



UNIVERSITÀ DI TRIESTE

Dipartimento di Matematica, Informatica e Geoscienze

Laurea Triennale in Intelligenza Artificiale e Data Analytics

TESI DI LAUREA

**Stima entropica del coefficiente di Lyapunov da stringhe binarie
associate a mappe dinamiche**

Relatore:

Prof. Fabio Anselmi

Correlatore:

Prof. Fabio Benatti

Laureando:

Tommaso Mingrone

A. A. 2023/2024

Abstract

Questa tesi propone uno studio sulla comprimibilità delle stringhe binarie generate da mappe dinamiche, basato sulla stima dell'entropia di Ziv, con un'attenzione particolare alla Arnold's Cat Map.

Il metodo utilizzato calcola il coefficiente di Lyapunov stimando la crescita dell'entropia H_d al variare di d , stimata dalla frequenza delle sottostringhe di lunghezza d all'interno di una stringa binaria associata alla dinamica descritta dalla mappa.

I risultati ottenuti per la mappa di Arnold mostrano che questo coefficiente consente di distinguere tra dinamiche caotiche e regolari.

L'algoritmo è stato inoltre testato sulla mappa di Pomeau–Manneville, un sistema che, a differenza della mappa di Arnold, presenta una dinamica intermittente in cui fasi di comportamento regolare si alternano a fasi di caos. L'analisi ha rivelato che la crescita dell'entropia delle sottostringhe non segue un andamento lineare rispetto alla loro lunghezza.

Questo apre la possibilità di classificare i sistemi dinamici in base al loro tasso di crescita dell'entropia, offrendo nuove prospettive nell'analisi della complessità dinamica.

Indice

Abstract	i
Indice	iii
Elenco delle figure	v
Elenco delle tabelle	vii
1 Introduzione	1
1.1 Motivazione e obiettivi	1
1.2 Panoramica della Ziv Entropy e delle sfide	3
1.3 Stringhe binarie ed entropia	4
1.4 Arnold's cat dataset: comportamento caotico e non caotico	6
1.4.1 Arnold's cat map	9
1.4.2 Generazione del dataset	9
2 Metodologia	11
2.1 Metodo basato sul conteggio delle frequenze	12
2.2 Costruzione della distribuzione delle sottostringhe	13
2.3 Calcolo diretto dell'entropia	13
2.4 Stima del coefficiente di Lyapunov tramite la crescita dell'entropia	14
3 Risultati e discussione	17
3.1 Scalabilità e impatto della dimensione delle sottostringhe sulla stima dell'entropia e del coefficiente di Lyapunov	22
3.2 Media dell'entropia su campioni estesi di mappe caotiche e regolari	24
4 Conclusioni e studi futuri	27
Bibliografia	31

Elenco delle figure

1.1	Confronto tra traiettorie caotiche regolari o non caotiche.	6
1.2	Percorso generato dalla Arnold's Cat Map.	10
2.1	Andamento del rapporto $\frac{H(d)}{d}$ per d da 2 a 100 nei due tipi di sistemi: caotico (a sinistra) e regolare o non caotico (a destra).	15
3.1	Andamento di $H(d)$ e del fit lineare per 10 diverse matrici caotiche.	20
3.2	Andamento di $H(d)$ e del fit lineare per 10 diverse matrici regolari o non caotiche.	21
3.3	$H(d)$ per $d = [2; 10], N = 1000$	22
3.4	$H(d)$ per $d = [2; 20], N = 20000$	22
3.5	$H(d)$ per $d = [2; 50], N = 50000$	22
3.6	$H(d)$ per $d = [2; 100], N = 100000$	22
3.7	$H(d)$ per $d = [2; 200]$ e $N = 500000$	23
3.8	Andamento medio dell'entropia $H(d)$ all'aumentare di d per sistemi caotici e regolari, calcolato su 50 mappe per ciascun tipo.	24
4.1	Andamento dell'entropia $H(d)$ per la Arnold's Cat Map: la curva mostra un plateau per valori elevati di d	29
4.2	Esempio per la mappa di Pomeau–Manneville: la curva non rivela un plateau nello stesso intervallo di d	29

Elenco delle tabelle

3.1	Entropia media per sistemi caotici e regolari al variare della dimensione d .	25
-----	---	----

Capitolo 1

Introduzione

La compressione di stringhe riduce lo spazio necessario per la memorizzazione e la trasmissione dei dati, permettendo un utilizzo più efficiente delle risorse digitali.

Storicamente, la compressione dati trova le sue radici nella teoria dell'informazione, sviluppata da Claude Shannon negli anni '40, che introdusse concetti fondamentali come l'entropia per quantificare l'informazione contenuta in una sorgente. Negli anni '70, algoritmi di compressione come Huffman Coding e Lempel-Ziv ne hanno rivoluzionato il campo, fornendo strumenti pratici per comprimere dati in modo efficiente e senza perdita.

Ad oggi, con la diffusione di grandi dataset e sistemi sempre più complessi, si applicano algoritmi di ottimizzazione per identificare pattern ricorrenti, analizzare strutture nei dati binari e ottimizzarne la compressione.

1.1 Motivazione e obiettivi

Studiare un sistema caotico è complesso poiché, nonostante la sua natura deterministica, presenta una forte sensibilità alle condizioni iniziali [6][12]. Questa caratteristica amplifica anche piccole variazioni, rendendo difficile prevedere l'evoluzione a lungo termine e identificare strutture ricorrenti nel comportamento apparentemente casuale.

L'entropia quantifica il disordine di un sistema e consente di identificare pattern nelle sue traiettorie. Nei sistemi discreti rappresentati da stringhe binarie, la complessità può essere stimata analizzando la distribuzione dei pattern di lunghezza variabile.

Questo studio propone un metodo per la stima del coefficiente di Lyapunov basato sulla Ziv Entropy, una misura derivata dall'entropia, applicata a sequenze binarie. L'approccio si basa sulla stima delle frequenze di occorrenza delle sottosequenze di lunghezza variabile all'interno delle stringhe generate. Per ciascuna lunghezza del pattern, vengono conteggiati i sottosegmenti distinti e calcolata la loro distribuzione di probabilità. Dopodiché l'entropia di Shannon quantifica il disordine della distribuzione, distinguendo tra sequenze altamente imprevedibili e strutture periodiche o regolari.

Per comprendere al meglio i successivi teoremi, è necessario definire il concetto di processo ergodico[2]. Si tratta di un processo casuale in cui osservando una singola evoluzione per un tempo sufficientemente lungo, si ottengono gli stessi risultati che si avrebbero calcolando la media su tante evoluzioni diverse del sistema. In altre parole, invece di dover analizzare molte versioni dello stesso processo, basta seguirne una per un tempo abbastanza lungo per capire il suo comportamento generale.

Dunque l'efficacia di questo metodo è supportata dal Teorema di Entropia di Ziv, Teorema 1.1.2, che stabilisce che, per un processo ergodico, se l'entropia del processo è h , la stima ottenuta, non può asintoticamente superare tale valore. A supporto di tale teoria vi è anche il Teorema 1.1.1.

Teorema 1.1.1 (Limite superiore di Lempel-Ziv[11]). Sia $X = \{X_t\}_{t \in \mathbb{N}}$ un processo stazionario ed ergodico con entropia h . Sia $C(X^n)$ la lunghezza della codifica di Lempel-Ziv per una sequenza di lunghezza n estratta da X . Sia inoltre A un alfabeto finito e definiamo $A^\infty = \{x = (x_1, x_2, x_3, \dots) \mid x_i \in A, \forall i \in \mathbb{N}\}$ come l'insieme di tutte le sequenze infinite su A . Allora, $\forall x \in A^\infty$:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} C(X^n) \log n \leq H(x), \quad \forall x \in A^\infty.$$

Ciò implica che la crescita della complessità compressiva di Lempel-Ziv è limitata superiormente dall'entropia della sorgente, ovvero la fonte informativa.

Teorema 1.1.2 (Teorema di Entropia di Ziv[11]). Se μ è un processo ergodico con entropia h , allora $H(x) \leq h$, quasi sicuramente.

Questo risultato, combinato con il Teorema del limite superiore di Lempel-Ziv, dimostra che nel caso di processi ergodici non è possibile ottenere una compressione più efficiente rispetto al limite teorico imposto dall'entropia.

In particolare, questi due risultati conducono alla Convergenza della complessità di Lempel-Ziv all'entropia della sorgente ergodica, garantendo che la stima dell'entropia ottenuta tramite compressione non superi il valore teorico h per processi ergodici.

Teorema 1.1.3 (Convergenza di Lempel-Ziv [11]). Sia $X = \{X_t\}_{t \in \mathbb{N}}$ un processo stazionario ed ergodico con entropia h . Sia $C(X^n)$ la lunghezza della codifica di Lempel-Ziv per una sequenza di lunghezza n estratta dal processo X . Allora:

$$\lim_{n \rightarrow \infty} \frac{1}{n} C(X^n) = h, \quad \text{quasi sicuramente.}$$

Quando $n \rightarrow \infty$, la complessità di Lempel-Ziv cresce in base all'entropia del sistema. Questo indica che il metodo di compressione fornisce una stima sempre più precisa dell'entropia effettiva.

