{Z}_N$ non può convergere a $h(\mu)$ (il processo limite non può essere 1-markoviano).

Consideriamo ora come misura iniziale quella di un processo di rinnovamento (*renewal process*) a media finita, ovvero un processo stazionario per il quale le distanze tra simboli '1' consecutivi sono variabili aleatorie indipendenti e identicamente distribuite (i.i.d.r.v.) con distribuzione $\{p_k\}_{k\geq 1}$, dove i p_k soddisfano $\sum_k p_k = 1$ e $E^0 = \sum_{k=1}^{\infty} kp_k < \infty$. Per una trattazione dettagliata dei processi di rinnovamento discreti, il lettore è invitato a consultare il testo di Feller [Fe].

L'entropia di tale processo si può calcolare nel modo seguente. Essa rappresenta (vedi capitolo 2) il numero di *bit* per carattere che servono per codificare il processo binario. La quantità $C = -\sum_k p_k \log p_k$ invece rappresenta il numero di *bit* per numero di salti tra simboli '1' consecutivi che servono per codificare il processo dei salti. In altre parole, la quantità C è l'entropia di una variabile aleatoria che descrive la lunghezza di un salto tra due simboli '1' consecutivi. Il legame tra queste due quantità si può vedere mediante le seguenti considerazioni. Con n salti si descrive una stringa di circa nE^0 caratteri, essendo E^0 la lunghezza media dei salti. Pertanto, per n tendente all'infinito, con nC *bit* si descrivono nE^0 caratteri e quindi

$$h(\mu) \overset{n\to\infty}{\approx} \frac{nC}{nE^0} = \frac{-\sum_k p_k \log p_k}{E^0}. \tag{3.20}$$

Per un calcolo esplicito dei marginali di μ_N abbiamo usato iterativamente le formule (3.21) qui sotto, le quali esprimono i valori marginali della misura trasformata da una sostituzione di coppie $xy \to \alpha$, con $x \neq y$, in funzione dei valori marginali della misura di partenza e dei fattori di normalizzazione dati dalla (3.22). Per la loro dimostrazione si consulti l'appendice A, equazioni (A.8) e (A.9).

Se $\underline{s}$ e $\underline{r}$ sono due stringhe finite tali che $\underline{s} = G^{\alpha}_{xy}(\underline{r})$ con $|\underline{s}| = n$ e $|\underline{r}| = k$, l'espressione della misura trasformata $\nu(\underline{s}) = \mathcal{G}^{\alpha}_{xy}\mu(\underline{s})$ è data da una delle seguenti:

$$\begin{aligned}
&s_1 \neq y,\ s_n \neq x, \quad \nu(\underline{s}) = Z\mu(\underline{r}) \\
&s_1 = y,\ s_n \neq x, \quad \nu(\underline{s}) = Z(\mu(\underline{r}) - \mu(xyr_2^k)) \\
&s_1 \neq y,\ s_n = x, \quad \nu(\underline{s}) = Z(\mu(\underline{r}) - \mu(r_1^{k-1}xy)) \\
&s_1 = y,\ s_n = x, \quad \nu(\underline{s}) = Z(\mu(\underline{r}) + \mu(xyr_2^{k-1}xy) - \mu(xyr_2^k) - \mu(r_1^{k-1}xy)).
\end{aligned} \tag{3.21}$$

Il fattore di normalizzazione $Z = Z^{\mu}_{xy}$ è dato in termini di μ da

$$Z = 1 + \nu(\alpha) = 1 + Z\mu(xy) = \frac{1}{1-\mu(xy)}. \tag{3.22}$$

Spiegando le (3.21) a parole, esse tengono conto del fatto che, essendo $\underline{s}$ e $\underline{r}$ stringhe finite facenti parte di successioni infinite, possono esserci degli effetti di bordo all'inizio e alla fine delle stringhe stesse. Pur essendo $S(*\underline{s}*) = *\underline{r}*$, potrebbe valere $\mu(\underline{r}) \neq \mathcal{G}\mu(\underline{s})$, e ciò nel caso in cui $\underline{s}$ abbia il primo o l'ultimo simbolo coinvolti nella sostituzione data.

Esemplificando secondo le convenzioni di questa sezione, mentre le stringhe che dopo la prima sostituzione ($10 \to 2$) iniziano con 2 derivano da tutte e sole quelle che iniziano con 10, quelle che dopo la sostituzione iniziano con

0 derivano da tutte e sole quelle che iniziano con 00, non 0, perché ovviamente se lo 0 fosse stato preceduto da un 1, con la trasformazione sarebbe scomparso.

Grazie alle (3.21) si ha

$$\begin{array}{lcccccc} \text{dopo } N \text{ sostituzioni} & 0 & 1 & 2 & \cdots & N & N+1 \\ & \updownarrow & \updownarrow & \updownarrow & \cdots & \updownarrow & \updownarrow \\ \text{dopo 0 sostituzioni} & 0^{N+1} & 11 & 101 & \cdots & 10^{N-1}1 & 10^N \end{array} \tag{3.23}$$

dove con 0^k si intende una stringa di k zeri consecutivi. Perciò, i valori marginali di μ_N sui simboli singoli sono uguali, a meno del fattore di normalizzazione $\overline{Z}_N$ da determinare, ai valori della misura iniziale μ_0 sulla stringa corrispondente.

Calcoliamo preliminarmente, perché utili nel seguito, $\mu_0(1)$ e $\mu_0(0^k)$. Il valore di $\mu_0(1)$ si può calcolare trovando la probabilità che ad un dato istante ci si trovi in un salto di lunghezza k, dividendo per k perché in tale salto ci sono $k-1$ zeri e 1 uno, e sommando su tutti i k:

$$\mu_0(1) = \sum_{k\geq 1} \frac{1}{k} \frac{kp_k}{\sum_{j\geq 1} jp_j} = \frac{1}{E^0}.$$

Ragionando in modo del tutto analogo si trova

$$\mu_0(0^k) = \frac{\sum_{j\geq 1} jp_{k+j}}{E^0}.$$

Si ha cosí:

$$\begin{aligned} \mu_N(0) &= \overline{Z}_N \mu_0(0^{N+1}) = \overline{Z}_N \frac{\sum_{j\geq 1} jp_{k+j}}{E^0} \\ \mu_N(1) &= \overline{Z}_N \mu_0(11) = \overline{Z}_N \frac{p_1}{E^0} \\ \mu_N(2) &= \overline{Z}_N \mu_0(101) = \overline{Z}_N \frac{p_2}{E^0} \\ &\vdots \\ \mu_N(N) &= \overline{Z}_N \mu_0(10^{N-1}1) = \overline{Z}_N \frac{p_N}{E^0} \\ \mu_N(N+1) &= \overline{Z}_N \mu_0(10^N) = \overline{Z}_N \frac{p_{N+1} + p_{N+2} + \ldots}{E^0} \end{aligned} \tag{3.24}$$

Poiché deve essere $\sum_{k=0}^{N+1} \mu_N(k) = 1$, si può trovare un valore esplicito del fattore di normalizzazione:

$$\overline{Z}_N = \frac{E^0}{p_1 + \ldots + p_N + p_{N+1} + 2p_{N+2} + 3p_{N+3} + \ldots},$$

che si può riscrivere come

$$\overline{Z}_N = \frac{E^0}{E^N}, \tag{3.25}$$

ponendo $E^N = \sum_{j=1}^{\infty} j p_j^N$ e

$$p_j^N = \begin{cases} p_1 + \ldots + p_{N+1} & j = 1 \\ p_{N+j} & j > 1 \end{cases} .$$

Il ragionamento per determinare i marginali di μ_N sulle coppie di simboli è del tutto analogo e i risultati sono riportati nelle seguenti tabelle. Nella prima, M_{ij} è la stringa iniziale (dopo 0 sostituzioni) corrispondente alla coppia di simboli ij che si ha dopo N sostituzioni.

$$M = \begin{array}{c|cccccc}
 & 0 & 1 & 2 & \ldots & N & N+1 \\
\hline
0 & 0^{N+2} & 0^{N+1}11 & 0^{N+1}101 & \ldots & 0^{N+1}10^{N-1}1 & 0^{N+1}10^N \\
1 & \varnothing & 111 & 1101 & \ldots & 110^{N-1}1 & 110^N \\
2 & \varnothing & 1011 & 10101 & \ldots & 1010^{N-1}1 & 1010^N \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
N & \varnothing & 10^{N-1}11 & 10^{N-1}101 & \ldots & 10^{N-1}10^{N-1}1 & 10^{N-1}10^N \\
N+1 & 10^{N+1} & 10^N11 & 10^N101 & \ldots & 10^N10^{N-1}1 & 10^N10^N
\end{array}$$

Nella prossima tabella sono riportati i valori marginali di μ_N sulle coppie di simboli. Il valore di posto (i, j) è $V_{ij} = \mu_N(ij) = \overline{Z}_N \mu_0(M_{ij})$.

$$V = \begin{array}{c|cccccc}
 & 0 & 1 & 2 & \ldots & N & N+1 \\
\hline
0 & \frac{\sum_{i\geq 1} i p_{N+2+i}}{E^N} & \frac{p_1 p_{N+2}^+}{E^N} & \frac{p_2 p_{N+2}^+}{E^N} & \ldots & \frac{p_N p_{N+2}^+}{E^N} & \frac{p_{N+1}^+ p_{N+2}^+}{E^N} \\
1 & 0 & \frac{p_1 p_1}{E^N} & \frac{p_1 p_2}{E^N} & \ldots & \frac{p_1 p_N}{E^N} & \frac{p_{N+1}^+ p_1}{E^N} \\
2 & 0 & \frac{p_2 p_1}{E^N} & \frac{p_2 p_2}{E^N} & \ldots & \frac{p_2 p_N}{E^N} & \frac{p_{N+1}^+ p_2}{E^N} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
N & 0 & \frac{p_N p_1}{E^N} & \frac{p_N p_2}{E^N} & \ldots & \frac{p_N p_N}{E^N} & \frac{p_{N+1}^+ p_N}{E^N} \\
N+1 & \frac{p_{N+2}^+}{E^N} & \frac{p_{N+1} p_1}{E^N} & \frac{p_{N+1} p_2}{E^N} & \ldots & \frac{p_{N+1} p_N}{E^N} & \frac{p_{N+1}^+ p_{N+1}}{E^N}
\end{array}$$

dove si è usata la notazione $p_k^+ = \sum_{i\geq k} p_i$.

Osservazione 3.5.1. *In particolare vale*

$$\mu_{N-1}(N0) = \overline{Z}_{N-1}\mu_0(10^N) = \frac{E^0}{E^{N-1}} \frac{\sum_{i\geq N+1} p_i}{E^0} = \frac{\sum_{i\geq N+1} p_i}{E^{N-1}},$$

da cui si ricava

$$Z_N = \frac{1}{1-\mu_{N-1}(N0)} = \frac{E^{N-1}}{E^{N-1} - \sum_{i\geq N+1} p_i} = \frac{E^{N-1}}{E^N},$$

in accordo con la (3.25).

Usando i risultati raccolti nelle (3.24), otteniamo

$$\begin{aligned} \frac{H_1(\mu_N)}{\overline{Z}_N} = -\frac{E^N}{E^0}\Bigg[& \frac{\sum_{i\geq 1} i p_{n+1+i}}{E^0} \log \frac{\sum_{i\geq 1} i p_{n+1+i}}{E^N} \\ & - \sum_{j=1}^{N} \frac{p_j}{E^0} \log \frac{p_j}{E^N} - \frac{\sum_{i\geq 1} p_{n+i}}{E^0} \log \frac{\sum_{i\geq 1} p_{n+i}}{E^N} \Bigg]. \end{aligned} \tag{3.26}$$

Ricordando che $\lim_{x\to 0^+} x \log x = 0$ e osservando che, per N tendente all'infinito, $\sum_{i\geq 1} p_{n+i} \to 0$, $\sum_{i\geq 1} i p_{n+1+i} \to 0$ e $E^N \to 1$, otteniamo infine

$$\lim_{N\to\infty} \frac{H_1(\mu_N)}{\overline{Z}_N} = \frac{-\sum_{j\geq 1} p_j \log p_j}{E^0}. \tag{3.27}$$

Dai risultati della tabella V otteniamo, utilizzando la notazione $p_k^+ = \sum_{i\geq k} p_i$,

$$\begin{aligned} \frac{H_2(\mu_N)}{\overline{Z}_N} = \frac{E^N}{E^0}\Bigg[& \sum_{i=1}^{N+1} \left(\sum_{j=1}^{N} \left(\frac{p_i p_j}{E^N} \log \frac{p_i p_j}{E^N} \right) + \frac{p_i p_{N+1}^+}{E^N} \log \frac{p_i}{E^N} \right) \\ & - \frac{p_{N+2}^+}{E^N} \log \frac{p_{N+2}^+}{E^N} - \frac{\sum_{k\geq 1} k p_{N+2+k}}{E^N} \log \frac{\sum_{k\geq 1} k p_{N+2+k}}{E^N} \\ & - \sum_{j=1}^{N} \frac{p_j p_{N+2}^+}{E^N} \log \frac{p_j p_{N+2}^+}{E^N} - \frac{p_{N+2}^+ p_{N+1}^+}{E^N} \log \frac{p_{N+2}^+ p_{N+1}^+}{E^N} \Bigg]. \end{aligned} \tag{3.28}$$

In tale somma tutti i termini hanno limite nullo per N tendente all'infinito, tranne $\sum_{i=1}^{N+1} \sum_{j=1}^{N} \frac{p_i p_j}{E^N} \log \frac{p_i p_j}{E^N}$, e da semplici calcoli otteniamo

$$\lim_{N\to\infty} \frac{H_2(\mu_N)}{\overline{Z}_N} = \frac{-2\sum_{j\geq 1} p_j \log p_j}{E^0}. \tag{3.29}$$

Pertanto, ricordando che $h_1 = H_2 - H_1$ e osservando che in questo caso si ha $H_2(\mu_N) = 2H_1(\mu_N)$ nel limite per N che va a infinito, per il particolare processo di questo esempio vale

$$\lim_{N\to\infty} \frac{h_1(\mu_N)}{\overline{Z}_N} = \lim_{N\to\infty} \frac{H_1(\mu_N)}{\overline{Z}_N} = h(\mu), \tag{3.30}$$

cioè non solo il processo diventa 1-markoviano $h_1(\mu_N)/\overline{Z}_N \to h(\mu)$ nel limite per il numero di sostituzioni che va all'infinito, ma esso diventa anche indipendente $H_1(\mu_N)/\overline{Z}_N \to h(\mu)$. Tale fatto è ovviamente conseguenza della scelta molto particolare della misura iniziale.

Quasto esempio fa sorgere una domanda naturale. Mentre il teorema centrale 3.4.1 dimostra che, sotto la condizione (3.15) di divergenza del fattore di normalizzazione, ogni processo ergodico diventa asintoticamente 1-markoviano, ci si può chiedere quali siano le condizioni sul processo e sulle sostituzioni affinché il processo limite diventi addirittura indipendente. Tuttavia, tale quesito non sembra essere di facile soluzione.

Capitolo 4

Metodi di stima dell'entropia

Nel capitolo 5 guarderemo il metodo delle sostituzioni di Grassberger da un punto di vista computazionale e confronteremo i risultati di tale metodo con altri tre metodi classici per la stima empirica dell'entropia. In questo capitolo presentiamo i risultati teorici sui quali tali tre metodi sono basati.

Il metodo delle frequenze empiriche si basa sul teorema 4.1.2, il quale formalizza il tentativo piú intuitivo, cioè quello di applicare direttamente la definizione di entropia stessa.

Il secondo metodo è quello dei tempi di ritorno, basato sul teorema 4.2.2, il quale mette in relazione l'entropia con la crescita esponenziale in n del tempo che passa tra due occorrenze consecutive di un blocco simbolico di lunghezza n.

Il terzo metodo si basa sull'uguaglianza, per certe classi di trasformazioni, tra entropia e esponente di Lyapunov (teorema 4.3.4).

In tutto il capitolo indicheremo con log il logaritmo in base 2.

4.1 Frequenze empiriche

Data una successione campione $x_1, x_2, \ldots, x_n$ estratta da un processo ergodico μ non conosciuto, si vuole stimare l'entropia h di μ. Una procedura semplice è quella di determinare la distribuzione empirica q_k dei k-blocchi non sovrapposti e prendere $H(q_k)/k$ come stima di h. Se k è fissato e $n \to \infty$ allora $H(q_k)/k$ converge quasi certamente a $H(\mu_k)/k$, quantità che converge a h per $k \to \infty$. Quindi esiste una scelta di $k = k(n)$ come funzione di n per la quale la stima converge quasi certamente a h. A prima vista, però, la scelta di $k(n)$ potrebbe dipendere strettamente dalla misura μ e non c'è motivo di ritenere *a priori* che esista una scelta universale di $k = k(n)$, con $k(n) \to \infty$, per la quale la distribuzione empirica è vicina alla vera distribuzione μ_k *per* *ogni* processo ergodico μ. Tuttavia, il prossimo teorema dovuto a Ornstein e Weiss [OW] (per la cui dimostrazione rimandiamo al testo di Shields [Sh]) implica proprio l'esistenza di una tale scelta universale.

Data la successione $x_1, x_2, \ldots, x_n$, la distribuzione empirica dei k-blocchi non sovrapposti è definita in modo naturale da

$$q_k(a_1^k|x_1^n) = \frac{|\{i \in [0, m-1] : x_{ik+1}^{ik+k} = a_1^k\}|}{m},$$

dove $n = km + r$, $0 \leq r < k$.

Teorema 4.1.1 (Teorema dell'entropia empirica)**.** *Sia μ una misura ergodica con entropia $h > 0$. Per ogni $\epsilon > 0$ e ogni k esiste un insieme $\mathcal{T}_k(\epsilon) \subseteq A^k$ con $|\mathcal{T}_k(\epsilon)| \leq 2^{k(h+\epsilon)}$, tale che per quasi ogni x esiste un $K = K(\epsilon, x)$ tale che se $k \geq K$ e $n \geq 2^{kh}$, allora*

(a) $q_k(\mathcal{T}_k(\epsilon)|x_1^n) > 1 - \epsilon$,

(b) $q_k(B|x_1^n) < \epsilon$, *per ogni* $B \subseteq A^k$ *con* $|B| \leq 2^{k(h-\epsilon)}$.

Da questo teorema discende il seguente.

Teorema 4.1.2. *Se μ è una misura ergodica di entropia $h > 0$, se $k(n) \to \infty$ per $n \to \infty$ e se $k(n) \leq (1/h) \log n$, allora*

$$\lim_{n\to\infty} \frac{1}{k(n)} H(q_{k(n)}(\cdot|x_1^n)) = h, \quad q.\ c. \tag{4.1}$$

Dimostrazione. Si fissi $\epsilon > 0$ e sia $\{\mathcal{T}_k(\epsilon) \subseteq A^k : n \geq 1\}$ un insieme che soddisfi la parte (a) del teorema 4.1.1. Sia U_k l'insieme di tutte le stringhe $a_1^k \in \mathcal{T}_k(\epsilon)$ per le quali $q_k(a_1^k|x_1^n) < 2^{-k(h+2\epsilon)}$, cosí che per tutti i $k \geq K$ e tutti gli $n \geq 2^{kh}$,

$$q_k(U_k|x_1^n) \leq |\mathcal{T}_k(\epsilon)| \cdot 2^{-k(h+2\epsilon)} \leq 2^{k(h+\epsilon)} 2^{-k(h+2\epsilon)} = 2^{-\epsilon k}. \tag{4.2}$$

Sia ora V_k l'insieme di tutte le stringhe a_1^k per le quali $q_k(a_1^k|x_1^n) > 2^{-k(h-2\epsilon)}$, e si osservi che $|V_k| \leq 2^{k(h-2\epsilon)}$. La parte (b) del teorema 4.1.1 implica che per quasi ogni x esiste un $K(x)$ tale che

$$q_k(V_k|x_1^n) \leq \epsilon, \quad \text{per } k \geq K(x), n \geq 2^{kh}. \tag{4.3}$$

Dalle disuguaglianze (4.1) e (4.2) e dalla parte (a) del teorema 4.1.1 segue che per quasi ogni x esiste $K_1(x)$ tale che

$$q_k(G_k|x_1^n) \geq 1 - 3\epsilon, \quad \text{per } k \geq K_1(x), n \geq 2^{kh},$$

dove $G_k = \mathcal{T}_k(\epsilon) \setminus (U_k \cup V_k)$. Quindi, l'insieme degli a_1^k per cui

$$2^{-k(h+2\epsilon)} \leq q_k(a_1^k|x_1^n) \leq 2^{-k(h-2\epsilon)},$$

o equivalentemente

$$h - 2\epsilon \leq -\frac{1}{k} \log q_k(a_1^k|x_1^n) \leq h + 2\epsilon, \tag{4.4}$$

misura secondo $q_k(\cdot|x_1^n)$ almeno $1-3\epsilon$. Moltiplicando nella (4.4) per la quantità $q_k(a_1^k|x_1^n)/\mu(G_n)$ e sommando su G_n si ottiene

$$h - 2\epsilon \leq -\frac{1}{k\mu(G_n)} \sum_{G_n} q_k(a_1^k|x_1^n) \log q_k(a_1^k|x_1^n) \leq h + 2\epsilon.$$

Poiché per $n \to \infty$ la misura di G_n tende a 1, si ha infine

$$h - 2\epsilon \leq H(q_{k(n)}(\cdot|x_1^n)) \leq h + 2\epsilon,$$

cioè la tesi. □

4.2 Tempi di ritorno

Wyner e Ziv hanno scoperto una connessione interessante fra entropia e i cosiddetti tempi di ritorno per processi ergodici. Essi hanno mostrato in [WZ] che il logaritmo del tempo di attesa perché si riosservino i primi n termini di una successione x è asintotico a nh in probabilità.

Presentiamo ora la formalizzazione di ciò e un'idea di dimostrazione.

Definizione 4.2.1. *Data una successione x estratta da un processo ergodico, si definisce l'n-esimo tempo di ritorno come*

$$R_n(x) = \min\{m \geq 1 : \, x_{m+1}^{m+n} = x_1^n\}.$$

Il teorema che lega i tempi di ritorno all'entropia è il seguente.

Teorema 4.2.2. *Se μ è un processo ergodico con entropia h, allora vale*

$$\lim_{n\to\infty} \frac{1}{n} \log R_n(x) = h, \quad q.\ c.$$

Dimostrazione. Definiamo i limiti superiore e inferiore,

$$\overline{r}(x) = \limsup_{n\to\infty} \frac{1}{n} \log R_n(x),$$

$$\underline{r}(x) = \liminf_{n\to\infty} \frac{1}{n} \log R_n(x).$$

Poiché $R_{n-1}(Tx) \leq R_n(x)$, i limiti superiore e inferiore sono entrambi subinvarianti, cioè

$$\overline{r}(Tx) \leq \overline{r}, \text{ e } \underline{r}(Tx) \leq \underline{r},$$

e perciò sono entrambi invarianti quasi certamente. Poiché la misura è ergodica per ipotesi, i limiti superiore e inferiore sono dunque costanti quasi ovunque, e quindi esistono due costanti $\overline{r}$ e $\underline{r}$ tali che

$$\overline{r}(x) = \overline{r} \text{ e } \underline{r}(x) = \underline{r}, \quad \text{q. o.}$$

Inoltre $\underline{r} \leq \overline{r}$, per cui la tesi si può dimostrare provando le due disuguaglianze $\underline{r} \geq h$ e $\overline{r} \leq h$.

Per la prima disuguaglianza proponiamo qui una dimostrazione dovuta a Wyner e Ziv. Per la seconda la dimostrazione è piú complicata e richiede nozioni che non approfondiamo in questa tesi, come i codici fedeli e la relazione tra il rapporto di compressione di un codice e l'entropia. Ci limiteremo a indicare per sommi capi la linea della dimostrazione, invitando a consultare il testo di Shields [Sh] per una presentazione completa.

Per stabilire $\overline{r} \leq h$ si fissi $\epsilon > 0$ e si definiscano gli insiemi

$$D_n = \{x : R_n(x) > 2^{n(h+\epsilon)}\}, \quad n \geq 1.$$

L'obiettivo è dimostrare che, per quasi ogni x, $x \notin D_n$ definitivamente. È sufficiente dimostrare ciò sotto l'ipotesi ulteriore che x sia una successione tipica per l'entropia. Infatti, se

$$\mathcal{T}_n = \{x : \mu(x_1^n) \geq 2^{-n(h+\epsilon/2)}\}$$

allora per quasi ogni x si ha $x \in \mathcal{T}_n$ definitivamente, per cui è sufficiente dimostrare che

$$x \notin D_n \cap \mathcal{T}_n, \quad \text{definitivamente, q. c.} \tag{4.5}$$

Per dimostrare ciò si fissi a_1^n e si considerino solo gli $x \in D_n$ per i quali $x_1^n = a_1^n$, cioè l'insieme $D_n(a_1^n) = D_n \cap [a_1^n]$. Si noti che, se $x \in D_n(a_1^n)$, allora $T^j x \notin [a_1^n]$ per $j \in [1, 2^{n(h+\epsilon)} - 1]$, per la definizione di $D_n(a_1^n)$, e pertanto gli insiemi

$$D_n(a_1^n), T^{-1}D_n(a_1^n), \ldots, T^{-2^{n(h+\epsilon)}+1}D_n(a_1^n)$$

sono disgiunti. Poiché tali insiemi hanno tutti la stessa misura, il fatto che siano disgiunti implica che

$$\mu(D_n(a_1^n)) \leq 2^{-n(h+\epsilon)}. \tag{4.6}$$

È ora semplice arrivare alla (4.5) dalla (4.6). Infatti, la cardinalità della proiezione di $D_n \cap \mathcal{T}_n$ su A^n è limitata superiormente dalla cardinalità di $\mathcal{T}_n$, la quale è limitata superiormente da $2^{n(h+\epsilon/2)}$, cosí che dalla (4.6) segue

$$\mu(D_n \cap \mathcal{T}_n) \leq 2^{n(h+\epsilon/2)} 2^{-n(h+\epsilon)} = 2^{-n\epsilon/2}.$$

Per il principio di Borel-Cantelli si ha infine che per quasi ogni x, definitivamente $x \notin D_n \cap \mathcal{T}_n$.

La disuguaglianza $\underline{r} \geq h$ si può dimostrare facendo vedere che, se non fosse vera, si potrebbe costruire un codice con un rapporto di compressione superiore all'entropia. L'esistenza di un tale codice non è però possibile (vedi Shields [Sh], p. 121 ss.) □

4.3 Esponente di Lyapunov

Nel teorema 4.3.4 si dimostrerà che l'entropia di una trasformazione C^1 a tratti T definita su un intervallo, che abbia una misura di probabilità invariante assolutamente continua, può essere calcolata prendendo la media di $\log |T'|$. Tale media è chiamata *esponente di Lyapunov* di T e misura la velocità di allontanamento esponenziale di due punti vicini sotto l'azione della trasformazione T.

Siano $X = [0, 1]$, $T : X \to X$ un'applicazione C^1 a tratti e μ una misura di probabilità T-invariante.

Definizione 4.3.1. *La quantità* $\int_0^1 \log |T'| \mathrm{d}\mu$ *è detta* esponente di Lyapunov *di* T.

Sebbene la conclusione del teorema 4.3.4 si pensa valga in casi molto generali, noi ne presentiamo qui la dimostrazione sotto le seguenti ipotesi restrittive (vedi Choe [Ch]). Qui e nel seguito della sezione si assumerà che valga $\mathrm{d}\mu = \rho(x)\mathrm{d}x$ con $A \leq \rho(x) \leq B$ per ogni $x \in [0, 1]$, per due costanti A, $B > 0$. Sia $\mathcal{P}$ una partizione generatrice di X, costituita da intervalli E_j con $\mu(E_j) > 0$ e $\sum_j \mu(E_j) = 1$. Sia

$$\mathcal{P}_n = \mathcal{P} \vee T^{-1}\mathcal{P} \vee \ldots \vee T^{-(n-1)}\mathcal{P},$$

cioè la partizione negli insiemi cilindrici

$$[i_1, i_2, \ldots, i_n] = E_{i_1} \cap T^{-1}E_{i_2} \cap \ldots \cap T^{-(n-1)}E_{i_n}. \tag{4.7}$$

Si assuma che esista un N tale che se $n \geq N$ allora (i) ogni insieme cilindrico della forma (4.7) è un intervallo e (ii) T è monotona e differenziabile con continuità su ogni insieme cilindrico in $\mathcal{P}_n$.

Lemma 4.3.2. *Sia λ la misura di Lebesgue su $[0,1]$. Allora*

$$\lambda([i_2,\dots,i_n]) = |T'(z_{i_1,\dots,i_n})|\lambda([i_1,\dots,i_n]),$$

per qualche $z_{i_1,\dots,i_n} \in [i_1,\dots,i_n]$ per n sufficientemente grande.

Dimostrazione. Poiché $T([i_1,\dots,i_n]) = [i_2,\dots,i_n]$ e T è iniettiva sull'insieme cilindrico $[i_1,\dots,i_n]$, vale

$$\lambda([i_2,\dots,i_n]) = \left|\int_{[i_1,\dots,i_n]} T'(x)\mathrm{d}x\right| = |T'(z_{i_1,\dots,i_n})|\lambda([i_1,\dots,i_n])$$

per qualche $z_{i_1,\dots,i_n} \in [i_1,\dots,i_n]$. □

Lemma 4.3.3. *Si supponga che una successione $\{b_n\}_{n=1}^{\infty}$ soddisfi*

$$\lim_{n\to\infty}(b_{n+1} - b_n) = C$$

per qualche costante C. Allora

$$\lim_{n\to\infty}\frac{b_n}{n} = C.$$

Dimostrazione. Fissato $\epsilon > 0$, si scelga N tale che se $n \geq N$ allora

$$|b_{n+1} - b_n - C| < \epsilon.$$

Allora

$$|b_{N+k} - b_N - kC| \leq \sum_{j=1}^{k} |b_{N+j} - b_{N+j-1} - C| < k\epsilon$$

e

$$\left|\frac{b_{N+k}}{N+k} - \frac{b_N}{N+k} - \frac{k}{N+k}C\right| < \frac{k}{N+k}\epsilon.$$

Mandando k all'infinito si ottiene

$$C - \epsilon \leq \liminf_{n\to\infty}\frac{b_n}{n} \leq \limsup_{n\to\infty}\frac{b_n}{n} \leq C + \epsilon.$$

Poiché ϵ è arbitrario si ha la tesi. □

Teorema 4.3.4. *Sia h l'entropia di T. Con le condizioni su T sopra elencate si ha*

$$h = \int_0^1 \log|T'(x)|\mathrm{d}\mu.$$

Dimostrazione. Si ponga

$$b_n = \sum_{i_1,\dots,i_n} \mu([i_1,\dots,i_n]) \log \lambda([i_1,\dots,i_n])$$

e

$$L_n = \sum_{i_1,\dots,i_n} \mu([i_1,\dots,i_n]) \log |T'(z_{i_1,\dots,i_n})|.$$

Il lemma 4.3.2 implica che

$$\begin{aligned} b_n &= \sum_{i_1,\dots,i_n} \mu([i_1,\dots,i_n]) \log \lambda([i_2,\dots,i_n]) - L_n \\ &= \sum_{i_2,\dots,i_n} \mu([i_2,\dots,i_n]) \log \lambda([i_1,\dots,i_n]) - L_n \\ &= b_{n-1} - L_n. \end{aligned}$$

Allora

$$\lim_{n\to\infty} (b_n - b_{n-1}) = -\lim_{n\to\infty} L_n = -\int_0^1 \log |T'| \mathrm{d}\mu.$$

Per il lemma 4.3.3 si ha

$$\lim_{n\to\infty} \frac{b_n}{n} = -\int_0^1 \log |T'| \mathrm{d}\mu$$

e allora

$$\begin{aligned} h &= -\lim_{n\to\infty} \frac{1}{n} \sum_{i_1,\dots,i_n} \mu([i_1,\dots,i_n]) \log \mu([i_1,\dots,i_n]) \\ &= -\lim_{n\to\infty} \frac{1}{n} \sum_{i_1,\dots,i_n} \mu([i_1,\dots,i_n]) \log \lambda([i_1,\dots,i_n]) \\ &= -\lim_{n\to\infty} \frac{b_n}{n} \\ &= \int \log |T'| \mathrm{d}\mu. \end{aligned}$$

La seconda uguaglianza segue dal fatto che

$$\log A + \log \lambda([i_1,\dots,i_n]) \leq \log \mu([i_1,\dots,i_n]) \leq \log B + \log \lambda([i_1,\dots,i_n])$$

e la dimostrazione è completa. □

Capitolo 5

Simulazioni al calcolatore

In questo capitolo presentiamo le trasformazioni considerate per le simulazioni, discutiamo le simulazioni al calcolatore fatte e infine ne illustriamo e commentiamo i risultati.

5.1 Trasformazioni dell'intervallo $[0, 1]$

Abbiamo applicato i metodi di stima dell'entropia a quattro trasformazioni dello spazio $\big([0,1], \mathcal{B}([0,1])\big)$.

La prima è la mappa logistica con fattore 4, definita da

$$\Lambda x = 4x(1-x), \quad x \in [0,1] \tag{5.1}$$

(vedi figura 5.1). È l'unico esempio di mappa continua che consideriamo.

Abbiamo poi considerato una mappa espandente a pezzi, discontinua in $\frac{1}{2}$ e con due rami suriettivi. Essa è definita da

$$Ex = \begin{cases} \frac{4x}{3-2x} & \text{se } x \in [0, \frac{1}{2}] \\ \frac{2x-1}{2-x} & \text{se } x \in]\frac{1}{2}, 1] \end{cases} \tag{5.2}$$

e per essa vale $E'(x) > k$ per ogni x, dove $k > 1$ è una costante (vedi figura 5.2).