Dal momento che la compressione Lempel-Ziv identifica pattern ripetuti all'interno della sequenza, la sua complessità misura la ridondanza dei dati. La convergenza alla soglia teorica dell'entropia implica che la quantità di informazione presente nella sequenza possa essere determinata attraverso metodi di compressione senza perdita, indipendentemente dalla distribuzione dei dati.

Nel caso preso in considerazione, la relazione tra la lunghezza del pattern e l'entropia viene studiata attraverso una regressione lineare, con l'obiettivo di stimare il coefficiente di Lyapunov, indicatore della tipologia di sistema. Siccome la Ziv Entropy dipende dalla distribuzione delle sottosequenze, la sua variazione riflette direttamente la complessità strutturale del sistema preso in esame.

1.2 Panoramica della Ziv Entropy e delle sfide

La Ziv Entropy è una misura della complessità di una sequenza discreta basata sulla distribuzione dei pattern distinti e sulla loro frequenza di occorrenza. Lempel e Ziv hanno dimostrato che, per processi ergodici, la complessità di compressione converge asintoticamente all'entropia della sorgente. Questo è garantito dai teoremi 1.1.1 e 1.1.2, introdotti precedentemente.

La Ziv Entropy viene, dunque, definita come il limite asintotico del rapporto tra la lunghezza della versione compressa di una sequenza e la sua lunghezza originale [13]:

$$H(S) = \lim_{n \rightarrow \infty} \frac{C(S_n) \log n}{n \log \alpha}$$

dove $C(S_n)$ indica la complessità della sequenza secondo il metodo di Lempel-Ziv, n è la sua lunghezza e α è la dimensione dell'alfabeto. Nel caso specifico di stringhe binarie (0,1), $\alpha = 2$.

L'algoritmo adottato si basa sulla Ziv Entropy per stimare la crescita dell'entropia rispetto alla lunghezza del pattern, sfruttando tale relazione tramite regressione lineare per ottenere il coefficiente di Lyapunov, indicatore della caoticità del sistema.

Nonostante la solidità teorica del metodo, la sua applicazione può presentare diverse criticità. La stima dell'entropia dipende dalla lunghezza della sequenza analizzata: dataset limitati possono introdurre distorsioni nella frequenza dei pattern, compromettendo la convergenza della misura. Inoltre, la presenza di rumore o fluttuazioni locali può alterare la distribuzione delle sottosequenze, portando a sovrastime o sottostime dell'entropia effettiva.

Un ulteriore aspetto riguarda la selezione della lunghezza dei pattern. Se troppo brevi, i pattern catturano solo variazioni locali senza riflettere la struttura globale della sequenza. Se troppo lunghi, riducono il numero di campioni utili per la stima della distribuzione, influenzando negativamente la stabilità dell'entropia calcolata.

Infine, l'affidabilità della regressione lineare utilizzata per stimare il coefficiente di Lyapunov dipende dalla qualità della misura dell'entropia. L'uso della compressione per approssimare l'entropia della sorgente si basa sulla validità della convergenza teorica dimostrata da Ziv. Tuttavia, con l'utilizzo di un compilatore, la convergenza può essere influenzata dal fatto che la memoria è finita, rendendo necessaria un'attenta calibrazione dei parametri del metodo per garantire una stima accurata della complessità del sistema.

1.3 Stringhe binarie ed entropia

Le stringhe binarie considerate sono sequenze composte esclusivamente da 0 e 1. Ogni stringa rappresenta quindi una serie di stati discreti che possono codificare informazioni o descrivere l'evoluzione del sistema. L'entropia di Shannon [7] è utilizzata per quantificare il grado di incertezza o complessità di una distribuzione di probabilità associata a una stringa binaria ed è definita come:

$$H = - \sum_{i=1}^n p_i \log(p_i),$$

dove:

- p_i è la probabilità associata al i -esimo valore della variabile aleatoria, con $p_i \geq 0$ e $\sum_{i=1}^n p_i = 1$;
- n è il numero totale di possibili valori distinti che la variabile aleatoria può assumere.

Nel caso specifico preso in analisi, p_i è la probabilità associata al pattern i -esimo nella distribuzione, calcolata come la frequenza relativa di occorrenza nella stringa analizzata; mentre n rappresenta il numero totale di pattern. Un'entropia alta indica una stringa casuale e non prevedibile, mentre un'entropia bassa suggerisce una struttura intrinseca o pattern ricorrenti, rendendo la stringa più comprimibile.

Poiché la comprimibilità di una sequenza è direttamente collegata alla sua entropia, è possibile dimostrare i risultati dei teoremi 1.1.1 e 1.1.2.

Dimostrazione. 1.1.1: Sia \mathcal{S}_d l'insieme di tutte le sottostringhe di lunghezza d estratte dalla sequenza x^n , con $p(s)$ la probabilità di occorrenza della sottostringa s . Dalla definizione di entropia di Shannon, si ha:

$$H_d(x) = - \sum_{s \in \mathcal{S}_d} p(s) \log p(s).$$

Poiché l'entropia media di un processo ergodico con distribuzione stazionaria è definita come il limite per $d \rightarrow \infty$:

$$h = \lim_{d \rightarrow \infty} \frac{H_d(x)}{d},$$

Ziv ha dimostrato che la complessità di Lempel-Ziv, la quale misura la lunghezza minima della rappresentazione compressa della sequenza, è legata a $H_d(x)$ dalla relazione:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} C(x^n) \log n \leq h.$$

La dimostrazione si basa su tre lemmi:

- Limite grezzo: stabilisce un vincolo superiore alla crescita della complessità compressiva nel caso peggiore.
- Limite entropico topologico: dimostra che il numero di parole distinte di lunghezza k fornisce un vincolo superiore all'entropia della sequenza.
- Limite di perturbazione: mostra che l'entropia topologica di una sequenza vicina fornisce un limite superiore alla complessità del codice.

Poiché la compressione Lempel-Ziv identifica pattern ripetuti e la struttura informativa all'interno della sequenza, la sua complessità fornisce una misura della ridondanza nei dati. La convergenza al valore teorico dell'entropia implica che la quantità di informazione presente nella sequenza possa essere determinata attraverso metodi di compressione senza perdita, indipendentemente dalla distribuzione della sorgente. Questo risultato conclude la dimostrazione. \square

Dimostrazione. 1.1.2: Sia $\epsilon > 0$ e consideriamo un insieme $T_k \subset A^k$ di cardinalità minore di $2^{k(h+\epsilon)}$, tale che la sua misura sia maggiore di $1 - \epsilon$. Sia x una sequenza tipica per la frequenza. L'idea della dimostrazione è che la maggior parte degli $(i + k - 1)$ -esimi blocchi della sequenza x appartengano a T_k per un numero limitato di indici i . Poiché lo stesso deve valere per blocchi non sovrapposti, esiste almeno un valore $s \in [0, k - 1]$ per cui il numero di blocchi non appartenenti a T_k può essere ridotto sostituendoli con un blocco fisso, ottenendo una sequenza y vicina a x con entropia topologica prossima a h . Consideriamo:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left| \{i \in [0, n - 1] : x_{i+1}^{i+k} \in T_k\} \right| > 1 - \epsilon.$$

Quindi esiste un intero $s \in [0, k - 1]$ tale che:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \left| \{i \in [0, n - 1] : x_{i+s+1}^{i+s+k} \in T_k\} \right| > 1 - \epsilon.$$

Definiamo y come la concatenazione della sequenza iniziale u e dei blocchi $v(i)$, con:

$$v(i) = \begin{cases} x_{i+s+1}^{i+s+k}, & \text{se } x_{i+s+1}^{i+s+k} \in T_k, \\ a^k, & \text{altrimenti,} \end{cases}$$

dove a^k è una sequenza di lunghezza k contenente solo il simbolo a . La costruzione garantisce che:

$$\bar{d}(x, y) < \epsilon.$$

Poiché l'universo dei blocchi di y contiene al massimo $1 + 2^{k(h+\epsilon)}$ elementi, applicando il lemma della costruzione di insiemi, otteniamo:

$$h_m(y) < h + \epsilon + o(m),$$

con $o(m) \rightarrow 0$ per $m \rightarrow \infty$, il che implica:

$$h(y) < h + \epsilon.$$

Dunque, facendo tendere ϵ a zero, si ottiene il risultato desiderato:

$$H(x) \leq h.$$

\square

1.4 Arnold's cat dataset: comportamento caotico e non caotico

Per comprendere al meglio le proprietà della mappa del gatto di Arnold, è utile distinguere tra comportamento caotico e non caotico o regolare del sistema. Il primo si manifesta con divergenza esponenziale tra traiettorie vicine, causata dalla sensibilità alle condizioni iniziali, mentre l'altro presenta traiettorie stabili o periodiche, prive di sensibilità significativa alle condizioni iniziali.

Consideriamo due configurazioni della matrice di trasformazione:

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \quad R = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix},$$

La matrice C genera un comportamento caotico, mentre R induce dinamiche non caotiche, regolari o stabili.

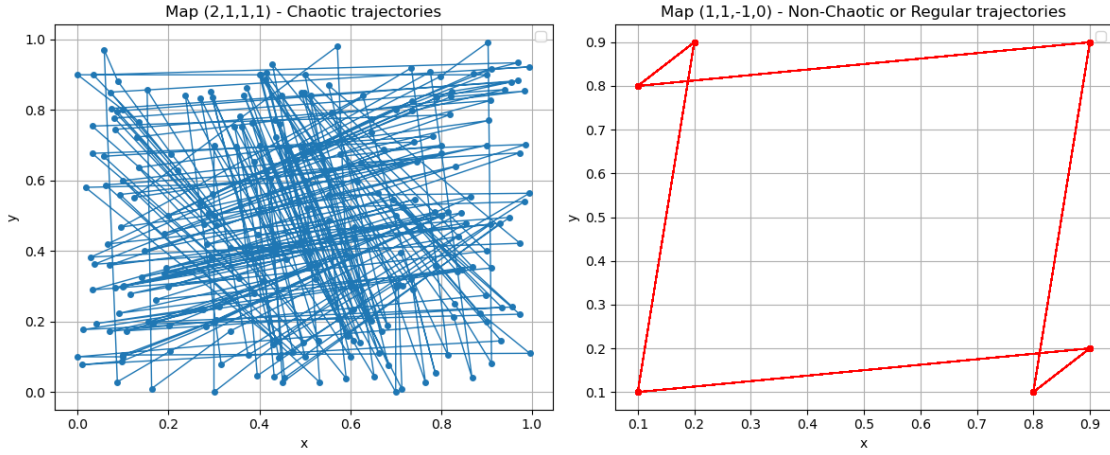


Figura 1.1: Confronto tra traiettorie caotiche e regolari o non caotiche.

Il coefficiente di Lyapunov [1] misura il tasso medio di divergenza esponenziale di traiettorie inizialmente vicine nello spazio delle fasi di un sistema dinamico. È definito come:

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \left(\frac{\|\delta x(t)\|}{\|\delta x(0)\|} \right),$$

dove:

- $\delta x(0)$ rappresenta una piccola perturbazione iniziale nello spazio delle fasi;
- $\delta x(t)$ è la separazione tra le due traiettorie dopo un tempo t ;
- $\|\cdot\|$ denota una norma (tipicamente euclidea).

Il valore di λ determina il comportamento del sistema: se $\lambda > 0$, il sistema è caotico; se $\lambda \leq 0$, il sistema è non caotico (regolare o stabile), con traiettorie che non divergono o convergono.

In particolare: Sia T una matrice 2×2 reale con autovalori λ_1, λ_2 (che supponiamo reali e distinti) e autovettori associati v_1, v_2 . Supponiamo che $\lambda_1 > 1$ e $0 < \lambda_2 < 1$.

Dati un vettore iniziale x_0 , possiamo decomporlo nella base degli autovettori:

$$x_0 = \alpha v_1 + \beta v_2.$$

Iterando l'operatore T , otteniamo:

$$T^n x_0 = \alpha \lambda_1^n v_1 + \beta \lambda_2^n v_2.$$

Poiché λ_1^n cresce esponenzialmente mentre $\lambda_2^n \rightarrow 0$, la traiettoria è dominata da $\lambda_1^n v_1$.

Definiamo la funzione di crescita esponenziale della norma:

$$\Lambda(x_0, n) = \frac{1}{n} \log \|T^n x_0\|.$$

Usando la decomposizione precedente:

$$\|T^n x_0\| = \|\alpha \lambda_1^n v_1 + \beta \lambda_2^n v_2\|.$$

Per $n \rightarrow \infty$, il termine dominante è $\alpha \lambda_1^n v_1$, quindi:

$$\Lambda(x_0, n) \rightarrow \log |\lambda_1|.$$

Il limite superiore definisce il coefficiente di Lyapunov massimo:

$$\lambda_{\max} = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|T^n x_0\| = \log |\lambda_1|.$$

Se $\lambda_{\max} > 0$, il sistema è sensibile alle condizioni iniziali, ovvero piccole perturbazioni iniziali si amplificano esponenzialmente.

Teorema 1.4.1 (Teorema di Oseledets [9]). Sia $f : M \rightarrow M$ una trasformazione differenziabile su una varietà differenziabile M e sia μ una misura invariante ergodica su M . Allora, per μ -quasi ogni punto $x \in M$, esiste una decomposizione dello spazio tangente $T_x M$ in sottospazi $E^i(x)$ e numeri reali λ_i tali che:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|Df^n(x)v\| = \lambda_i$$

per ogni $v \in E^i(x) \setminus \{0\}$. I numeri λ_i sono detti esponenti di Lyapunov.

Il Teorema di Oseledets garantisce che per una trasformazione T su uno spazio misurabile con misura μ , esistono esponenti di Lyapunov $\lambda_1 \geq \lambda_2$ tali che:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|T^n x_0\| = \lambda_{\max}.$$

In questo caso, si ha una matrice 2×2 con un esponente di Lyapunov positivo ($\log \lambda_1$), che implica separazione esponenziale delle orbite e quindi comportamento caotico.

Un altro risultato importante riguarda il Teorema di Pesin, stabilisce un legame quantitativo tra l'entropia di Kolmogorov-Sinai e gli esponenti di Lyapunov positivi.

Teorema 1.4.2 (Formula di Pesin [8]). Sia $f : M \rightarrow M$ una trasformazione differenziabile di classe C^1 su una varietà differenziabile compatta M , e sia μ una misura invariante ergodica assolutamente continua rispetto alla misura di Lebesgue. Allora, l'entropia di Kolmogorov-Sinai del sistema è data dalla somma degli esponenti di Lyapunov positivi pesati dalla misura invariante:

$$h_{\text{KS}}(f) = \int_M \sum_{\lambda_i(x) > 0} \lambda_i(x) d\mu(x).$$

Questa relazione [4] fornisce un collegamento diretto tra il tasso di crescita dell'informazione e la divergenza delle traiettorie nel sistema dinamico. Nel caso di sistemi uniformemente iperbolici, dove gli esponenti di Lyapunov sono costanti per quasi tutti i punti della varietà, la formula di Pesin, come si riduce a:

$$h_{\text{KS}}(T) = \sum_{\lambda_i > 0} \lambda_i.$$

Nel caso in analisi, si prende in considerazione la Arnold's Cat Map, che è un caso particolare di sistema iperbolico con due esponenti di Lyapunov: uno positivo (λ_1) e uno negativo (λ_2).

Poiché $\lambda_2 < 0$, l'entropia di Kolmogorov-Sinai è determinata solo dall'esponente positivo:

$$h_{\text{KS}}(T) = \lambda_1.$$

Questa relazione indica che l'entropia rappresenta la velocità con cui il sistema crea nuove informazioni ed è strettamente legata alla rapidità con cui le traiettorie inizialmente vicine si allontanano nel tempo.

1.4.1 Arnold's cat map

La mappa del gatto di Arnold [5] è una trasformazione lineare definita sul toro bidimensionale T^2 , che rappresenta lo spazio quoziente $\mathbb{R}^2/\mathbb{Z}^2$, descritta come:

$$\Gamma : T^2 \rightarrow T^2, \quad \Gamma(x, y) \rightarrow (2x + y, x + y) \mod 1.$$

dove $(x, y) \in T^2$. In notazione matriciale come:

$$\Gamma \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \mod 1.$$

Dove la matrice della trasformazione,

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},$$

descrive una mappa iperbolica con le seguenti proprietà:

- Autovalori: $\lambda_1 = \frac{3+\sqrt{5}}{2} = 2.618$ (espansione) e $\lambda_2 = \frac{3-\sqrt{5}}{2} = 0.382$ (contrazione), inoltre $\lambda_1 \lambda_2 = 1$.
- Determinante: Pari a 1, il che implica che la mappa preserva l'area.

1.4.2 Generazione del dataset

La mappa deforma lo spazio tramite stiramenti, piegamenti e un'operazione modulo, riportando i punti nel toro unitario, rappresentabile come il quadrato $[0, 1) \times [0, 1)$. Questo dominio è suddiviso in quattro sotto-regioni, ciascuna con la seguente classificazione:

- Quadrante 1: $0 \leq x \leq 0.5, 0 \leq y \leq 0.5$;
- Quadrante 2: $0.5 < x \leq 1, 0.5 < y \leq 1$;
- Gli altri punti appartengono ai restanti quadranti.

Si assegna così un valore binario (0 o 1) a ogni punto in base alla regione di appartenenza. Questo approccio consente di tradurre traiettorie continue in una rappresentazione discreta.

Il dataset è stato costruito iterando la mappa per un numero N di passi, partendo da coordinate iniziali (x_0, y_0) con valori casuali perturbati casualmente con un piccolo offset (± 0.01). Ad ogni iterazione, le nuove coordinate (x', y') sono calcolate attraverso:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \mod 1.$$

Questo processo genera una stringa binaria rappresentativa del percorso del sistema nel tempo.