Un esempio di mappa con due rami non suriettivi è la mappa di tipo Lorenz definita da

$$Lx = \begin{cases} 1 - \left(\frac{-6x+3}{4}\right)^{\frac{3}{4}} & \text{se } x \in [0, \frac{1}{2}[\\ \left(\frac{6x-3}{4}\right)^{\frac{3}{4}} & \text{se } x \in [\frac{1}{2}, 1] \end{cases} . \tag{5.3}$$

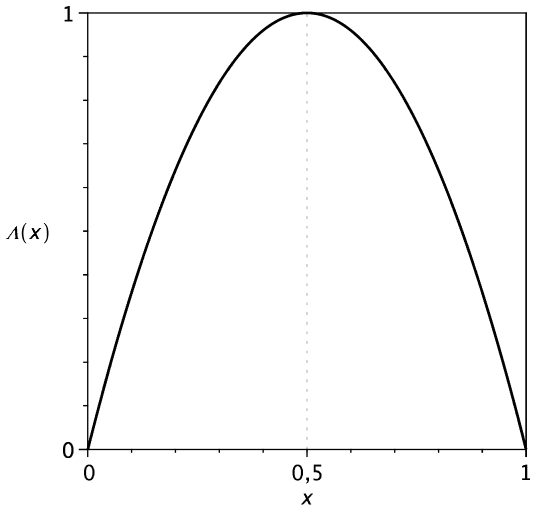


Figura 5.1: La mappa logistica $\Lambda(x) = 4x(1-x)$.

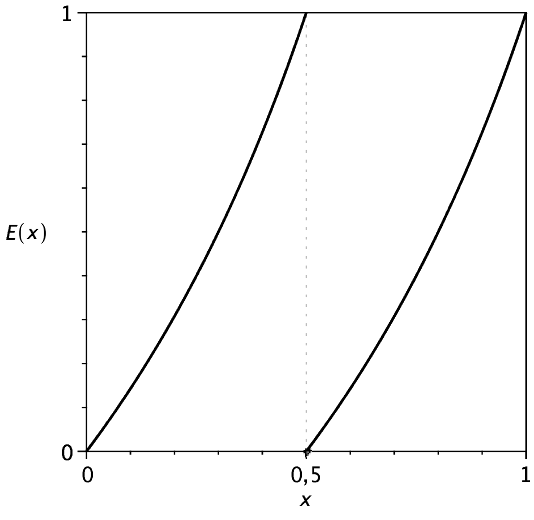


Figura 5.2: La mappa espandente a pezzi E definita da $E(x) = \frac{4x}{3-2x}$ per $0 \leq x \leq \frac{1}{2}$ e da $E(x) = E(x - \frac{1}{2})$ per $\frac{1}{2} < x \leq 1$.

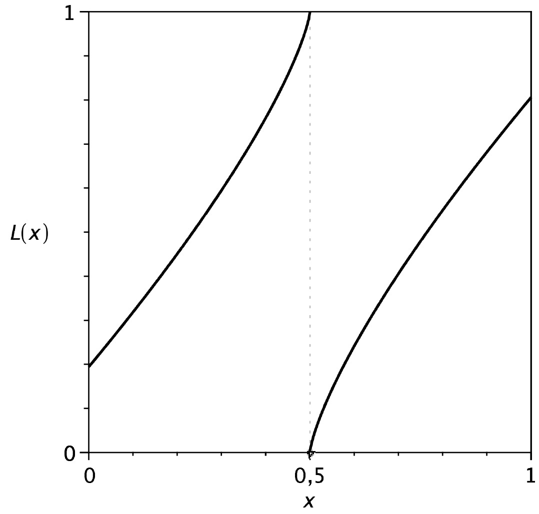


Figura 5.3: Una mappa di tipo Lorenz, con rami non suriettivi e derivata uniformemente maggiore di 1. Essa è definita da $L(x) = \left(\frac{6x-3}{4}\right)^{\frac{3}{4}}$ per $\frac{1}{2} \leq x \leq 1$ e $L(x) = 1 - L(1-x)$ per $0 \leq x < \frac{1}{2}$.

Anche per essa vale $L' > 1$ uniformemente per ogni $x \in [0,1] \setminus \{\frac{1}{2}\}$ e $L'\left(\frac{1}{2}^{\pm}\right) = +\infty$ (vedi figura 5.3).

Per la quarta trasformazione abbiamo preso in considerazione la classe delle mappe di Manneville (vedi [PM]) definite da

$$M_s x = x + x^s \pmod 1, \quad s \in \mathbb{R}^+. \tag{5.4}$$

Tali mappe sono di notevole interesse in fisica e hanno caratteristiche diverse al variare del parametro s. Noi abbiamo focalizzato l'attenzione sui valori $1 < s < 2$, per i quali le mappe ammettono una misura di probabilità invariante assolutamente continua. Nella figura 5.4 è riportato il grafico della mappa di Manneville con esponente $s = \frac{3}{2}$.

La mappa logistica si conosce molto bene. Di essa è nota esplicitamente l'unica misura invariante, che è ergodica e assolutamente continua con densità $\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}$. Pur essendo ρ illimitata, Choe dimostra (vedi [Ch]) che la conclusione del teorema 4.3.4 vale anche in questo caso. Si sa inoltre (vedi ad esempio [Ch]) che il sistema dinamico $\left([0,1], \mathcal{B}([0,1]), \rho(x)\mathrm{d}x, \Lambda\right)$ è isomorfo allo *shift* di Bernoulli con alfabeto $\{0,1\}$ e parametri $(\frac{1}{2}, \frac{1}{2})$. Tale isomorfismo è ottenuto mediante il passaggio alla dinamica simbolica considerando la partizione generatrice $\{[0, \frac{1}{2}],]\frac{1}{2}, 1]\}$.

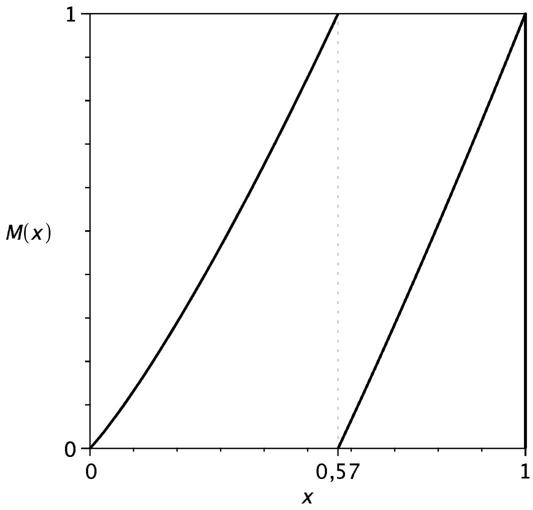


Figura 5.4: La mappa di Manneville con esponente $s = \frac{3}{2}$. La derivata è strettamente maggiore di 1 in ogni $x \in]0,1]$ mentre $M'(0^+) = 1$.

Le mappe espandente a pezzi e di tipo Lorenz possiedono anch'esse un'unica misura di probabilità invariante, la quale è assolutamente continua rispetto alla misura di Lebesgue ed ergodica (vedi Viana [Vi]).

Le mappe di Manneville con esponente $1 < s < 2$ hanno una misura invariante assolutamente continua con densità illimitata per $x \to 0^+$ ma integrabile (vedi Isola [Is]).

Ricordiamo che il passaggio da una trasformazione $T : [0,1] \to [0,1]$ alla dinamica simbolica si effettua fissando una partizione (vedi sezione 1.5.3). Affinché la trasformazione di *shift* sullo spazio di successioni cosí ottenuto costituisca una t.c.m. isomorfa alla trasformazione T, occorre ed è sufficiente che la partizione considerata sia generatrice (proposizione 1.5.4). Nel caso che T sia una delle trasformazioni espandente a pezzi, di tipo Lorenz e di tipo Manneville sono generatrici le partizioni $\{[0,c],]c,1]\}$, dove $c \in]0,1[$ è quel valore tale che $T(c^-) = 1$ e $T(c^+) = 0$ (si veda a tal proposito Buzzi [Bu]).

5.2 Simulazioni

Per le simulazioni al calcolatore abbiamo usato il linguaggio di programmazione *C*, con la libreria *GMP - GNU Multiple Precision* per la gestione di numeri in virgola mobile con precisione arbitraria. Il lettore interessato può visitare il sito Internet `http://gmplib.org/` dove si trova la documentazione completa.

Tali simulazioni sono consistite nel calcolare, per le trasformazioni considerate, un certo numero di iterazioni partendo da un punto pseudo-casuale, nel calcolo dell'esponente di Lyapunov, nella creazione di stringhe simboliche derivanti dalle partizioni scelte, nell'applicazione a tali stringhe dei metodi di stima dell'entropia descritti nei capitoli 3 e 4. Nei casi da noi considerati, le partizioni sono costituite da due elementi cui abbiamo associato, come consuetudine, i simboli 0 e 1.

È necessario prestare attenzione nell'eseguire esperimenti al calcolatore che presuppongono l'iterazione di una data trasformazione, poiché l'accuratezza delle cifre diminuisce man mano che si eseguono le iterazioni. Gli errori dovuti all'uso da parte dei calcolatori di una precisione finita si amplificano, fino a quando le iterazioni calcolate ben poco hanno a che fare con i punti da cui si era partiti. Si può mostrare (vedi Choe [Ch], cap. 9) che, partendo con un certo numero d di cifre binarie significative, dopo n iterazioni rimangono in media $d - nh$ cifre corrette, dove h è l'entropia della trasformazione. Si può fare anche un ragionamento piú preciso, che permette di determinare fino a quando si può iterare la trasformazione avendo la certezza che i simboli binari fino a quel momento determinati siano corretti.

Nell'eseguire il calcolo delle iterazioni abbiamo fatto due prove. Nella prima abbiamo usato per ogni numero gestito un numero di cifre molto alto (variabile da 56000 a 100000 a seconda delle trasformazioni), tale da garantire la correttezza teorica di 100000 simboli binari. L'ovvio prezzo da pagare per la correttezza teorica del calcolo è la lentezza dell'esecuzione del programma e il conseguente basso numero di iterazioni calcolabili in tempi accettabili. Nella seconda prova abbiamo usato un numero di cifre piú basso (1000), riuscendo ad ottenere per ognuna delle mappe una stringa simbolica lunga 33 milioni di *bit*. Al vantaggio di avere successioni simboliche piú lunghe si contrappone lo svantaggio (almeno in teoria) di avere dati affetti da errore. Tuttavia, applicando i vari metodi di stima alle successioni simboliche da 100 mila e 33 milioni di *bit* abbiamo ottenuto risultati uguali nella sostanza. Interpretiamo ciò dicendo che, anche con una precisione di calcolo teoricamente non sufficiente, non viene alterata in maniera apprezzabile la dinamica delle mappe. Vista la sostanziale uguaglianza dei risultati delle simulazioni nei due casi, riporteremo e commenteremo i dati ottenuti dall'applicazione dei metodi di stima alle successioni simboliche lunghe 33 milioni di *bit*.

Il calcolo dell'esponente di Lyapunov (ovvero la media rispetto alla misura invariante di $\log_2 |T'|$) è stato eseguito calcolando 100 milioni di iterazioni utilizzando una precisione pari a 1000 cifre binarie.

Il metodo di stima delle frequenze empiriche basato sul teorema 4.1.2 è stato applicato (seguendo le notazioni del medesimo teorema) con $k = 1, 2, \ldots$ e $n(k) = \lceil 2^{k\overline{h}} \rceil$, dove $\overline{h}$ è il valore dell'entropia alla terza cifra decimale dato dal calcolo dell'esponente di Lyapunov e $\lceil x \rceil$ denota il minimo intero maggiore di o uguale a x. Sono stati considerati valori di k crescenti fin tanto che il tempo di esecuzione al calcolatore è rimasto contenuto entro poche decine di minuti. Come valore massimo abbiamo considerato $k = 25$. Osserviamo che il tempo richiesto cresce esponenzialmente con k, giacché l'algoritmo richiede un numero di operazioni al passo k proporzionale a 2^k.

Poiché l'ipotesi del teorema 4.1.2 assume che si abbia $k(n) \leq \frac{1}{h} \log_2 n$ (o, equivalentemente, $n(k) \geq 2^{hk}$), dove h è l'entropia della trasformazione, impostando nella formula di $k(n)$ (ovvero di $n(k)$) un valore $\tilde{h} \geq h$ l'ipotesi è vera *a fortiori*. Si potrebbe pertanto pensare di considerare dei valori $k(n)$ che crescano con n piú lentamente (o, ciò che è lo stesso, dei valori di $n(k)$ che crescano piú velocemente con k). Abbiamo fatto delle prove in questo senso, riuscendo ad ottenere delle approssimazioni decisamente migliori del valore di riferimento. Tuttavia, i valori $\tilde{h} \geq h$ piú opportuni sono stati determinati empiricamente a seconda della trasformazione considerata, senza ravvisare in essi un criterio che li suggerisca *a priori*.

Il metodo dei tempi di ritorno di stringhe simboliche (vedi teorema 4.2.2) è stato eseguito mediando i dati su un campione di 1000 stringhe. Tale versione dell'algoritmo di stima dà risultati piú attendibili rispetto a quella che considera i tempi di ritorno delle sole stringhe simboliche iniziali, giacché quest'ultima risente fortemente della scelta pseudo-casuale del punto iniziale. Nella versione media, inoltre, i dati sono piú regolari e non presentano, come invece i dati relativi alla versione singola, rami di monotonia intervallati da salti di ampiezza apprezzabile.

Il metodo delle sostituzioni di coppie di Grassberger è stato applicato in conformità al capitolo 3 e all'appendice A. Vi sono solo due precisazioni da fare. Quando si sostituisce una coppia formata da simboli uguali, la somma infinita in (A.12) è stata approssimata troncandola ai primi $\lfloor \log_2 l(s) \rfloor$ addendi, dove $l(s)$ è la lunghezza del vettore simbolico dopo s sostituzioni di coppie e $\lfloor x \rfloor$ indica il massimo intero minore di o uguale a x. Come seconda osservazione, facciamo presente che nelle simulazioni si sono eseguite in tutti i casi 280 sostituzioni di coppie. Dai dati ottenuti, tale numero sembra sufficiente.

5.3 Risultati delle simulazioni

Commenteremo diffusamente i risultati nel caso della mappa logistica e si potrà osservare come negli altri casi i dati si prestino spesso a un'analisi

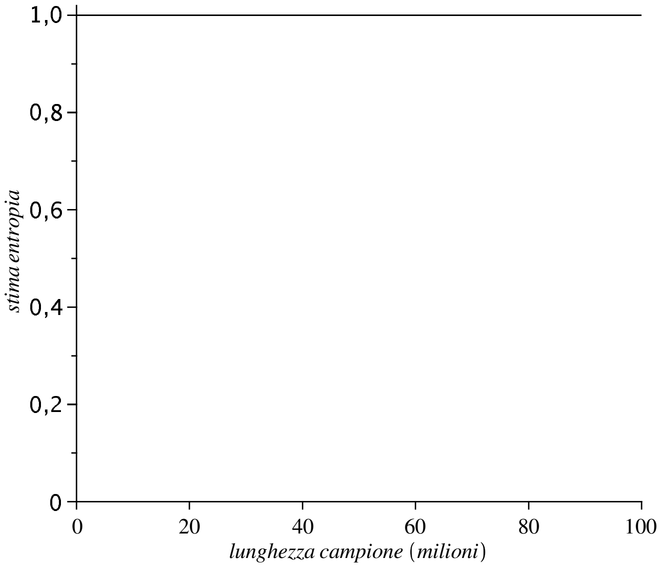


Figura 5.5: Stima dell'esponente di Lyapunov per la mappa logistica Λ.

analoga. Quando non dovesse essere cosí ci soffermeremo per evidenziare i dettagli.

5.3.1 Mappa logistica Λ

Il calcolo dell'esponente di Lyapunov si stabilizza molto velocemente, fino già alla sesta cifra decimale dopo 10 milioni di iterazioni, tanto che i dati riportati nel grafico in figura 5.5 non sono distinguibili da una linea retta.

Nel grafico di figura 5.6 sono riportati i medesimi dati. Nonostante l'enorme ingrandimento di scala sulle ordinate, i dati risultano quasi perfettamente allineati sul valore 1.

Vista la notevole precisione e stabilizzazione dell'esponente di Lyapunov, esso verrà preso come valore 'vero' dell'entropia con cui comparare i dati ottenuti con gli altri metodi di stima.

Calcolando la distribuzione empirica dei blocchi simbolici non sovrapposti, otteniamo dei dati poco soddisfacenti (vedi figura 5.7).

La crescita dei valori stimati è estremamente lenta, e nei limiti del numero di iterazioni considerate essi arrivano a indicare un valore di stima molto piú basso rispetto al valore di riferimento. Si arguisce dal grafico in figura 5.7 che per osservare valori vicini a 1 bisognerebbe disporre di un numero molto piú alto di iterazioni.

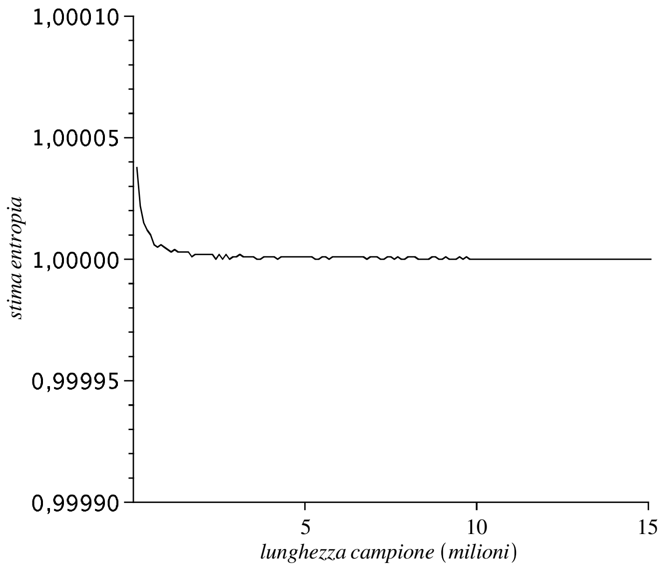


Figura 5.6: Stima dell'esponente di Lyapunov per la mappa logistica Λ (zoom).