Un esempio di output è una sequenza binaria come: 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1; dove ogni cifra rappresenta il quadrante in cui si trova il punto ad ogni iterazione.

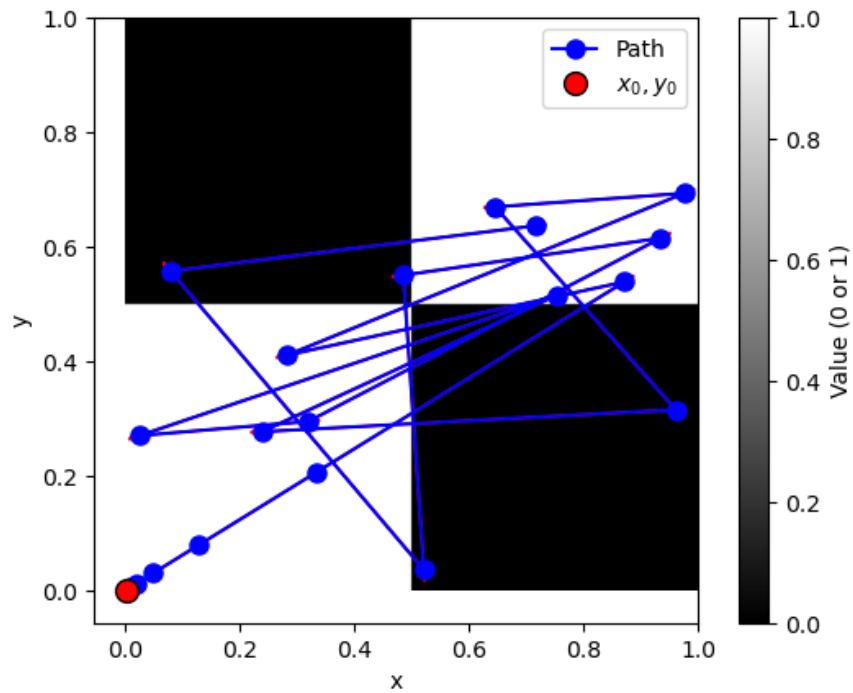


Figura 1.2: Percorso generato dalla Arnold's Cat Map.

Capitolo 2

Metodologia

La valutazione delle sequenze binarie generate dalla mappa di Arnold viene effettuata attraverso il calcolo dell'entropia di Shannon, basato sul conteggio esatto delle frequenze delle sottostringhe di lunghezza variabile secondo un intervallo definito tra la lunghezza minima e massima della sottosequenza; consentendo una stima precisa della distribuzione delle sottostrutture presenti nella sequenza.

Il metodo impiegato prevede l'estrazione di tutte le sottosequenze di lunghezza d e la determinazione delle rispettive frequenze di occorrenza. A partire da questa distribuzione, viene calcolata l'entropia $H(d)$ per quantificarne la complessità. L'intero procedimento viene applicato su un intervallo di valori di d , che definisce la lunghezza della sottostringa, permettendo lo studio della dinamica del sistema.

L'andamento di $H(d)$ in funzione di d è utilizzato per stimare il coefficiente di Lyapunov. In particolare, l'entropia viene calcolata per tutti i valori di d e i risultati ottenuti vengono interpolati tramite regressione lineare per determinare il coefficiente angolare α , della relazione $H(d)$. Questo è direttamente correlato all'esponente di Lyapunov e fornisce una misura quantitativa del comportamento del sistema. Inoltre, si ricerca la presenza di un plateau entropico, che indicherebbe una stazionarietà dell'entropia e potrebbe suggerire la comprimibilità della sequenza esaminata, anche nei sistemi caotici deterministici.

Di seguito, si descrivono nel dettaglio le fasi operative adottate, dalla costruzione della distribuzione delle frequenze al calcolo dell'entropia, fino alla stima del coefficiente di Lyapunov.

2.1 Metodo basato sul conteggio delle frequenze

Questo metodo esamina la struttura della sequenza binaria estraendo tutte le possibili sottostringhe di lunghezza d e contando le loro occorrenze. Una volta ottenuta la distribuzione completa delle frequenze, viene calcolata l'entropia di Shannon per ciascun valore di d , quantificando la diversità dei pattern presenti nella sequenza.

Successivamente, l'andamento dell'entropia in funzione di d viene modellato attraverso una trasformazione logaritmica, permettendo di applicare una regressione lineare della forma $H(d) \approx \alpha d + \beta$. Il coefficiente angolare α della retta di regressione viene infine utilizzato come parametro per identificare la complessità della sequenza.

Algorithm 1 Frequency-Based Method

Require: `binary_list`, `min_dim`, `max_dim`

Ensure: `dims`, `entropies`, `alpha`

```
1: Inizializzazione
2: Definisci una lista vuota dims
3: Definisci una lista vuota entropies
4: for  $d$  da min_dim a max_dim do
5:   Conteggio delle sottostringhe
6:   Inizializza un Counter vuoto
7:   total_windows  $\leftarrow$  lunghezza(binary_list) - d + 1
8:   for  $i$  da 0 a total_windows - 1 do
9:     pattern  $\leftarrow$  binary_list[i : i + d]
10:    Aggiorna counter[pattern]  $\leftarrow$  counter[pattern] + 1
11:   end for
12:   Calcolo dell'entropia di Shannon
13:    $H \leftarrow 0$ 
14:   for pattern in counter do
15:      $p \leftarrow \frac{\text{counter}[\text{pattern}]}{\text{total\_windows}}$ 
16:      $H \leftarrow H - p \log_2 p$ 
17:   end for
18:   Aggiungi  $d$  a dims
19:   Aggiungi  $H$  a entropies
20: end for
21: Regressione lineare su  $H(d)$ 
22: Definisci valid  $\leftarrow$  valori  $(d, H)$  con  $H > 0$ 
23: Applica un fit lineare  $H(d) \approx \alpha d + \beta$ 
24: dims, entropies,  $\alpha$ 
```

2.2 Costruzione della distribuzione delle sottostringhe

La distribuzione delle sottostringhe viene ottenuta estraendo e conteggiando tutte le possibili sequenze di lunghezza d presenti nella sequenza binaria S , generata dalla mappa del gatto di Arnold. Sia $S = (s_1, s_2, \dots, s_N)$ una sequenza binaria di lunghezza N , con $s_i \in \{0, 1\}$. Si definisce una sottostringa di lunghezza d come una sequenza ordinata di d bit consecutivi estratti da S , ovvero:

$$s_i = (s_i, s_{i+1}, \dots, s_{i+d-1})$$

per $i = 1, 2, \dots, N - d + 1$. Il numero totale di sottostringhe di lunghezza d che possono essere estratte dalla sequenza S è dato da:

$$\text{total_windows} = N - d + 1$$

La lunghezza d non è fissata, bensì varia nell'intervallo $[\text{min_dim}, \text{max_dim}]$, dove:

- min_dim è la lunghezza minima delle sottostringhe analizzate;
- max_dim è la lunghezza massima considerata.

Viene analizzato ogni valore di d in questo intervallo, estraendo le relative sottostringhe.

Per ciascuna sottostringa s , viene calcolata la frequenza assoluta, ovvero il numero di volte in cui essa compare nella sequenza. Definiamo quindi la distribuzione di probabilità di ciascuna sottostringa come:

$$p(s) = \frac{\text{counter}[s]}{\text{total_windows}}$$

dove:

- $\text{counter}[s]$ rappresenta il numero di occorrenze della sottostringa s ;
- total_windows è il numero totale di sottostringhe estratte dalla sequenza binaria.

L'implementazione dell'algoritmo segue una scansione lineare della sequenza binaria, in cui ogni finestra di d elementi viene trasformata in una tupla immutabile e registrata in una struttura dati di tipo Counter. Infine si utilizza questa distribuzione per il calcolo dell'entropia.

2.3 Calcolo diretto dell'entropia

L'entropia è calcolata secondo la definizione di Shannon precedentemente descritta, applicata al caso specifico:

$$H(d) = - \sum_{s \in \mathcal{S}_d} p(s) \log_2 p(s),$$

dove \mathcal{S}_d è l'insieme di tutte le sottostringhe di lunghezza d .

L'algoritmo scorre tutte le sottostringhe della sequenza, accumulando il contributo di ogni pattern alla somma dell'entropia. Il calcolo viene ripetuto per ogni dimensione di d nell'intervallo definito.

2.4 Stima del coefficiente di Lyapunov tramite la crescita dell'entropia

L'idea alla base è che, per un sistema dinamico ergodico con una distribuzione stabile delle sottosequenze, l'entropia di Shannon calcolata sulle sottostringhe di lunghezza d cresce in modo lineare rispetto a d per grandi valori di d . Questo comportamento è analogo a quello osservato nella definizione degli esponenti di Lyapunov, dove si analizza la crescita logaritmica delle separazioni tra traiettorie.

Se la crescita dell'entropia è asintoticamente lineare, possiamo scrivere:

$$H(d) \approx \alpha d + \beta, \quad \text{per } d \gg 1.$$

Dividendo entrambi i lati per d e passando al limite:

$$\lim_{d \rightarrow \infty} \frac{H(d)}{d} = \lim_{d \rightarrow \infty} \left(\alpha + \frac{\beta}{d} \right).$$

Poiché $\frac{\beta}{d} \rightarrow 0$ per $d \rightarrow \infty$, otteniamo:

$$\lim_{d \rightarrow \infty} \frac{H(d)}{d} = \alpha.$$

L'entropia $H(d)$ delle sottostringhe viene analizzata in funzione della loro lunghezza d , mostrando un comportamento che può essere modellato dalla relazione lineare definita precedentemente.

Per determinare i coefficienti α e β , si applica il metodo dei minimi quadrati, che minimizza la somma degli errori quadratici tra i valori osservati di $H(d)$ e quelli stimati dal modello lineare. Si definisce il problema di ottimizzazione:

$$\min_{\alpha, \beta} \sum_i (H(d_i) - (\alpha d_i + \beta))^2.$$

Questo metodo fornisce la miglior stima lineare possibile di $H(d)$ in senso dei minimi quadrati, risolvendo il sistema lineare associato. Il modello è espresso in forma matriciale come:

$$\mathbf{X} = \begin{bmatrix} d_1 & 1 \\ d_2 & 1 \\ \vdots & \vdots \\ d_n & 1 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} H(d_1) \\ H(d_2) \\ \vdots \\ H(d_n) \end{bmatrix}.$$

I coefficienti $\boldsymbol{\theta} = (\alpha, \beta)$ sono quindi ottenuti risolvendo:

$$\boldsymbol{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{H}.$$

L'implementazione pratica di questo calcolo avviene con la funzione `np.polyfit`, che restituisce direttamente α e β :

$$\alpha, \beta = \text{np.polyfit}(\text{dims}, \text{entropies}, 1).$$

Nei sistemi caotici, l'entropia segue un andamento lineare rispetto a d , descritto dalla relazione:

$$H(d) \approx \lambda d,$$

dove λ rappresenta il coefficiente di Lyapunov, che quantifica la separazione esponenziale tra traiettorie inizialmente vicine. Confrontando le due espressioni, si ottiene:

$$\lambda \approx \alpha.$$

Questa relazione permette di stimare λ direttamente dall'analisi della crescita dell'entropia. Il valore di α fornisce quindi un'indicazione della dinamica del sistema: se positivo e significativo, indica un comportamento caotico, mentre valori bassi o nulli suggeriscono una dinamica regolare o non caotica.

Si prendono da esempio le mappe $C = (2, 1, 1, 1)$ e $R = (1, 1, -1, 0)$ con parametri: $N = 100000$, $\text{min_dim} = 2$, $\text{max_dim} = 100$. E' possibile visualizzare, si veda Figura 2.1, il rapporto tra $H(d)$ e la lunghezza d delle sottostringhe, permettendo di distinguere tra sistemi caotici e regolari. Nei sistemi regolari, di seguito a destra, $H(d)$ presenta un comportamento oscillatorio con valori ripetuti, mentre nei sistemi caotici, nel grafico a sinistra, cresce rapidamente per poi raggiungere un plateau. In particolare per la configurazione regolare R si nota un coefficiente angolare della retta di regressione $\alpha \approx 0$. Considerando $\alpha \approx \lambda$, il valore $\lambda \leq 0$ conferma la regolarità del sistema. Nello specifico per ogni dimensione di d si trova lo stesso risultato pari a $H(d) = 1.5850$, dato da $H(d) \approx 0.0000 * d + 1.5850$. Mentre per la configurazine caotica C , $\alpha = 0.064$, mostrando un $\lambda > 0$ e dunque un comportamento caotico che si stabilizza in un plateau per $d > 26$ pari a $H(d) = 16.6082$, dato da $H(d) \approx 0.0639 * d + 12.1711$.

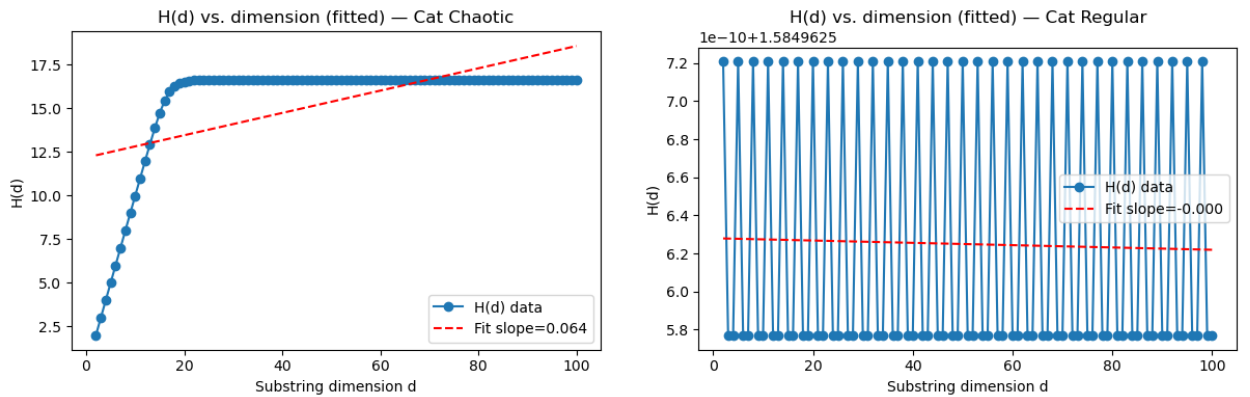


Figura 2.1: Andamento del rapporto $\frac{H(d)}{d}$ per d da 2 a 100 nei due tipi di sistemi: caotico (a sinistra) e regolare o non caotico (a destra).

Nei sistemi regolari, la distribuzione delle sottostringhe è dominata da una ridondanza strutturale che impedisce la crescita indefinita di $H(d)$. La periodicità delle traiettorie nel sistema si traduce in una ripetizione delle distribuzioni di probabilità delle sottostringhe. Poiché la diversità delle sottostringhe non aumenta con d , la pendenza della regressione lineare risulta trascurabile:

$$\lim_{d \rightarrow \infty} \frac{H(d)}{d} \approx 0.$$

L'assenza di crescita entropica riflette la natura deterministica e comprimibile del sistema, confermando l'assenza di separazione esponenziale delle traiettorie.

Nei sistemi caotici invece, $H(d)$ cresce rapidamente per bassi valori di d , fino a raggiungere un plateau. Questa transizione indica che, oltre una certa lunghezza, tutte le sottostringhe possibili si distribuiscono con una probabilità stazionaria, stabilizzando la crescita dell'entropia:

$$\lim_{d \rightarrow \infty} H(d) \approx H_{\max}.$$

La presenza del plateau è una diretta conseguenza dell'ergodicità del sistema e della rapida dispersione dell'informazione nelle traiettorie.

Il coefficiente angolare α della regressione lineare è proporzionale all'esponente di Lyapunov λ_{\max} ; se maggiore di 0, indica un sistema caotico, confermando quanto descritto in precedenza.

Capitolo 3

Risultati e discussione

Si vuole mostrare che quanto riportato precedentemente è valido per altre mappe, considerando sempre il gatto di Arnold. Per garantire la corretta generazione delle mappe del gatto di Arnold come caotiche o regolari, sono stati adottati specifici criteri. Di seguito vengono elencati i vincoli [3] per la generazione di entrambe le tipologie di mappe.

Vincoli per le mappe caotiche:

- Determinante: deve essere $\det(\Gamma) = 1$ per garantire che la trasformazione sia invertibile e che preservi l'area.
- Autovalori irrazionali e distinti: gli autovalori della matrice devono essere irrazionali e distinti: $\lambda_{1,2} = \frac{\text{Tr}(\Gamma) \pm \sqrt{\text{Tr}(\Gamma)^2 - 4}}{2}$, con il discriminante $\text{Tr}(\Gamma)^2 - 4 > 0$, dove $\text{Tr}(\Gamma)$ è la traccia della matrice, ovvero la somma degli elementi sulla diagonale principale.
- Iperbolicità del sistema: deve esserci una chiara separazione tra direzione stabile e instabile, ovvero: $\lambda_1 > 1$, $\lambda_2 < 1$. Ciò implica la presenza di un esponente di Lyapunov positivo, segno distintivo del caos deterministico.
- Prodotto degli autovalori: per assicurare la conservazione dell'area, gli autovalori devono rispettare la condizione: $\lambda_1 \lambda_2 = 1$. Questo viene garantito già dal vincolo sul determinante.