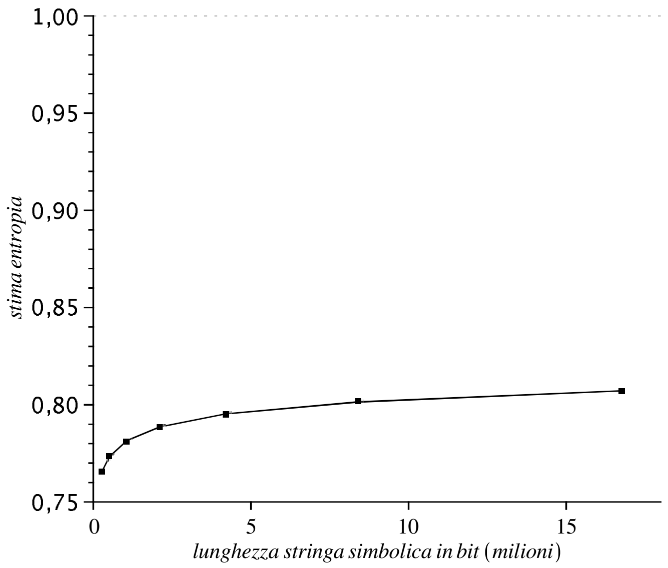


Figura 5.7: Stima dell'entropia per la mappa Λ mediante il calcolo delle frequenze empiriche dei blocchi simbolici.

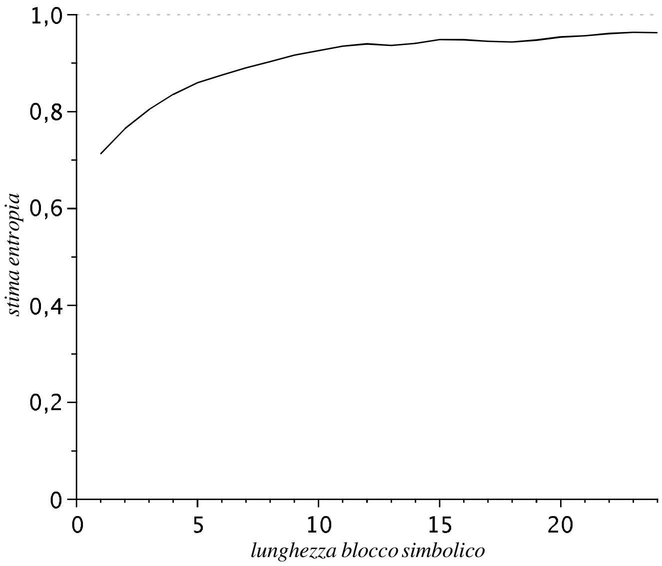


Figura 5.8: Stima dell'entropia della mappa Λ mediante il calcolo dei tempi di ritorno di stringhe simboliche.

La stima ottenuta mediante il metodo dei tempi di ritorno presenta un andamento crescente al crescere della lunghezza l del blocco considerato (vedi figura 5.8). Complessivamente, i valori massimi ottenuti approssimano meglio il valore dell'entropia dato dall'esponente di Lyapunov rispetto al metodo del calcolo delle entropie empiriche.

Il metodo di Grassberger fornisce stime dell'entropia che crescono e si avvicinano velocemente al valore dell'entropia nel giro delle prime decine di sostituzioni. I valori di stima raggiungono un massimo in corrispondenza della 157-esima sostituzione, dopo la quale essi decrescono lentamente (vedi figure 5.9 e 5.10).

5.3.2 Mappa espandente a pezzi E

Anche in questo caso il calcolo dell'esponente di Lyapunov si stabilizza velocemente, anche se in misura minore rispetto al caso della mappa logistica. Nel corso dei 100 milioni di iterazioni considerate, il valore stimato dell'entropia si assesta fino alla terza cifra decimale e anche la quarta subisce oscillazioni molto ridotte. Posizionando i dati in un grafico con intervallo di osservazione delle ordinate $[0, 1]$ (figura 5.11), dopo i primi 7-8 milioni di iterazioni essi non sono distinguibili da una linea retta.

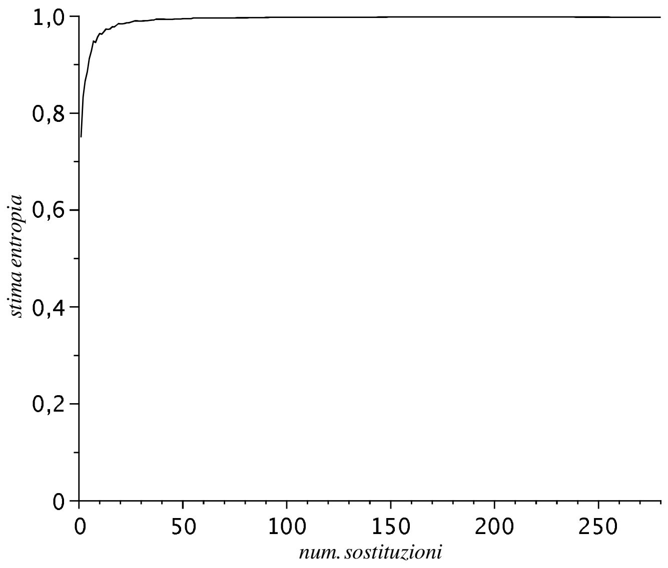


Figura 5.9: Stima dell'entropia della mappa logistica Λ mediante il metodo delle sostituzioni di coppie di Grassberger.

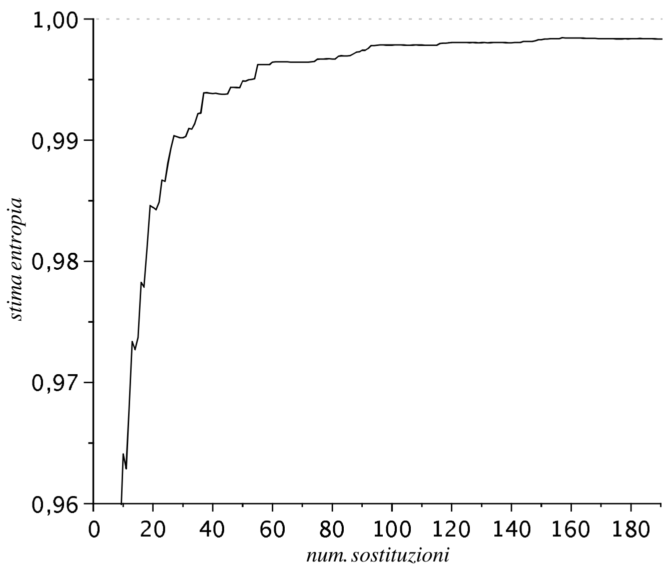


Figura 5.10: Stima dell'entropia della mappa logistica Λ mediante il metodo delle sostituzioni di coppie di Grassberger (zoom).

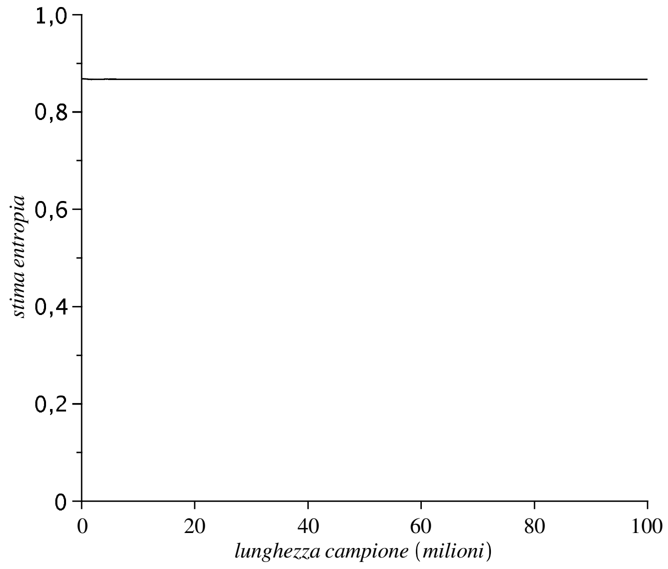


Figura 5.11: Stima dell'esponente di Lyapunov per la mappa espandente a pezzi E.

Nel grafico di figura 5.12 riportiamo i medesimi dati ma con intervallo di osservazione sulle ordinate notevolmente ridotto.

Anche in questo caso, considerate la precisione e stabilizzazione dell'esponente di Lyapunov, esso verrà preso come valore di riferimento per l'entropia.

Il metodo del calcolo della distribuzione empirica dei blocchi simbolici non sovrapposti fornisce dei dati qualitativamente analoghi al caso corrispondente per la mappa logistica. Tali dati crescono molto lentamente e restano molto al di sotto del valore stimato dall'esponente di Lyapunov (vedi figura 5.13).

La stima ottenuta mediante il metodo dei tempi di ritorno (vedi figura 5.14) è anche in questo caso migliore rispetto a quella ottenuta dal calcolo delle probabilità empiriche dei blocchi simbolici, in termini di una migliore approssimazione del valore di riferimento.

Il metodo delle sostituzioni di coppie di Grassberger, analogamente al caso della mappa logistica, fornisce stime dell'entropia che crescono e si avvicinano velocemente al valore di riferimento (vedi figure 5.15 e 5.16). I valori di stima raggiungono un massimo intorno alla 170-esima sostituzione, e poi decrescono lentamente.

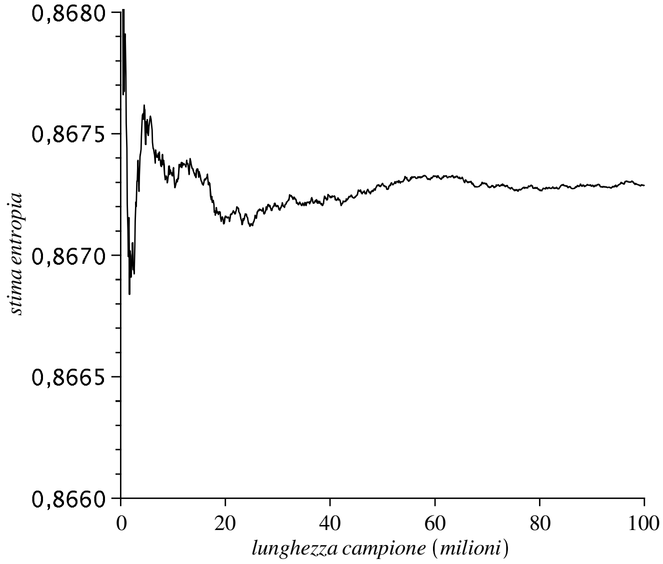


Figura 5.12: Stima dell'esponente di Lyapunov per la mappa espandente a pezzi E (zoom).

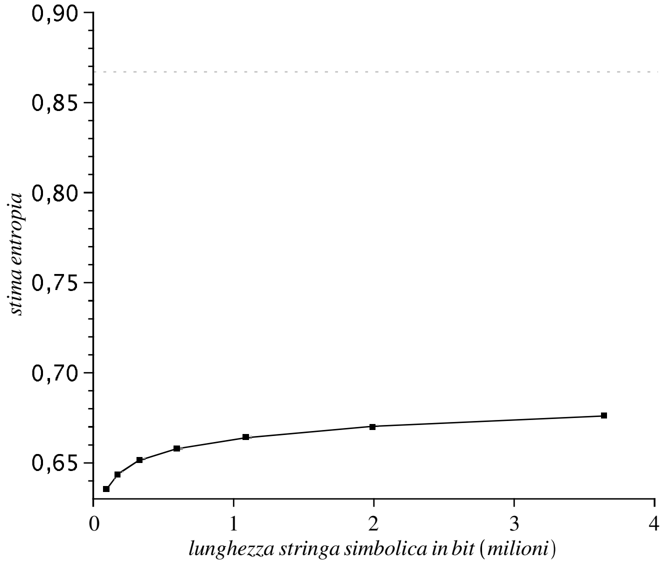


Figura 5.13: Stima dell'entropia per la mappa E mediante il calcolo delle frequenze empiriche dei blocchi simbolici.

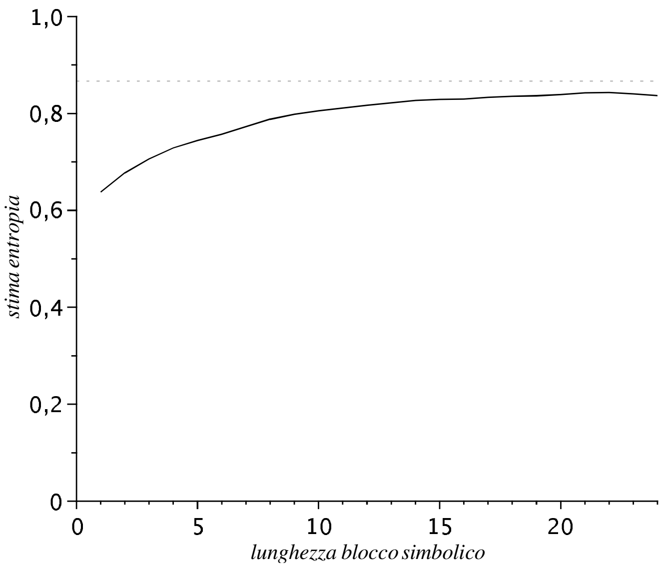


Figura 5.14: Stima dell'entropia della mappa E mediante il calcolo dei tempi di ritorno di stringhe simboliche.

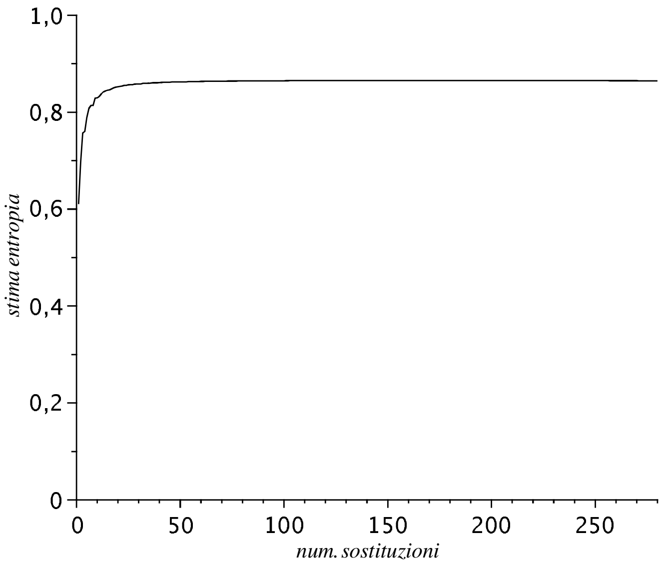


Figura 5.15: Stima dell'entropia della mappa espandente a pezzi E mediante il metodo delle sostituzioni di coppie di Grassberger.

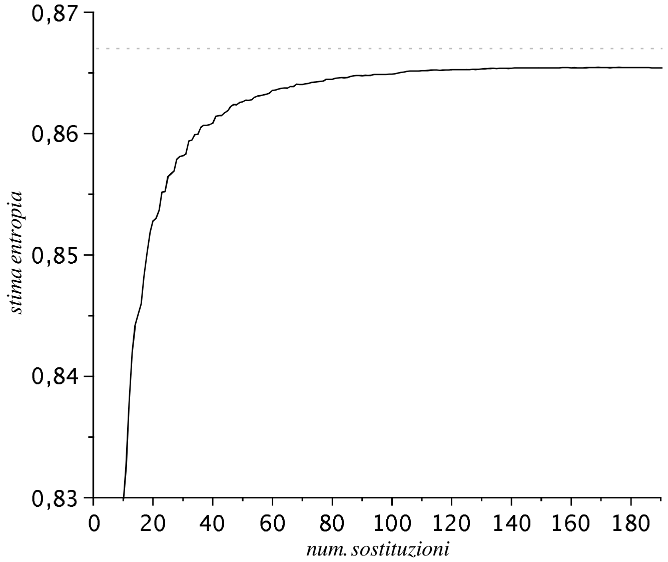


Figura 5.16: Stima dell'entropia della mappa espandente a pezzi E mediante il metodo delle sostituzioni di coppie di Grassberger (zoom).

5.3.3 Mappa di tipo Lorenz L

La stima dell'esponente di Lyapunov per la mappa L ha un andamento molto simile al caso della mappa espandente a pezzi E. Nei 100 milioni di iterazioni calcolate essa si stabilizza velocemente fino alla quarta cifra decimale, come si può osservare dal grafico in figura 5.18. In figura 5.17 si nota come senza ingrandimenti degli assi i dati sembrano perfettamente allineati.

Nuovamente, il valore determinato dalla stima dell'esponente di Lyapunov verrà preso come termine di paragone per i dati ottenuti con gli altri metodi di stima.

Calcolando la distribuzione empirica dei blocchi simbolici non sovrapposti, otteniamo i dati riportati nel grafico di figura 5.19. Valgono anche in questo caso le corrispondenti considerazioni fatte per le mappe Λ e E.

Il calcolo dei tempi di ritorno fornisce, come per le mappe Λ e E, stime migliori rispetto al metodo del calcolo delle entropie empiriche dei blocchi simbolici (vedi figura 5.20).

Il metodo NSRPS stima l'entropia in modo qualitativamente analogo ai casi delle mappe logistica e espandente a pezzi. I valori di stima crescono fino a raggiungere un massimo in corrispondenza della 159-esima sostituzione, dopo la quale essi decrescono seppure molto lentamente (vedi figure 5.21

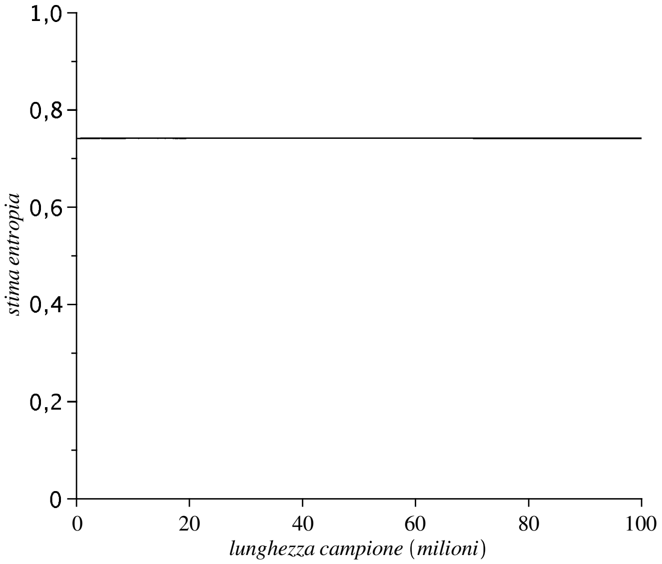


Figura 5.17: Stima dell'esponente di Lyapunov per la mappa di tipo Lorenz L.

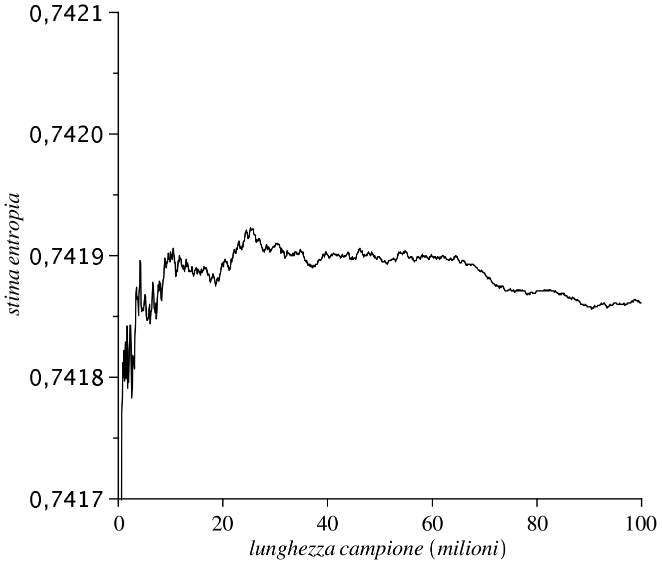


Figura 5.18: Stima dell'esponente di Lyapunov per la mappa di tipo Lorenz L (zoom).

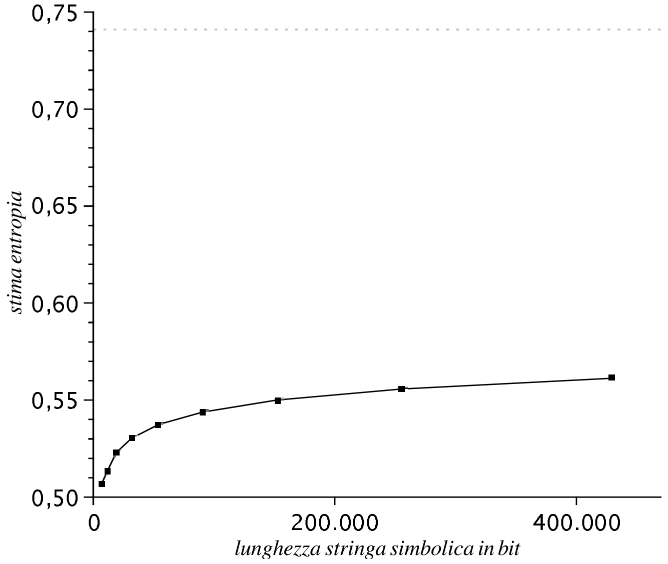


Figura 5.19: Stima dell'entropia per la mappa L mediante il calcolo delle frequenze empiriche dei blocchi simbolici.

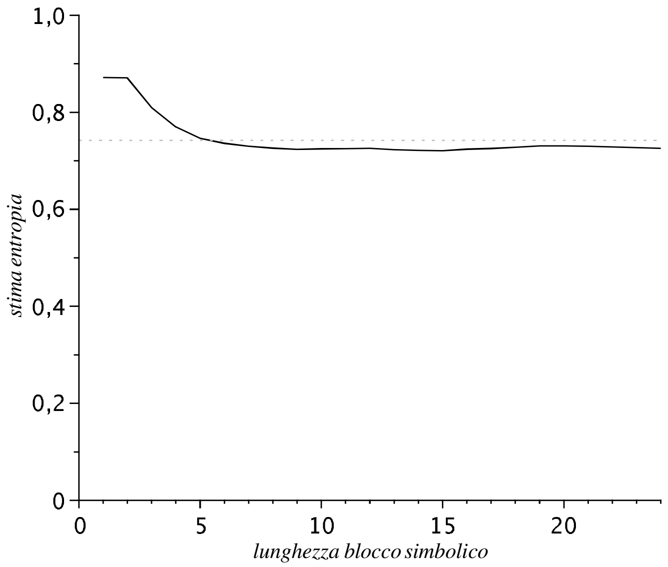


Figura 5.20: Stima dell'entropia della mappa L mediante il calcolo dei tempi di ritorno di stringhe simboliche.

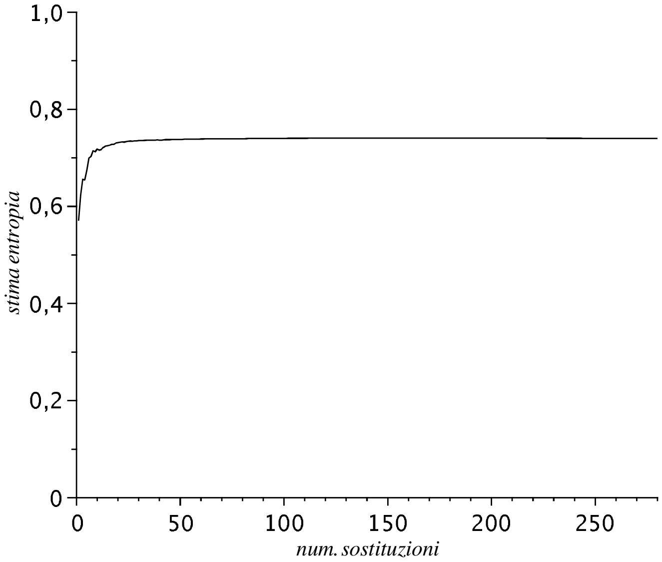


Figura 5.21: Stima dell'entropia della mappa di tipo Lorenz L mediante il metodo delle sostituzioni di coppie di Grassberger.

e 5.22).

5.3.4 Mappa di Manneville M con esponente $\frac{3}{2}$

La stabilizzazione della stima dell'esponente di Lyapunov nel caso della mappa di Manneville è un po' piú lenta rispetto agli altri tre esempi considerati. I dati inseriti nel grafico di figura 5.23 si presentano quasi su una retta orizzontale e uno zoom sull'asse delle ordinate (vedi figura 5.24) mostra che nell'arco dei 100 milioni di iterazioni calcolate la convergenza della stima si stabilizza fino alla terza cifra decimale. Tuttavia osserviamo che negli altri tre esempi si registrava una stabilizzazione superiore. Questa maggiore oscillazione dei valori di stima è dovuta al fatto che la derivata $M'(x)$ è uguale a 1 in 0 ed è molto vicina a 1 per valori molto piccoli di x. Quando, iterando la trasformazione, capita che un'iterazione cada molto vicino a 0, le successive iterate si manterranno per molto tempo vicine a 0. Si ha cosí che la stima dell'esponente di Lyapunov è continuamente alterata dalla presenza di lunghi blocchi di iterazioni per cui la quantità da mediare $\log_2 |M'|$ è pressoché nulla. Si spiegano cosí le frequenti ripide diminuzioni della quantità stimata osservabili in figura 5.24. In termini di successioni simboliche, il fenomeno

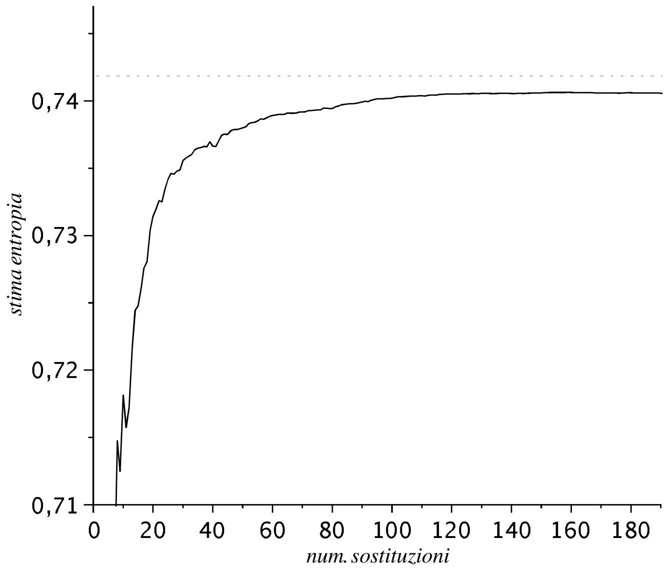


Figura 5.22: Stima dell'entropia della mappa di tipo Lorenz L mediante il metodo delle sostituzioni di coppie di Grassberger (zoom).

appena osservato si traduce nella presenza di lunghe sequenze costituite da molti zeri consecutivi.

Nonostante la stima dell'entropia come esponente di Lyapunov sia piú imprecisa rispetto ai casi delle mappe Λ, E, L, tre cifre significative stabilizzate sono comunque sufficienti perché possiamo prendere tale stima come valore di riferimento per il raffronto tra i vari metodi.

Il calcolo della distribuzione empirica dei blocchi simbolici non sovrapposti dà dei risultati insoddisfacenti (vedi figura 5.25). I valori di stima che si ottengono con questo metodo sono molto bassi e la loro crescita molto lenta, come del resto avviene nel caso delle altre tre trasformazioni prese in considerazione.

Il metodo dei tempi di ritorno dà una stima che, differentemente da quanto accadeva per le mappe Λ, E, L, resta molto al di sotto del valore di riferimento (vedi figura 5.26). Per ciascuna delle altre tre mappe lo scarto tra il valore di riferimento e gli ultimi valori stimati col metodo dei tempi di ritorno è inferiore a $0,036$. Per la mappa di Manneville invece tale scarto è circa $0,076$. Tale fatto si spiega con l'osservazione seguente. Nel caso della mappa di Manneville M ci sono molti blocchi simbolici di misura molto piccola e per i quali quindi i tempi di ritorno sono mediamente molto alti. La nostra successione simbolica lunga 33 milioni di *bit* è troppo corta perché in essa tali

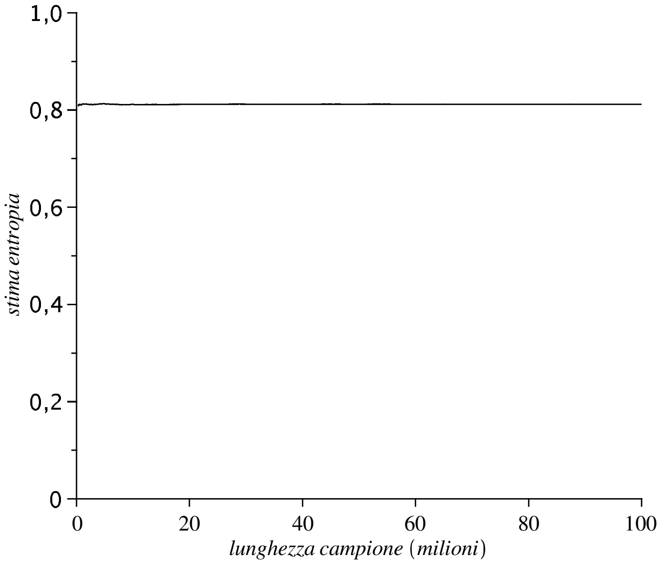


Figura 5.23: Stima dell'esponente di Lyapunov per la mappa di Manneville M.

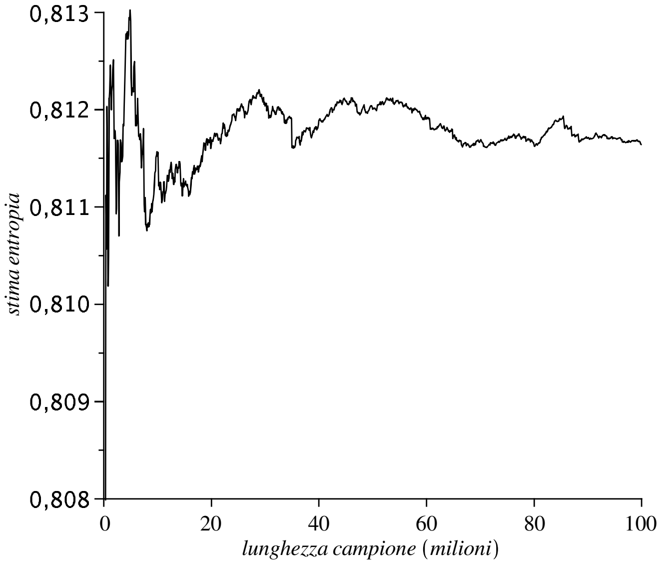


Figura 5.24: Stima dell'esponente di Lyapunov per la mappa di Manneville M (zoom).

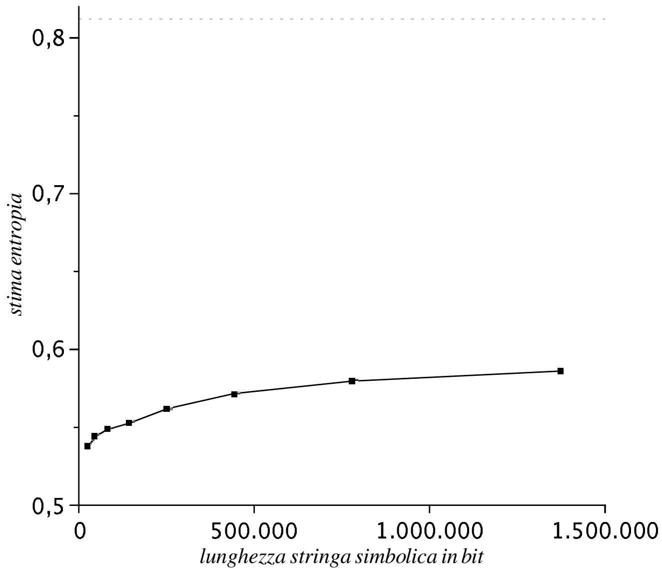


Figura 5.25: Stima dell'entropia per la mappa M mediante il calcolo delle frequenze empiriche dei blocchi simbolici.

blocchi di misura piccola si ripresentino. Segue da ciò che la media dei tempi di ritorno che noi calcoliamo su un campione di 1000 stringhe è sottostimata, ed essa è tanto piú sottostimata quanto piú sono numerosi i blocchi simbolici del campione di bassa probabilità. Nel caso della mappa di Manneville, per il blocco lungo 24 simboli abbiamo registrato sempre almeno 90 casi su 1000 in cui il blocco non ricompare entro la fine della successione. Molti di meno erano i casi analoghi per le mappe Λ, E, L. Osserviamo che con tale medesima motivazione Choe spiega un analogo fenomeno riguardo allo *shift* di Bernoulli di parametri $(p, 1-p)$ con $p \neq \frac{1}{2}$ (vedi [Ch], capitolo 12).

Il metodo di Grassberger fornisce per la mappa M dei risultati differenti rispetto alle altre tre mappe (vedi figure 5.27 e 5.28). In questo caso infatti le stime dell'entropia crescono e raggiungono un valore massimo molto al di sotto del valore di riferimento. Lo scarto è circa $0,048$, valore decisamente piú alto rispetto a quanto si aveva per le altre tre mappe, per le quali lo scarto era inferiore a $0,002$. Nonostante questa diversità, anche in questo caso i valori di stima raggiungono un massimo (questa volta in corrispondenza della 153-esima sostituzione) per poi decrescere molto lentamente.