Vincoli per le mappe regolari o non caotiche:

- Determinante: deve essere $\det(\Gamma) = \pm 1$ per garantire che la trasformazione sia invertibile e conservativa.
- Autovalori complessi: gli autovalori della matrice di trasformazione devono essere complessi coniugati della forma: $\lambda_{1,2} = \frac{\text{Tr}(\Gamma) \pm i\sqrt{|\Delta|}}{2}$, con $\Delta = \text{Tr}(\Gamma)^2 - 4\det(\Gamma) < 0$.
- Modulo degli autovalori: gli autovalori devono avere modulo unitario, cioè: $|\lambda_1| = |\lambda_2| = 1$. Questo assicura che il sistema mostri solo un moto oscillatorio senza crescita esponenziale delle traiettorie, evitando quindi il caos.
- Evitare valori degeneri: Per evitare comportamenti quasi caotici o degenerazioni periodiche con lungo periodo, gli autovalori non devono essere vicini a $\lambda = 1$ o $\lambda = -1$.

Seguendo queste proprietà, sono state generate 10 mappe caotiche e altrettante regolari, rappresentate da quadruple di interi (a, b, c, d) selezionate in un intervallo prestabilito, evitando la matrice nulla per evitare dinamiche banali. Per ciascuna di queste, si è quindi prodotta una sequenza binaria utilizzando l'Arnold's Cat Map per i seguenti parametri: $N = 100000$, $min_dim = 2$, $max_dim = 100$. Su tale sequenza è stata applicata la metodologia Frequency-Based, stimando così i parametri α e β mediante regressione lineare $H(d) \approx \alpha d + \beta$.

Di seguito vengono esplicitati gli pseudocodici che riportano le proprietà generali per generare mappe caotiche o non caotiche. Si tiene conto di un intervallo di valori compreso tra $[-5, 5]$ per la quadrupla (a, b, c, d) in entrambe le tipologie, per semplicità di calcolo. Inoltre, per le mappe regolari, vengono specificate delle tolleranze nei vincoli sugli autovalori. Si impone che il loro modulo sia 1 con una tolleranza di $atol = 1e - 20$ e che non siano troppo vicini a $\lambda = 1$ o $\lambda = -1$; con una distanza minima di ± 0.05 in modo da evitare dinamiche quasi caotiche o oscillazioni anomale.

Algorithm 2 Generation of chaotic maps

Require: n : numero di matrici caotiche da generare

Ensure: Lista di matrici caotiche (a, b, c, d)

```

1: Inizializza una lista vuota chaotic_maps
2: while  $\text{len}(\text{chaotic\_maps}) < n$  do
3:   Estrai casualmente quattro interi  $(a, b, c, d)$  tra -5 e 5
4:   Calcola il determinante  $\det_{\Gamma} = a \cdot d - b \cdot c$ 
5:   if  $\det_{\Gamma} \neq 1$  then
6:     continue
7:   end if
8:   Calcola la traccia  $\text{trace} = a + d$ 
9:   Calcola il discriminante  $\text{discriminant} = \text{trace}^2 - 4$ 
10:  if discriminant è un quadrato perfetto then
11:    continue
12:  end if
13:  Calcola gli autovalori della matrice  $\Gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 
14:  if max autovalori  $> 1$  e min autovalori  $< 1$  then
15:    Aggiungi  $(a, b, c, d)$  a chaotic_maps
16:  end if
17: end while
18: return chaotic_maps
```

Algorithm 3 Generation of regular maps

Require: n : numero di matrici regolari da generare

Ensure: Lista di matrici regolari (a, b, c, d)

```

1: Inizializza una lista vuota regular_maps
2: while len(regular_maps) < n do
3:   Estrai casualmente quattro interi  $(a, b, c, d)$  tra -5 e 5
4:   Calcola il determinante  $det_{\Gamma} = a \cdot d - b \cdot c$ 
5:   if  $det_{\Gamma} \notin \{-1, 1\}$  then
6:     continue
7:   end if
8:   Calcola la traccia  $trace = a + d$ 
9:   Calcola il discriminante  $discriminant = trace^2 - 4 \cdot det_{\Gamma}$ 
10:  if  $discriminant \geq 0$  then
11:    continue
12:  end if
13:  Calcola gli autovalori della matrice  $\Gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 
14:  if modulo degli autovalori  $\approx 1$  con tolleranza  $10^{-6}$  then
15:    if nessun autovalore vicino a 1 o -1 (distanza > 0.05) then
16:      Aggiungi  $(a, b, c, d)$  a regular_maps
17:    end if
18:  end if
19: end while
20: return regular_maps

```

Dai risultati grafici ottenuti, si osserva che nei sistemi caotici, si veda Figura 3.1 l'entropia $H(d)$ cresce rapidamente per piccoli valori di d , per poi stabilizzarsi su un plateau. Questo comportamento mostra che, oltre una certa soglia di d non viene acquisita ulteriore informazione. Utilizzando $N = 100000$ si garantisce una stima stabile dell'entropia. Come già descritto in precedenza, una sequenza troppo breve, la frequenza delle sottostringhe potrebbe subire variazioni non rappresentative del sistema. I valori $min_dim = 2$ e $max_dim = 100$ sono stati scelti per evidenziare sia la fase di crescita iniziale che la successiva saturazione. Se max_dim fosse stato inferiore, il plateau non sarebbe stato rilevabile, mentre valori più alti sarebbero stati ridondanti, poiché l'entropia risultava già stabilizzata.

Per queste configurazioni, il plateau emerge a partire da $d > 20$, suggerendo che oltre tale soglia $H(d)$ non subisce variazioni significative. Inoltre, il coefficiente α , ottenuto dall'adattamento lineare della crescita iniziale di $H(d)$, fornisce una stima del massimo esponente di Lyapunov, confermando la natura caotica del sistema poiché $\alpha > 0$.

In quelli regolari, in Figura 3.2, invece, $H(d)$ non mostra una crescita sostenuta, ma segue un andamento caratterizzato da oscillazioni periodiche o si mantiene costante, indicando un sistema completamente deterministico e prevedibile. In questi casi, le traiettorie nello spazio delle fasi sono tipicamente periodiche o quasi-periodiche, il che significa che la dinamica segue schemi ripetitivi senza un'esplorazione caotica delle configurazioni possibili. Di conseguenza, il numero di sottostringhe distinte rimane limitato e l'entropia non cresce significativamente. Questo è coerente con il fatto che nei sistemi non caotici il massimo esponente di Lyapunov è minore o uguale a zero, confermando l'assenza di divergenza esponenziale tra traiettorie inizialmente vicine.

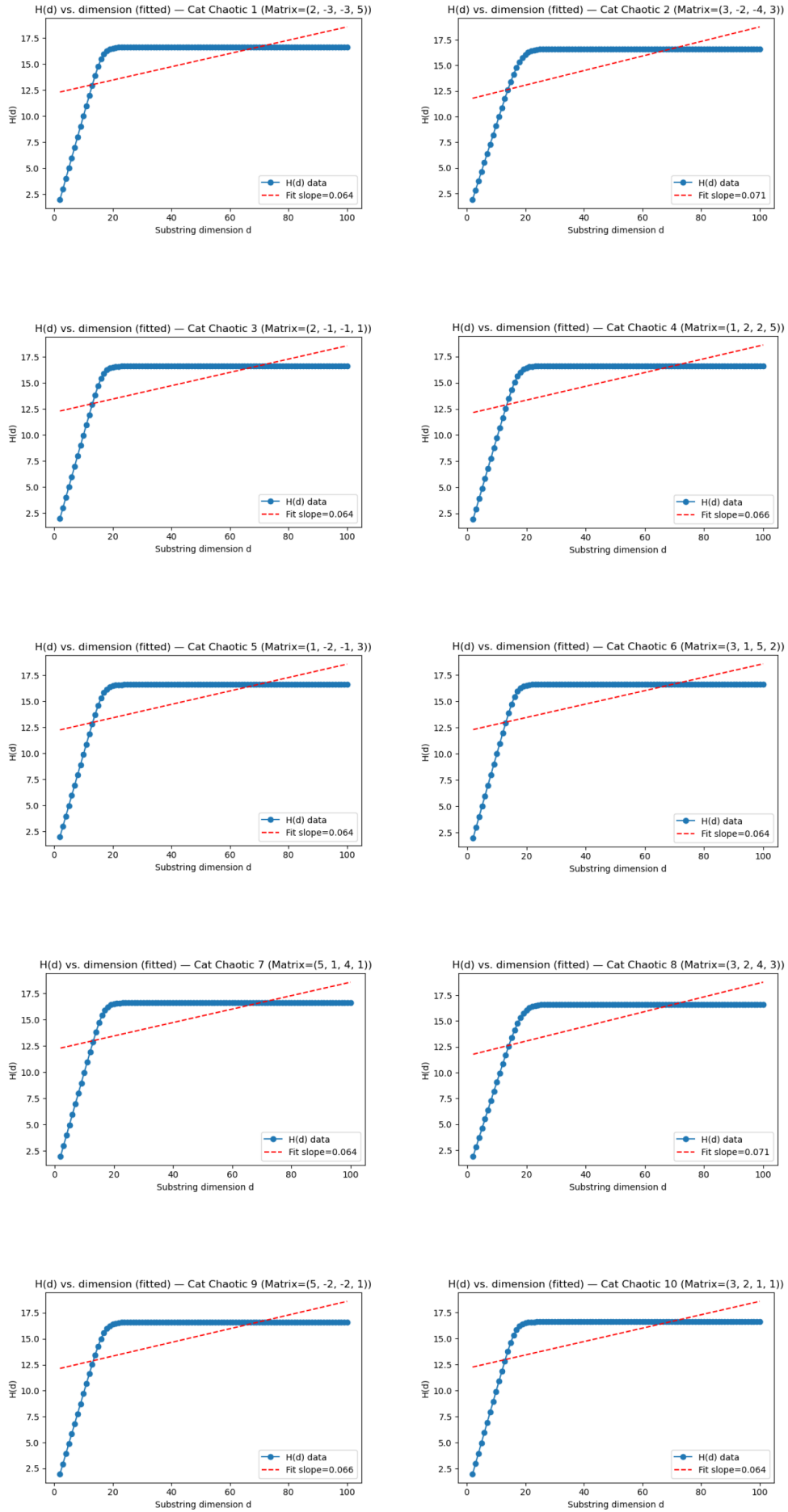


Figura 3.1: Andamento di $H(d)$ e del fit lineare per 10 diverse matrici caotiche.

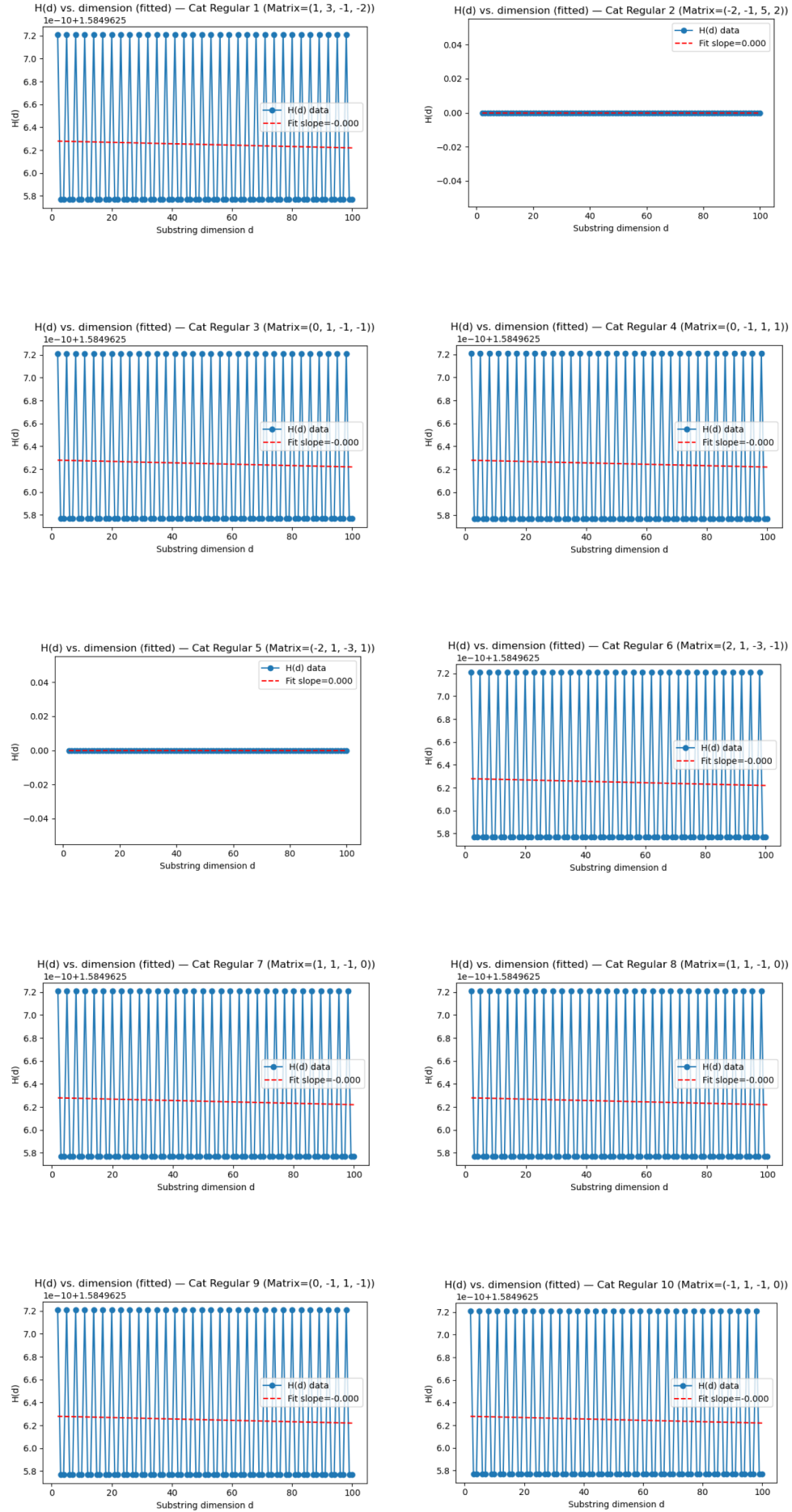


Figura 3.2: Andamento di $H(d)$ e del fit lineare per 10 diverse matrici regolari o non caotiche.

3.1 Scalabilità e impatto della dimensione delle sottostringhe sulla stima dell'entropia e del coefficiente di Lyapunov

Si analizza la crescita dell'entropia in un sistema caotico descritto dalla Arnold's Cat Map, utilizzando la stessa matrice di trasformazione $C = (2, 1, 1, 1)$, già considerata in precedenza. L'obiettivo è osservare eventuali differenze di comportamento tra sequenze molto corte e molto lunghe, verificando come scala l'entropia e a quale valore si stabilizza il coefficiente di Lyapunov. Per individuare il plateau dell'entropia, è necessario considerare sequenze sufficientemente lunghe e sottostringhe di dimensione adeguata, in modo da ottenere stime rappresentative della dinamica caotica.

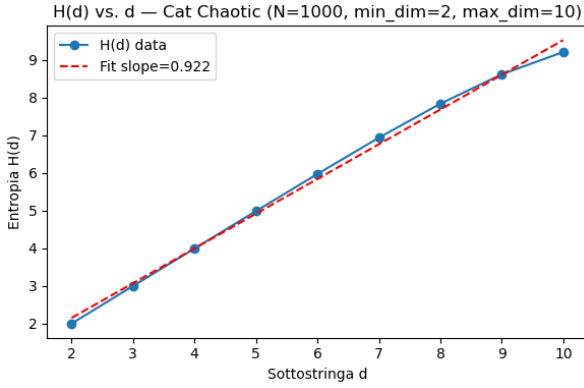


Figura 3.3: $H(d)$ per $d = [2; 10], N = 1000$

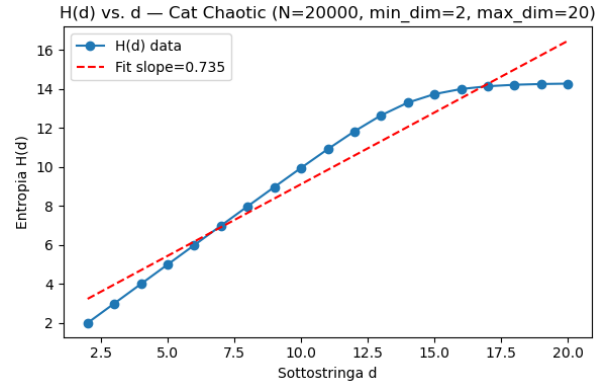


Figura 3.4: $H(d)$ per $d = [2; 20], N = 20000$

Per una sequenza breve come in Figura 3.3, con $N = 1000$, si osserva una crescita quasi lineare con coefficiente di Lyapunov $\lambda \approx 0.922$, ma il plateau non è visibile a causa della limitata lunghezza della sequenza. Con $N = 20000$ in Figura 3.4, la crescita iniziale segue una pendenza inferiore con $\lambda \approx 0.735$, e per $d > 15$ si nota un rallentamento, suggerendo un avvicinamento al plateau, che tuttavia non è ancora completamente evidente.