Per capire il motivo di questa peggiore approssimazione del valore di riferimento da parte del metodo di Grassberger, abbiamo provato a variare alcuni parametri. Innanzi tutto, ricordiamo che nel programma che imple-

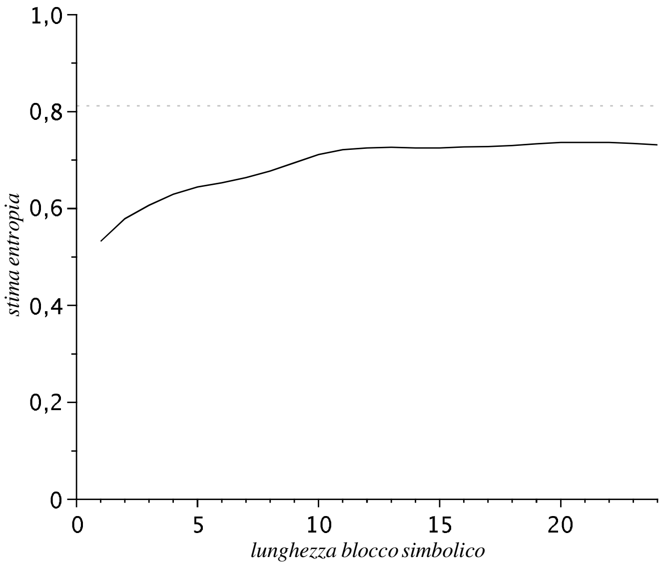


Figura 5.26: Stima dell'entropia della mappa M mediante il calcolo dei tempi di ritorno di stringhe simboliche.

menta l'algoritmo abbiamo troncato una somma infinita a un numero di addendi pari a $\lfloor \log_2 l(s) \rfloor$ (vedi sezione 5.2). Per capire se questo potesse incidere sui valori stimati abbiamo provato ad aumentare il numero di addendi calcolati. Tuttavia, a tale aumento di precisione teorica non è corrisposta una variazione rilevabile entro la sesta cifra decimale. Siamo quindi portati a concludere che l'approssimazione della somma infinita con una somma finita non è responsabile del fatto che la stima ottenuta è insoddisfacente.

Abbiamo poi provato ad applicare il metodo NSRPS a stringhe simboliche di lunghezza piú corta, sia nel caso della mappa di Manneville che negli altri tre casi. Al diminuire della lunghezza della stringa abbiamo registrato una diminuzione del valore massimo stimato e una diminuzione del numero di sostituzioni in corrispondenza del quale tale massimo è raggiunto. Ciò lascia supporre che si avrebbero migliori risultati di approssimazione del valore di riferimento applicando il metodo delle sostituzioni a una stringa simbolica piú lunga. Il motivo perché questo sia il caso della sola mappa di Manneville tra quelle considerate risiede probabilmente nelle caratteristiche molto particolari della mappa M.

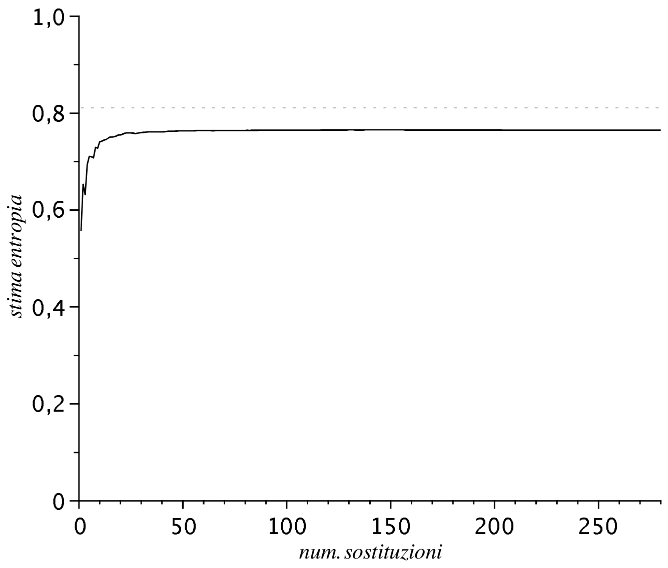


Figura 5.27: Stima dell'entropia della mappa di Manneville M mediante il metodo delle sostituzioni di coppie di Grassberger.

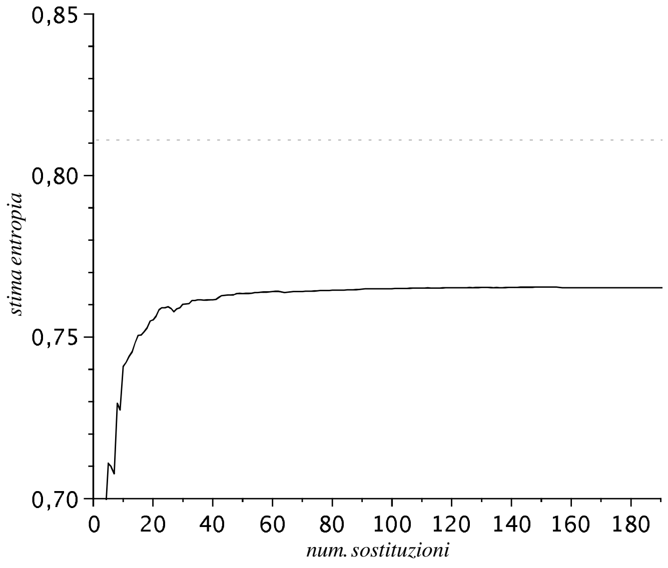


Figura 5.28: Stima dell'entropia della mappa di Manneville M mediante il metodo delle sostituzioni di coppie di Grassberger (zoom).

5.4 Tempi di esecuzione delle simulazioni

Esponiamo in questa sezione alcuni commenti importanti sul tempo di esecuzione dei programmi delle simulazioni.

Nella sezione 5.3 abbiamo confrontato e valutato i vari metodi di stima dell'entropia studiati in virtú della loro migliore o peggiore approssimazione di un valore preso come riferimento. Non abbiamo fatto menzione dei tempi di esecuzione dei programmi, cosa che facciamo qui di seguito. Ai fini di una valutazione piú completa di un metodo computazionale di stima, la questione del tempo che occorre per l'applicazione dello stesso è tutt'altro che oziosa. Un metodo di stima che desse buoni o ottimi risultati di approssimazione ma che richiedesse dei tempi sproporzionatamente grandi non potrebbe essere valutato complessivamente bene.

Riporteremo i tempi approssimativi impiegati nell'esecuzione dei programmi su uno stesso calcolatore. Per una trattazione piú precisa dovrebbe eseguirsi un'analisi della complessità degli algoritmi. Per i nostri fini è tuttavia sufficiente una comparazione relativa tra i tempi dei vari metodi.

La velocità dei programmi che stimano l'esponente di Lyapunov dipende in modo consistente dal grado di precisione utilizzato e dal numero di operazioni necessarie nel calcolo di ogni iterazione. Tuttavia, come abbiamo già avuto modo di osservare, anche con una precisione di calcolo non eccessivamente alta (1000 cifre binarie) i risultati non differiscono quantitativamente in modo significativo da quelli ottenuti con precisione molto piú elevata. I programmi che abbiamo realizzato per la stima dell'esponente di Lyapunov (vedi sezione 5.2) impiegano circa 10-12 minuti, le variazioni essendo dovute alla trasformazione considerata.

Quello che ci interessa veramente però è il confronto fra i tempi di esecuzione dei programmi che realizzano gli altri tre metodi di stima, poiché, come già abbiamo osservato nell'*Introduzione*, questi si basano su operazioni eseguite su successioni simboliche e hanno pertanto un carattere essenzialmente diverso rispetto al calcolo dell'esponente di Lyapunov.

Il metodo che calcola le frequenze empiriche dei blocchi simbolici è fortemente influenzato dal valore $\tilde{h}$ che si usa nella relazione $k(n) = \left\lfloor \frac{1}{\tilde{h}} \log_2 n \right\rfloor$ o nella relazione $n(k) = \left\lceil 2^{\tilde{h}k} \right\rceil$. L'implementazione del metodo richiede un numero di operazioni al passo k dell'ordine di $n_k^{1/\tilde{h}}$.

Se si usa il valore dell'entropia $\tilde{h} = h$ e si esegue l'algoritmo per valori di n equidistanziati, i tempi richiesti sono enormemente alti. Per esempio, se $n_k = k \cdot 10^5$, $k = 1, 2, \ldots, 330$, il numero di operazioni richieste (e quindi anche il tempo di esecuzione del programma) è proporzionale a $\sum_{k=1}^{330} (k \cdot 10^5)^{1/h} \approx 10^{5/h} 330^{1/h+1}$. Il tempo necessario ha quindi una crescita con legge di potenza con esponente pari a 1 piú l'inverso dell'entropia. Ciò rende l'algoritmo estremamente lento, ed esso è tanto piú inefficiente quanto piú piccola è l'entropia h della trasformazione, tanto da risultare inapplicabile in tempi ragionevoli per valori di h inferiori a circa $0,9$.

Abbiamo superato questa difficoltà eseguendo il calcolo per valori di n i cui logaritmi siano equidistanziati. In particolare abbiamo considerato $k = 1, 2, \ldots, 25$ e $n(k) = \lceil 2^{\tilde{h}k} \rceil$, dove $\tilde{h} = h + 10^{-3}$ e h è il valore dell'entropia alla terza cifra decimale dato dalla stima dell'esponente di Lyapunov.

Gli algoritmi che implementano il metodo dei tempi di ritorno e quello delle sostituzioni di coppie di Grassberger sono decisamente piú efficienti. La loro esecuzione ha richiesto 5-6 minuti per il calcolo dei tempi di ritorno di 1000 blocchi simbolici e circa 3 minuti per effettuare 280 sostituzioni di coppie simboliche. In questi due algoritmi il numero di operazioni eseguite è lineare nella lunghezza della successione simbolica.

Le piccole variazioni registrate per il calcolo dei tempi di ritorno nei casi delle quattro trasformazioni sono imputabili al fatto che è talvolta necessario scorrere tutta la successione simbolica (quando il dato blocco simbolico non ricompare entro i 33 milioni di simboli) e talvolta no.

Per le sostituzioni di coppie, le piccole variazioni dei tempi di esecuzione sono da attribuirsi principalmente a due fattori. In primo luogo, ricordiamo che quando la sostituzione è del tipo $xx \to \alpha$, cioè quando si sostituiscono due simboli uguali, il programma stima la somma infinita nella (3.8) calcolando i primi $\lfloor \log_2 l(s) \rfloor$ addendi, dove $l(s)$ è la lunghezza del vettore simbolico dopo s sostituzioni di coppie (vedi sezione 5.2). A tal fine il programma scorre la successione simbolica $\lfloor \log_2 l(s) \rfloor - 1$ volte per determinare le frequenze empiriche dei blocchi di lunghezza i, per $2 \leq i \leq \lfloor \log_2 l(s) \rfloor$. Il secondo fattore che determina piccole differenze dei tempi di esecuzione del programma al variare della mappa considerata è il fatto che ogni sostituzione riduce la lunghezza della successione simbolica e ovviamente ciò avviene in misura diversa per mappe diverse.

Conclusioni

In questa tesi abbiamo studiato da un punto di vista teorico e computazionale un metodo di stima dell'entropia di processi simbolici e di sistemi dinamici ergodici.

Tale metodo si basa sul lavoro di Grassberger [Gr] e sui risultati ottenuti e formalizzati da Benedetto, Caglioti, Gabrielli in [BCG].

Il metodo di Grassberger (o metodo NSRPS, da Non-Sequential Recursive Pair Substitution) consiste nella sostituzione ricorsiva di coppie di simboli (le piú frequenti) con simboli singoli nuovi. Il risultato fondamentale, esposto nel teorema 3.4.1, asserisce che l'applicazione ripetuta della procedura di sostituzione trasforma, nel limite per il numero di sostituzioni che tende a infinito, ogni processo ergodico in un processo 1-markoviano.

Il nostro approfondimento teorico è consistito nello studio di un esempio molto particolare di processo (vedi sezione 3.5), proposto nell'articolo [BCG] dove gli autori elencano i risultati senza riportare i calcoli. Per il processo di tale esempio e per le sostituzioni di coppie considerate vale la conclusione del teorema 3.4.1 pur non valendone l'ipotesi, come abbiamo dimostrato nella sezione 3.5. In particolare, le sostituzioni considerate non sono relative a coppie di simboli tra quelle con probabilità massima. Ciò nonostante, il processo diventa ugualmente 1-markoviano nel limite per il numero di sostituzioni che tende a infinito, e anzi diventa addirittura indipendente (0-markoviano). Trovare condizioni sufficienti e necessarie sul processo e sulle sostituzioni affinché valga la tesi del teorema 3.4.1 resta una questione aperta, di soluzione apparentemente non semplice. Ciò nondimeno, ci sembra questa una delle piú interessanti direzioni da investigare per quanto riguarda il metodo NSRPS.

Guardando al metodo NSRPS da un punto di vista applicativo, il teorema centrale 3.4.1 suggerisce un algoritmo per la stima dell'entropia di processi simbolici ergodici, che abbiamo implementato per ottenere delle simulazioni al calcolatore nel caso di trasformazioni ergodiche dell'intervallo unitario (vedi capitolo 5). Per una valutazione del metodo di stima studiato, lo abbiamo confrontato con altri tre (vedi capitolo 4). Due di questi sono anch'essi applicabili nell'ambito della dinamica simbolica (vedi sezioni 4.1 e 4.2), mentre un altro è il calcolo dell'esponente di Lyapunov (vedi sezione 4.3) e presuppone la conoscenza esplicita della trasformazione considerata.

Dalle simulazioni realizzate al calcolatore (vedi capitolo 5) emerge che la stima dell'esponente di Lyapunov, ottenuta mediando la quantità $\log_2 |T'|$ su un'orbita molto lunga, dà ottimi risultati sia in termini di approssimazione del valore dell'entropia sia in termini del tempo di esecuzione dell'algoritmo. In particolare, abbiamo osservato per tutte e quattro le trasformazioni considerate una notevole stabilizzazione del valore stimato. Nel caso della mappa logistica, tale valore di stima era perfettamente in accordo col valore vero dell'entropia, noto da argomentazioni teoriche. In virtú di tale precisione, il valore stimato dell'esponente di Lyapunov è stato preso come valore dell'entropia delle altre tre trasformazioni dell'intervallo $[0, 1]$, per le quali il valore teorico dell'entropia non è noto.

Abbiamo altresí visto che cercare di stimare l'entropia calcolando le frequenze empiriche dei blocchi (vedi teorema 4.1.2) su stringhe simboliche di lunghezza crescente in modo lineare è un metodo assolutamente inefficiente in termini di tempo. Un'analisi dell'algoritmo ha mostrato che il numero di operazioni necessarie è estremamente alto e aumenta all'aumentare del blocco considerato seguendo una legge di potenza con esponente pari a $1+\frac{1}{h}$ (vedi sezione 5.4). Abbiamo dovuto pertanto eseguire le stime in corrispondenza di valori crescenti in modo esponenziale (vedi sezione 5.4).

Oltre alla lentezza di tale algoritmo, abbiamo osservato che le stime ottenute con esso crescono in maniera estremamente lenta al crescere del blocco simbolico considerato, e per il numero di iterazioni che abbiamo considerato i valori massimi restano molto al di sotto del valore di riferimento.

Il metodo di stima dato dal calcolo dei tempi di ritorno di stringhe simboliche finite (vedi sezioni 4.2 e 5.2) ha dato, per tre delle quattro mappe considerate, risultati molto piú soddisfacenti rispetto al metodo delle frequenze empiriche in termini di approssimazione del valore di riferimento. Nel caso della mappa di Manneville M i risultati di approssimazione non sono buoni, e il motivo di ciò risiede nella particolarità della mappa e nella lunghezza non sufficiente dell'orbita campione considerata, come spiegato nella sezione 5.3.4. I tempi di esecuzione dell'algoritmo sono piuttosto contenuti (vedi sezione 5.4).

L'applicazione del metodo delle sostituzioni di coppie di Grassberger (NSRPS) ha dato ottimi risultati di approssimazione per tre delle quattro trasformazioni prese in esame. Come nel caso dei tempi di ritorno, risultati non soddisfacenti (anche se migliori rispetto ai metodi delle frequenze empiriche e dei tempi di ritorno) sono stati ottenuti per la mappa di Manneville (vedi sezione 5.3.4). Sempre come nel caso dei tempi di ritorno, anche il tempo di esecuzione dei programmi al calcolatore è stato decisamente contenuto (vedi sezione 5.4) e ciò fa sí che il metodo NSRPS sia interessante anche da un punto di vista applicativo computazionale.

Il metodo delle sostituzioni di coppie simboliche si è dimostrato, nelle simulazioni eseguite, un valido strumento di stima dell'entropia per alcune trasformazioni ergodiche dell'intervallo. Nel confronto con gli altri due metodi di stima basati sulle successioni simboliche, esso è risultato piú accurato ed efficiente nell'approssimazione del valore di riferimento e nel tempo di esecuzione dell'algoritmo.

Avendo preso sempre come valore di riferimento per l'entropia quello dato dalla stima dell'esponente di Lyapunov, che converge velocemente e in maniera molto precisa, ci si potrebbe chiedere quale importanza possa rivestire la stima mediante il metodo di Grassberger. Se esso infatti semplicemente 'approssima' il valore che si ottiene stimando l'esponente di Lyapunov, non è forse conveniente utilizzare sempre quest'ultimo metodo? La risposta l'avevamo già accennata nell'*Introduzione*. Il calcolo o la stima dell'esponente di Lyapunov presuppongono che si sia in presenza di una trasformazione di qualche intervallo e che essa sia nota esplicitamente. Poiché gli altri tre metodi di stima -e in particolare quello di Grassberger- agiscono invece soltanto su processi simbolici, possono essere applicati ad una classe di casi piú ampia. Non di poca importanza a fini applicativi sono tutti quei casi in cui la sorgente non è nota e si rilevano empiricamente successioni simboliche.

Sopra abbiamo individuato un'interessante direzione di indagine riguardo agli aspetti teorici delle sostituzioni ricorsive di coppie simboliche. Un'altra direzione di studio, questa interessante soprattutto a fini applicativi e altrettanto importante, ci sembra essere quella della determinazione del numero piú opportuno di sostituzioni da eseguire a seconda del particolare processo in esame.

Appendice A

Risultati tecnici sul metodo NSRPS

Riportiamo in questa appendice le dimostrazioni dei risultati che abbiamo solo enunciato o di cui ci siamo serviti nel capitolo 3. Seguiamo [BCG] e nel corso di tutta l'appendice indicheremo con log il logaritmo in base 2.

A.1 Trasformazioni delle misure

Dimostrazione del teorema 3.2.2

Si fornisce qui di seguito una dimostrazione informale.

Il fatto che i limiti siano costanti quasi ovunque segue dalla legge forte dei grandi numeri. Ciò implica l'ergodicità di $\mathcal{G}\mu$ e di $\mathcal{S}\nu$ (vedi teorema I.4.2 pag. 44 di [Sh]). Le condizioni di compatibilità per le famiglie dei marginali si verificano facilmente. La formula (3.6) è conseguenza della (3.3).

Dimostrazione della proposizione 3.2.3

Nel caso $x \neq y$ si ha

$$|G(w_1^n)| = n - \#\{xy \subseteq w_1^n\}$$

cosí che

$$\frac{n}{|G(w_1^n)|} = \frac{1}{1 - \frac{\#\{xy \subseteq w_1^n\}}{n}}$$

e la (3.7) segue dalla legge forte dei grandi numeri. Nel caso $x = y$ si ha

$$|G(w_1^n)| = n - \sum_{k=2}^{n} \#\{*x^k* \subseteq w_1^n\} \left[\frac{k}{2}\right],$$

dove $[\cdot]$ denota la parte intera e $\#\{*x^k* \subseteq w_1^n\}$ è il numero dei blocchi di esattamente k simboli x consecutivi contenuti in w_1^n ($*$ rappresenta una possibile occorrenza di una lettera generica diversa da x). Vale

$$\#\{*x^k* \subseteq w_1^n\} = \#\{x^k \subseteq w_1^n\} - 2\#\{x^{k+1} \subseteq w_1^n\} + \#\{x^{k+2} \subseteq w_1^n\}.$$

Ora, si ha

$$\frac{n}{|G(w_1^n)|} = \frac{1}{1 - \sum_{k=2}^{n}(-1)^k \left(\frac{\#\{x^k \subseteq w_1^n\}}{n}\right)},$$

che converge al membro di destra della (3.8) per ogni misura ergodica μ diversa dalla misura concentrata sulla successione di tutti x (in tal caso, chiaramente $Z = 2$).

La formula (3.9) segue da

$$S(z_1^n) = n + \#\{\alpha \subseteq z_1^n\}$$

e dalla legge forte dei grandi numeri.

La formula (3.10) si deduce dalla (3.3).

$\mathcal{S}\nu$ in termini di ν

Si consideri la sostituzione $\alpha \to xy$. Si ha che

$$W = \lim_{n\to\infty} \frac{|S(z_1^n)|}{n} = \lim_{n\to\infty} \sum_{|\underline{z}|=n} \nu(\underline{z}) \frac{|S(\underline{z})|}{n},$$

e vale

$$\begin{aligned} \mathcal{S}\nu(\underline{r}) &\stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{\#\{\underline{r} \subseteq S(z_1^n)\}}{|S(z_1^n)|} = \lim_{n\to\infty} \frac{\#\{\underline{r} \subseteq S(z_1^n)\}}{Wn} \\ &= \lim_{n\to\infty} \frac{1}{Wn} \sum_{|\underline{z}|=n} \nu(\underline{z})\#\{\underline{r} \subseteq S(z_1^n)\}. \end{aligned} \tag{A.1}$$

Si supponga ora $|\underline{r}| = k$ e si consideri, per $n \geq k$,

$$\begin{aligned} D_n &= \sum_{|\underline{z}|=n} \nu(\underline{z})\#\{\underline{r} \subseteq S(\underline{z})\} - \sum_{|\underline{z}|=n-1} \nu(\underline{z}) \\ &= \sum_{|\underline{z}|=n} \nu(\underline{z})\mathbb{1}(\underline{r} = S(\underline{z})_1^k) + \sum_{|\underline{z}|=n-1} \nu(\alpha\underline{z})\mathbb{1}(\underline{r} = yS(\underline{z})_1^{k-1}). \end{aligned} \tag{A.2}$$

I due termini di questa somma possono essere riscritti come

$$\sum_{|\underline{z}|=n} \nu(\underline{z})\mathbb{1}(\underline{r} = S(\underline{z})_1^k) = \sum_{\underline{s}:S(\underline{s})=\underline{r}} \Big(\nu(\underline{s}) + \nu(s_1^{|\underline{s}|-1}\alpha)\mathbb{1}(r_k = x)\Big) \tag{A.3}$$

e

$$\begin{aligned}\sum_{|\underline{z}|=n-1} \nu(\alpha\underline{z})\mathbb{1}(\underline{r} = yS(\underline{z})_1^{k-1}) &= \\ &= \sum_{\underline{s}:S(\underline{s})=\underline{r}} \Big(\nu(\alpha s_2^{|\underline{s}|}) + \nu(\alpha s_2^{|\underline{s}|-1}\alpha)\mathbb{1}(r_k = x)\Big)\, \mathbb{1}(r_1 = y).\end{aligned} \tag{A.4}$$

Pertanto D_n è costante per $n \geq k$ e

$$\lim_{n\to\infty} \frac{1}{W} \sum_{|\underline{z}|=n} \nu(\underline{z})\#\{\underline{r} \subseteq S(z_1^n)\} = \frac{1}{W} D_k. \tag{A.5}$$

Mettendo insieme le (A.1)÷(A.5), si ottiene l'espressione per $\mathcal{S}\nu$:

$$\begin{aligned}\mathcal{S}\nu(\underline{r}) = \frac{1}{W} \sum_{\underline{s}:S(\underline{s})=\underline{r}} &\Big(\nu(\underline{s}) + \nu(s_1^{|\underline{s}|-1}\alpha)\mathbb{1}(r_k = x) + \nu(\alpha s_2^{|\underline{s}|})\mathbb{1}(r_1 = y) \\ &+ \nu(\alpha s_2^{|\underline{s}|-1}\alpha)\mathbb{1}(r_1 = y)\mathbb{1}(r_k = x)\Big).\end{aligned} \tag{A.6}$$

$\mathcal{G}\mu$ in termini di μ

Poiché la mappa S inverte G, per trovare l'espressione di $\mathcal{G}\mu$ si può invertire l'espressione di $\mathcal{S}\mathcal{G}\mu = \mu$. Sia ν la misura $\mathcal{G}\mu$.

La somma su $\underline{s}$ nell'equazione (A.6) si riduce a $\underline{s} = G(\underline{r})$, cioè $\nu(\underline{s}) = 0$ se $\underline{s} \notin G(A^*)$. Questa riduzione rende esplicitamente invertibile la (A.6), ma bisogna distinguere i casi $x \neq y$ e $x = y$.

Caso $x \neq y$

Sia $\underline{r} \in A^*$ e siano z, $w \in A$ tali che $z \neq x$ e $w \neq y$. Dalla (A.6) si ottiene:

$$\begin{aligned}&W\mu(w\underline{r}z) = \nu(G(w\underline{r}z)) \\ &W\mu(w\underline{r}x) = \nu(G(w\underline{r})x) + \nu(G(w\underline{r})\alpha) \\ &W\mu(y\underline{r}z) = \nu(yG(\underline{r}z)) + \nu(\alpha G(\underline{r}z)) \\ &W\mu(y\underline{r}x) = \nu(yG(\underline{r})x) + \nu(\alpha G(\underline{r})x) + \nu(yG(\underline{r})\alpha) + \nu(\alpha G(\underline{r})\alpha).\end{aligned} \tag{A.7}$$

Sia ora $\underline{s} = G(\underline{r})$ con $|\underline{s}| = n$ e $|\underline{r}| = k$. L'espressione di $\nu(\underline{s}) = \mathcal{G}\mu(\underline{s})$ può essere calcolata dalle equazioni (A.7), ottenendo

$$\begin{aligned}&s_1 \neq y,\ s_n \neq x, \quad \nu(\underline{s}) = W\mu(\underline{r}) \\ &s_1 = y,\ s_n \neq x, \quad \nu(\underline{s}) = W(\mu(\underline{r}) - \mu(xyr_2^k)) \\ &s_1 \neq y,\ s_n = x, \quad \nu(\underline{s}) = W(\mu(\underline{r}) - \mu(r_1^{k-1}xy)) \\ &s_1 = y,\ s_n = x, \quad \nu(\underline{s}) = W(\mu(\underline{r}) + \mu(xyr_2^{k-1}xy) - \mu(xyr_2^k) - \mu(r_1^{k-1}xy)).\end{aligned} \tag{A.8}$$

Ora si può calcolare $Z = W$ (cfr. equazione (3.10)) in termini di μ:

$$Z = 1 + \nu(\alpha) = 1 + Z\mu(xy) = \frac{1}{1-\mu(xy)}. \tag{A.9}$$

Si osservi infine che le equazioni (A.8) possono essere sintetizzate in

$$\mathcal{G}\mu(\underline{s}) = Z \sum_{a,b\in A: a\underline{s}b\in G(A^*)} \mu(a\underline{s}b). \tag{A.10}$$

Caso $x = y$
Procedendo come prima si ottendono espressioni esplicite per $\nu(\underline{s})$, le quali però sono piú complicate. Come prima, sia $\underline{s} \in G(A^*)$, $|\underline{s}| = n > 0$, $G(\underline{r}) = \underline{s}$, $|\underline{r}| = k$. Siano $s_1, s_n \neq x$. Indicando con $\underline{a}^p$ la stringa costituita da p simboli a consecutivi, le stringhe in $G(A^*)$ sono del tipo

$$\underline{\alpha}^p\underline{x}^\pi\underline{s}\underline{\alpha}^q\underline{x}^\sigma \text{ e } \underline{\alpha}^p\underline{x}^\pi, \quad \text{con } p,q \geq 0 \text{ e } \pi,\sigma = 0,1.$$

L'espressione di $\mathcal{G}\mu = \nu$ in termini di μ è data da:

$$\begin{array}{rlr}
\nu(\underline{s}\underline{\alpha}^q) & = Z\mu(\underline{r}\underline{x}^{2q}) & \text{per } q \geq 0 \\
\nu(\underline{s}\underline{\alpha}^q x) & = Z(\mu(\underline{r}\underline{x}^{2q+1}) - \mu(\underline{r}\underline{x}^{2q+2})) & \text{per } q \geq 0 \\
\nu(\underline{\alpha}^p) & = Z\sum_{j=0}^\infty (-1)^j \mu(\underline{x}^{2p+j}) & \text{per } p \geq 1 \\
\nu(\underline{\alpha}^p \underline{x}) & = Z(\mu(\underline{x}^{2p+1}) - 2\sum_{j=2}^\infty (-1)^j \mu(\underline{x}^{2p+j})) & \text{per } p \geq 1 \\
\nu(\underline{\alpha}^p\underline{x}^\pi\underline{s}\underline{\alpha}^q) & = Z\sum_{j=0}^\infty (-1)^j \mu(\underline{x}^{2p+\pi+j}\underline{r}\underline{x}^{2q}) & \text{per } p+\pi \geq 1, q \geq 0 \\
\nu(\underline{\alpha}^p\underline{x}^\pi\underline{s}\underline{\alpha}^q x) & = Z\sum_{j=0}^\infty (-1)^j \cdot & \\
& (\mu(\underline{x}^{2p+\pi+j}\underline{r}\underline{x}^{2q+1}) - \mu(\underline{x}^{2p+\pi+j}\underline{r}\underline{x}^{2q+2})) & \text{per } p+\pi \geq 1, q \geq 0
\end{array} \tag{A.11}$$

Ora è possibile calcolare Z in termini di μ:

$$Z = 1 + \nu(\alpha) = 1 + Z\sum_{j=0}^\infty (-1)^j \mu(\underline{x}^{2+j}) = \frac{1}{1-\sum_{j=2}^\infty (-1)^j \mu(\underline{x}^j)}. \tag{A.12}$$

Dimostrazione della proposizione 3.2.4

Questa proposizione è una conseguenza dell'equazione (A.6), cioè $|\underline{s}| \leq |\underline{r}|$ se $S(\underline{s}) = \underline{r}$.

Dimostrazione della proposizione 3.2.5

Questa proposizione è una conseguenza del fatto che l'espressione esplicita (A.6) di $\mu = \mathcal{S}\nu$ in termini di ν può essere invertita (in modo unico) se ν rispetta i vincoli di coppia dati da G, come segue dalle equazioni (A.7)÷(A.11). L'espressione di ν in termini di μ è esattamente $\mathcal{G}\mu$, e allora $\nu = \mathcal{G}\mu = \mathcal{G}\mathcal{S}\nu$.

A.2 Trasformazioni delle entropie

Dimostrazione del teorema 3.3.1

Il teorema segue dal fatto che G è un codice fedele e S è un codice fedele se ristretto al supporto di $\mathcal{G}\mu$. Si indichi con $C = \{C_n\}_{n\in\mathbb{N}}$ una successione di codici universali nell'alfabeto A e $C' = \{C'_n\}_{n\in\mathbb{N}}$ una successione di codici universali nell'alfabeto A' (si vedano i teoremi II.1.1 e II.1.2, pag. 122 di [Sh]). La composizione $C' \circ G$ è una successione di codici fedeli in A. Da ciò si deduce che su un insieme di misura uno secondo μ vale

$$h(\mathcal{G}\mu) = \lim_{n\to\infty} \frac{C'_{|G(w_1^n)|}(G(w_1^n))}{|G(w_1^n)|} = \lim_{n\to\infty} \frac{n}{|G(w_1^n)|} \frac{C'_{|G(w_1^n)|} \circ G(w_1^n)}{n} \geq Zh_1(\mu). \tag{A.13}$$

Similmente, anche $C \circ S$ è una successione di codici fedeli in A'. Da ciò si deduce che su un insieme di misura uno secondo μ vale

$$h(\mu) = \lim_{n\to\infty} \frac{C_n(w_1^n)}{n} = \lim_{n\to\infty} \frac{|G(w_1^n)|}{n} \frac{C_n \circ S(G(w_1^n))}{|G(w_1^n)|} \geq \frac{h(\mathcal{G}\mu)}{Z}. \tag{A.14}$$

Dimostrazioni dei teoremi 3.3.2 e 3.3.3

Nel corso di queste dimostrazioni, l'azione di G (e quindi di $\mathcal{G}$) verrà divisa in tre parti, mediante l'introduzione di due nuovi simboli $b1, b2 \notin A$. Data una stringa:

passo 1: si sostituirà, partendo dalla sinistra, ogni occorrenza di xy con xb_1. Questa operazione definisce una mappa $R : A^* \to A_R^*$, dove $A_R = A \cup \{b_1\}$. Si indicherà con $\mathcal{R}$ la corrispondente mappe tra le misure, definita con lo stesso spirito del teorema 3.2.2;

passo 2: si sostituirà ogni occorrenza di xb_1 con b_2b_1. Questa operazione definisce una mappa $L : A_R^* \to A_L^*$, dove $A_L = A_R \cup \{b_2\}$. Si indicherà con $\mathcal{L}$ la corrispondente mappe tra le misure;

passo 3: si sostituirà ogni occorrenza di b_2b_1 con α. Questa operazione definisce una mappa $C : A_L^* \to A_C^*$, dove $A_C = A_L \cup \{\alpha\}$. Si indicherà con $\mathcal{C}$ la corrispondente mappe tra le misure.

Da queste definizioni si ha

$$C(L(R(\underline{w}))) = G(\underline{w}), \text{ e quindi } \mathcal{CLR}\mu = \mathcal{G}\mu.$$

Per mezzo di questa fattorizzazione di G, vengono separati gli effetti dell'accorciamento delle stringhe (passo 3) dall'effetto delle sostituzioni parziali dei simboli (passi 1 e 2).

Lemma A.2.1.

$$h_1(\mathcal{R}\mu) \leq h_1(\mu) \tag{A.15}$$

La dimostrazione del lemma si trova poco oltre.

La medesima disuguaglianza vale anche per $\mathcal{LR}\mu$. Infatti, si può definire L anche considerando le sostituzioni iniziando da destra, con $x \neq b_1$. Cosí, $L(\underline{w}) = (R'(\underline{w}^r))^r$, dove $\underline{w}^r = (w_1 \dots w_k)^r = w_k \dots w_1$ e R' è la sostituzione, da sinistra, di $b_1 x$ con $b_1 b_2$. La mappa R' agisce nello stesso modo di R, e quindi il lemma A.2.1 vale per la corrispondente mappa tra le misure $\mathcal{R}'$, e quindi anche per $\mathcal{L}$. In questo modo si prova che

$$h_1(\mathcal{LR}\mu) \leq h_1(\mu).$$

Il terzo passo lascia invariata h_1, a meno della normalizzazione, come asserito nel seguente lemma (la dimostrazione si trova poco piú giú).

Lemma A.2.2. *Se $\rho \in \mathcal{E}(A_L)$ soddisfa*

$$\rho(b_2 w) = \rho(z b_1) = 0 \textit{ per } w \neq b_1, z \neq b_2, \tag{A.16}$$

allora

$$h_1(\mathcal{C}\rho) = W h_1(\rho), \tag{A.17}$$

dove

$$W = \frac{1}{1 - \rho(b_2 b_1)} = 1 + \mathcal{C}\rho(\alpha). \tag{A.18}$$

La dimostrazione del teorema 3.3.2 si conclude osservando che la misura $\rho = \mathcal{LR}\mu$ verifica le condizioni (A.16), per cui $h_1(\mathcal{G}\mu) \leq W h_1(\mu)$, dove $W = Z$ perché $W = 1 + \mathcal{C}\rho(\alpha) = 1 + \mathcal{G}\mu(\alpha) = W_\alpha^{\mathcal{G}\mu} = Z_{xy}^\mu$ (cfr. equazione (3.10)).

Per dimostrare il teorema 3.3.3 è sufficiente osservare che il lemma A.2.1 vale anche per h_k e che per h_k vale il seguente analogo del lemma A.2.2, che sarà dimostrato piú giú.

Lemma A.2.3. *Sotto le medesime ipotesi del lemma A.2.2, vale*

$$h_k(\mathcal{C}\rho) = W h_k(\rho). \tag{A.19}$$

Dimostrazione del lemma A.2.1

Sia $\xi = \mathcal{R}\mu$. La misura μ può essere espressa in termini di ξ nel modo seguente:

$$\mu(\underline{w}) = \sum_{\underline{z}:R(\underline{z})=\underline{w}} \xi(\underline{w}).$$

Questa formula verrà usata ora per esprimere le probabilità dei simboli e delle coppie di simboli.

Caso $x \neq y$

Sia $p \in A$:

$$\begin{array}{lll}
\mu(y) = \xi(y) + \xi(b_1), & \mu(p) = \xi(p) & \text{per } p \neq y, \\
\mu(yp) = \xi(yp) + \xi(b_1p), & \mu(pq) = \xi(pq) & \text{per } p \neq x, p \neq y \\
\mu(xy) = \xi(xb_1), & \mu(xp) = \xi(xp) & \text{per } p \neq y.
\end{array}$$

Da un calcolo diretto segue

$$\begin{aligned}
h_1(\mu) - h_1(\xi) = &- \sup_{p\in A}(\xi(yp) + \xi(b_1p)) \log \frac{\xi(yp) + \xi(b_1p)}{\xi(y) + \xi(b_1)} \\
&+ \sum_{p\in A} \left(\xi(yp) \log \frac{\xi(yp)}{\xi(y)} + \xi(b_1p) \log \frac{\xi(b_1p)}{\xi(b_1)} \right).
\end{aligned} \tag{A.20}$$

Il lemma sarà provato dimostrando che vale

$$\left(\xi(yp) \log \frac{\xi(yp)}{\xi(y)} + \xi(b_1p) \log \frac{\xi(b_1p)}{\xi(b_1)} \right) \geq (\xi(yp) + \xi(b_1p)) \log \frac{\xi(yp) + \xi(b_1p)}{\xi(y) + \xi(b_1)}.$$

Dividendo per $\xi(yp)+\xi(b_1p)$ e ponendo $\beta = \frac{\xi(y)}{\xi(y)+\xi(b_1)}$, $\gamma = \frac{\xi(yp)}{\xi(yp)+\xi(b_1p)}$, l'ultima disuguaglianza si può riscrivere come

$$\gamma \log \frac{\beta}{\gamma} + (1-\gamma) \log \frac{1-\beta}{1-\gamma} \leq 0,$$

che è sempre verificata.

Caso $x = y$

Sia $p \in A$, $p \notin x$... $p \in A, p \notin x$:

$$\begin{array}{ll}
\mu(x) = \xi(x) + \xi(b_1), & \mu(p) = \xi(p), \\
\mu(xx) = \xi(xb_1) + \xi(b_1x), & \mu(xp) = \xi(xp) + \xi(b_1p), \\
\mu(pq) = \xi(pq) \text{ per } q \in A, & \mu(px) = \xi(px).
\end{array}$$

La differenza tra le entropie condizionate di ordine 1 è

$$\begin{aligned}
h_1(\mu) - h_1(\xi) = &- \sum_{\substack{p\in A\\ p\neq x}} (\xi(xp)+\xi(b_1p)) \log \frac{\xi(xp)+\xi(b_1p)}{\xi(x)+\xi(b_1)} \\
&- (\xi(xb_1)+\xi(b_1x)) \log \frac{\xi(xb_1)+\xi(b_1x)}{\xi(x)+\xi(b_1)} \\
&+ \sum_{\substack{p\in A\\ p\neq x}} \left(\xi(xp) \log \frac{\xi(xp)}{\xi(x)} + \xi(b_1p) \log \frac{\xi(b_1p)}{\xi(b_1)} \right) \\
&+ \xi(xb_1) \log \frac{\xi(xb_1)}{\xi(x)} + \xi(b_1x) \log \frac{\xi(b_1x)}{\xi(b_1)}.
\end{aligned}$$

Il fatto che tale differenza sia positiva si dimostra argomentando come nel caso $x \neq y$.

Si osservi infine che nella medesima maniera si può dimostrare che $h_k(\xi) \leq h_k(\mu)$.

Dimostrazione del lemma A.2.2

Siano $\nu = \mathcal{C}\rho$ e $W = 1 + \nu(\alpha)$. È facile scrivere ρ in termini di ν. Siano $p, q \neq b_1, b_2$. Le probabilità dei simboli e delle coppie di simboli sono date da:

$$\begin{array}{ll}
W\rho(b_1) = W\rho(b_2) = \nu(\alpha) & W\rho(p) = \nu(p) \\
W\rho(pb_1) = W\rho(b_2q) = 0 & W\rho(pq) = \nu(pq) \\
W\rho(pb_2) = \nu(p\alpha) & W\rho(b_1q) = \nu(\alpha q) \\
W\rho(b_2b_1) = \nu(\alpha) & W\rho(b_1b_2) = \nu(\alpha\alpha).
\end{array}$$

Tramite calcolo diretto si trova:

$$\begin{aligned}
H_1(\rho) &= - \sum_{p\in A_C\setminus\alpha} \frac{\nu(p)}{W} \log \frac{\nu(p)}{W} - 2\frac{\nu(\alpha)}{W} \log \frac{\nu(\alpha)}{W} \\
&= \frac{H_1(\nu)}{W} + \frac{\log W}{W} - \frac{\nu(\alpha)}{W} \log \frac{\nu(\alpha)}{W}, \\
H_2(\rho) &= - \sum_{p,q\in A_C} \frac{\nu(pq)}{W} \log \frac{\nu(pq)}{W} - \frac{\nu(\alpha)}{W} \log \frac{\nu(\alpha)}{W} \\
&= \frac{H_2(\nu)}{W} + \frac{\log W}{W} - \frac{\nu(\alpha)}{W} \log \frac{\nu(\alpha)}{W}.
\end{aligned}$$

Pertanto,

$$h_1(\rho) = H_2(\rho) - H_1(\rho) = \frac{h_1(\nu)}{W}.$$

Dimostrazione del lemma A.2.3

Prima di tutto, ecco alcune definizioni. Sia $\underline{w} = w_1^l$. Si può identificare $\underline{w}$ con il sottoinsieme cilindrico $K_{\underline{w}} \subseteq A^{\mathbb{Z}}$ definito da

$$K_{\underline{w}} = \{\underline{x} \in A^{\mathbb{Z}} : z_{-l} = w_1, x_{-l+1} = w_2, \ldots, x_{-1} = w_l\}.$$

Sia $P \subseteq A^*$ un insieme finito. Si dice che P è una partizione se

$$\{K_{\underline{w}}\}_{\underline{w} \in P} \text{ è una partizione di } A^{\mathbb{Z}}, \text{ cioè } \begin{cases} (1) & K_{\underline{w}} \cap K_{\underline{z}} = \varnothing \text{ se } \underline{w} \neq \underline{z}, \\ (2) & \bigcup_{\underline{w} \in P} K_{\underline{w}} = A^{\mathbb{Z}} \end{cases}$$

La condizione (1) dice che nessuna stringa di P è un suffisso di un'altra stringa di P. Se solo la condizione (1) è verificata, P è detta una semi-partizione. Si vede facilmente che ogni semi-partizione si può completare per ottenere una partizione. Inoltre, se il minimo delle lunghezze delle stringhe in P è l, si può completare P usando stringhe di lunghezza maggiore di o uguale a l.

Se P è una partizione, si definisce l'*entropia P-condizionata* come

$$h_P(\mu) = -\sum_{\substack{\underline{w} \in P \\ a \in A}} \mu(\underline{w}a) \log \frac{\mu(\underline{w}a)}{\mu(\underline{w})}.$$

Se P e Q sono due partizioni, si dice che P è piú fine di Q se ogni stringa di P termina con una stringa di Q. Se P è piú fine di Q, allora vale la formula (la dimostrazione è alla fine della sezione)

$$h_P(\mu) \leq h_Q(\mu). \tag{A.21}$$

Si noti che

$$P = \left\{\underline{s} \in A_L^* \middle| |C(\underline{s})| = k\right\}$$

è una semi-partizione, e che mediante calcolo diretto si ottiene

$$h_k(\nu) = W h_P(\rho),$$

dove $\nu = \mathcal{C}\rho$. In particolare, si è usato che se $\underline{s} \in A_L^*$ allora $\rho(\underline{s}b_2) = \rho(\underline{s}b_2b_1)$ e che se l'ultimo simbolo di $\underline{s}$ è diverso da b_2 allora $\rho(\underline{s}b_1) = 0$. Sia infine $\overline{P}$ un completamento di P; allora

$$h_k(\nu) = W h_P(\rho) \leq W h_{\overline{P}}(\rho).$$

La lunghezza delle stringhe in P è maggiore di o uguale a k e si può costruire $\overline{P}$ in modo che lo stesso valga per $\overline{P}$. Per la (A.21) si ha in conclusione che

$$h_k(\nu) \leq W h_{A_L^k}(\rho) = W h_k(\rho).$$

Dimostrazione della disuguaglianza (A.21)

Sia $\underline{w} \in Q$ e sia $X_{\underline{w}} \subseteq P$ il sottoinsieme delle stringhe che terminano con $\underline{w}$. Da questa definizione:

$$P = \bigcup_{\underline{w} \in Q} X_{\underline{w}}, \quad \mu(\underline{w}) = \sum_{\underline{r} \in X_{\underline{w}}} \mu(\underline{r}).$$

La funzione $\Phi(x) = x \log x$ è convessa, allora se $\lambda_i \geq 0$ e $\sum \lambda_i = 1$, si ha $\Phi(\sum \lambda_i x_i) \leq \sum \lambda_i x_i \log x_i$. Ora,

$$-h_Q(\mu) = \sum_{\underline{w} \in Q} \mu(\underline{w}) \sum_{a \in A} \mu(a|\underline{w}) \log \mu(a|\underline{w}),$$

e

$$\mu(a|\underline{w}) = \frac{\mu(\underline{w}a)}{\mu(\underline{w})} = \sum_{\underline{r} \in X_{\underline{w}}} \frac{\mu(\underline{r}a)}{\mu(\underline{w})} = \sum_{\underline{r} \in X_{\underline{w}}} \frac{\mu(\underline{r}a)}{\mu(\underline{r})} \frac{\mu(\underline{r})}{\mu(\underline{w})}.$$

Indicando con $x_{\underline{r}}^a = \mu(\underline{r}a)/\mu(\underline{r})$ e con $\lambda_{\underline{r}} = \mu(\underline{r})/\mu(\underline{w})$ e osservando che $\sum_{\underline{r} \in X_{\underline{w}}} \lambda_{\underline{r}} = 1$, si ottiene

$$\begin{aligned}
-h_Q(\mu) &= \sum_{\underline{w} \in Q} \sum_{a \in A} \mu(\underline{w}) \Phi\left(\sum_{\underline{r} \in X_{\underline{w}}} \lambda_{\underline{r}} x_{\underline{r}}^a \right) \\
&\leq \sum_{\underline{w} \in Q} \sum_{\underline{r} \in X_{\underline{w}}} \sum_{a \in A} \mu(\underline{w}) \frac{\mu(\underline{r})}{\mu(\underline{w})} \mu(a|\underline{r}) \log \mu(a|\underline{r}) \\
&= \sum_{\underline{r} \in P} \sum_{a \in A} \mu(\underline{r}) \mu(a|\underline{r}) \log \mu(a|\underline{r}) = -h_P(\mu).
\end{aligned}$$

Bibliografia

[Ab] M. Abate, *Sistemi Dinamici Discreti*, dispense del corso tenuto presso l'Università di Pisa nell'A.A. 2005-2006. Su Internet in `http://www.dm.unipi.it/~abate/matdid/dispense/` (22 novembre 2007).

[BCG] D. Benedetto, E. Caglioti, D. Gabrielli, *Non sequential recursive pair substitution: some rigorous results*, Journal of Statistical Mechanics (2006) P09011.

[Bil] P. Billingsley, Probability and measure, Wiley, New York, 1979.

[Bir] G. D. Birkhoff, Proof ot the ergodic theorem, Proceedings of the National Academy of Sciences USA **17** (1931), 656-660.

[Br] L. Breiman, *Probability theory*, Addison-Wesley, Reading, MA, 1968.

[Bu] J. Buzzi, *A minicourse on entropy theory on the interval*, `arXiv:math/0611337v1 [math.DS]`, 2006.

[Ch] G. H. Choe, *Computational ergodic theory*, Springer-Verlag, Berlin, 2005.

[EJM] W. Ebeling, M. A. Jiménez-Montaño, *On grammars, complexity, and information measures of biological macromolecules*, Mathematical Biosciences **52** (1980), 53-71.

M. A. Jiménez-Montaño, *On the syntactic structure of protein sequences and the concept of grammar complexity*, Bulletin of Mathematical Biology **42** (1984), 641-659.

M. A. Jiménez-Montaño, W. Ebeling, T. Pöschel, *SYNTAX: a computer program to compress a sequence and to estimate its information content*, `arXiv:cond-mat/0204134`, 2002.

P. E. Rapp, I. D. Zimmermann, E. P. Vining, N. Cohen, A. M. Albano, M. A. Jiménez-Montaño, *Tha algoritmic complexity of neural spike trains increases during focal seizures*, The Journal of Neuroscience **14** (1994), 4731-4739.

[Fe] W. Feller, *An introduction to probability theory and its applications* (vol. I, 3a ediz.), Wiley, New York, 1968.

[Gr] P. Grassberger, *Data compression and entropy estimates by non-sequential recursive pair substitution*, `arXiv:physics/0207023`, 2002.

[Is] S. Isola, *On systems with finite ergodic degree*, Far East Journal of Dynamical Systems **5** (2003), 1-62.

[KH] A. Katok, B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, 1995.

[Kh] A. Ya. Khintchine, *Mathematical Foundations of Information Theory*, Dover, New York, 1957.

[Ko] A. N. Kolmogorov, *A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces*, Doklady Akademii Nauk SSSR **119** (1958), 861-864. (In russo.)

[Kr] W. Krieger, *On entropy and generators of measure-preserving transformations*, Transactions of the American Mathematical Society **149** (1970), 453-464.

[Ma] R. Mañé, *Ergodic theory and differentiable dynamics*, Springer-Verlag, Berlin, 1987.

[OW] D. Ornstein, B. Weiss, *How sampling reveals a process*, The Annals of Probability **18** (1990), 905-930.

[Pa] W. Parry, *Entropy and generators in ergodic theory*, Benjamin, New York, 1969.

[Pe] K. Petersen, *Ergodic theory*, Cambridge University Press, 1983.

[PM] Y. Pomeau, P. Manneville, *Intermittent transition to turbulence in dissipative dynamical systems*, Communications in Mathematical Physics **74** (1980), 189-197.

[Sh] P. C. Shields, *The ergodic theory of discrete sample paths*, American Mathematical Society, 1996.

[Si] Ya. G. Sinai, *On the notion of entropy of a dynamical system*, Doklady Akademii Nauk SSSR **124** (1959), 768-771. (In russo.)

[Th] M. Thaler, *Estimates of the invariant densities of endomorphisms with indifferent fixed points*, Israel Journal of Mathematics **37** (1980), 303-314.

[Vi] M. Viana, *Stochastic dynamics of deterministic systems*, Lecture Notes XXI Brazilian Mathematics Colloquium, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1997. Su Internet in `http://www.impa.br/~viana/` (22 novembre 2007).

[Wa1] P. Walters, *An introduction to ergodic theory*, Springer-Verlag, New York, 1982.

[Wa2] P. Walters, *Invariant measures and equilibrium states for some mappings which expand distances*, Transactions of the American Mathematical Society **236** (1978), 121-153.

[WZ] A. D. Wyner, J. Ziv, *Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression*, IEEE Transactions on Information Theory **35** (6) (1989), 1250-1258.