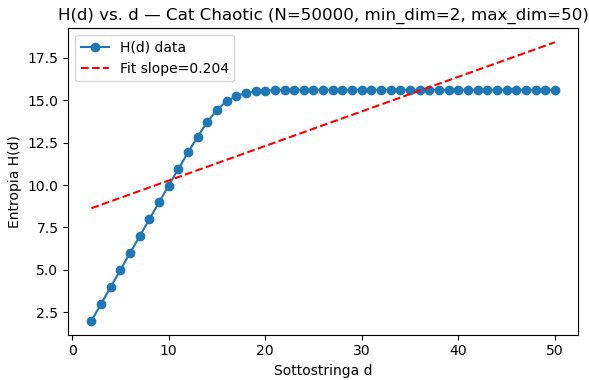


Figura 3.5: $H(d)$ per $d = [2; 50], N = 50000$

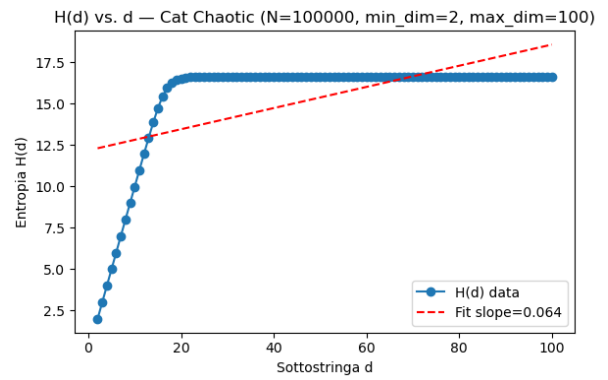


Figura 3.6: $H(d)$ per $d = [2; 100], N = 100000$

Nel caso di $N = 50000$, Figura 3.5, l'entropia cresce rapidamente fino a circa $d = 15$, dove si osserva un evidente plateau. Il fit lineare mostra un coefficiente di crescita più basso rispetto ai casi precedenti con $\lambda \approx 0.204$.

Per $N = 100000$, Figura 3.6, la tendenza è simile, con il plateau che si manifesta già intorno a $d \approx 20$ e si stabilizza su un valore pari a $H(d) = 16.6082$. L'inclinazione della retta di best-fit è ancora inferiore $\lambda \approx 0.064$, suggerendo che per lunghezze di sequenza maggiori l'entropia tende a saturarsi più rapidamente.

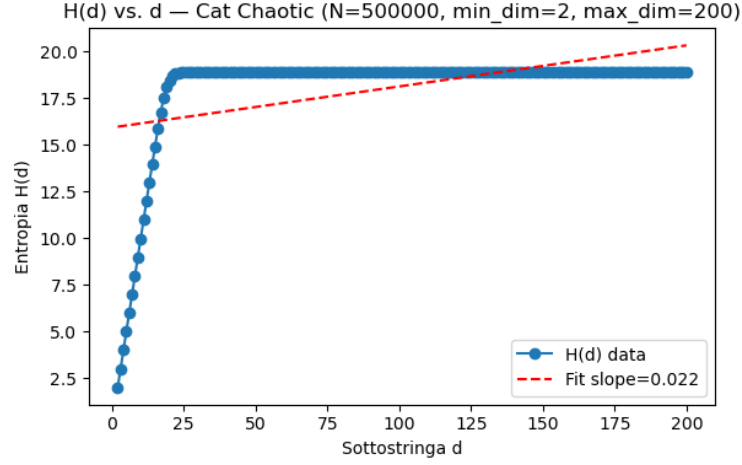


Figura 3.7: $H(d)$ per $d = [2; 200]$ e $N = 500000$

Il comportamento mostrato in figura 3.7 è sostanzialmente analogo a quello in Figura 3.6. Inoltre, aumentando il numero di punti a $N = 500000$ e considerando sottostringhe fino a $d = 200$, il calcolo diventa significativamente più oneroso dal punto di vista computazionale. Poiché il plateau dell'entropia è già chiaramente visibile con i parametri: $N = 100000, min_dim = 2, max_dim = 100$; quest'ultima configurazione risulta essere la scelta più efficace, motivando i parametri scelti per mostrare il fenomeno senza eccessivi costi computazionali.

3.2 Media dell'entropia su campioni estesi di mappe caotiche e regolari

Successivamente si generano altre 50 mappe del gatto di Arnold caotiche e altrettante regolari, con i medesimi parametri $N = 100000$, $min_dim = 2$, $max_dim = 100$. Viene fatta una media delle $H(d)$ per ogni categoria, in modo da ottenere un andamento rappresentativo e stabile dell'entropia per i sistemi caotici e regolari, evitando possibili fluttuazioni dovute a singole istanze specifiche.

Come si nota in Figura 3.8, la curva in rosso rappresenta l'andamento medio dell'entropia $H(d)$ all'aumentare di d per i sistemi caotici, mentre quella in blu riguarda i sistemi regolari. Dai dati riportati, si osserva che i sistemi caotici mostrano una crescita molto rapida di $H(d)$ (ad esempio 16.6090 a $d = 30$, si veda Tabella 3.1), che raggiunge un valore massimo, un plateau. Al contrario, i sistemi regolari restano pressoché invariati su un livello più basso, pari a 1.2257. Questa larga differenza mostra come l'andamento medio dell'entropia, anche considerando 50 mappe diverse per ciascun sistema, sia rappresentativo del comportamento tipico di entrambe le categorie. Nei sistemi caotici, la complessità e la varietà dei pattern emergono chiaramente, portando a un'ampia gamma di sottostringhe e a un plateau dell'entropia per valori elevati di d . Nei sistemi regolari, invece, la struttura più rigida e prevedibile si riflette in un'entropia che rimane costantemente bassa, confermando la limitata diversità delle sottostringhe generate.

Per confermare i risultati del grafico, nella Tabella 3.1 sono riportati i valori medi dell'entropia per entrambi i sistemi in funzione della dimensione d .

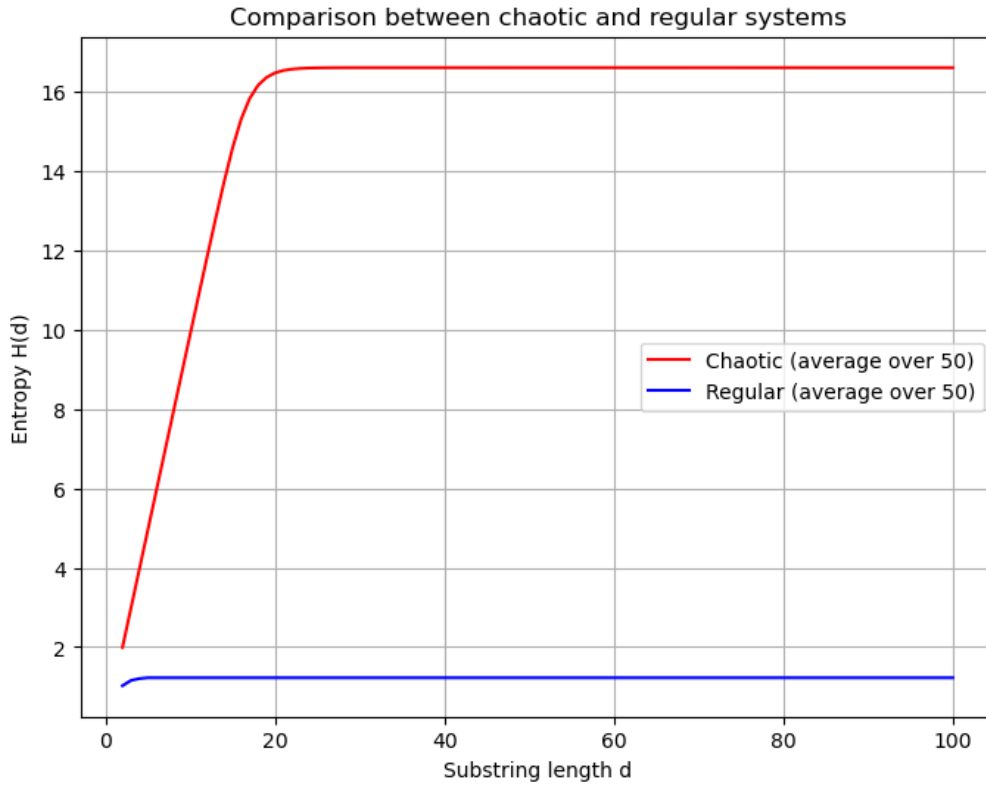


Figura 3.8: Andamento medio dell'entropia $H(d)$ all'aumentare di d per sistemi caotici e regolari, calcolato su 50 mappe per ciascun tipo.

Tabella 3.1: Entropia media per sistemi caotici e regolari al variare della dimensione d .

Dimensione d	Entropia Media (Caotiche)	Entropia Media (Regolari)
2	1.9933	1.0257
3	2.9833	1.1581
4	3.9727	1.2057
5	4.9617	1.2257
6	5.9503	1.2257
7	6.9385	1.2257
8	7.9260	1.2257
9	8.9125	1.2257
10	9.8971	1.2257
11	10.8781	1.2257
12	11.8516	1.2257
13	12.8099	1.2257
14	13.7365	1.2257
15	14.5919	1.2257
16	15.3067	1.2257
17	15.8284	1.2257
18	16.1645	1.2257
19	16.3636	1.2257
20	16.4760	1.2257
21	16.5379	1.2257
22	16.5713	1.2257
23	16.5892	1.2257
24	16.5987	1.2257
25	16.6037	1.2257
26	16.6064	1.2257
27	16.6077	1.2257
28	16.6084	1.2257
29	16.6088	1.2257
30	16.6090	1.2257
31	16.6091	1.2257
32	16.6091	1.2257
33	16.6092	1.2257
34	16.6092	1.2257
35	16.6091	1.2257
36	16.6091	1.2257
37	16.6091	1.2257
38	16.6091	1.2257
39	16.6091	1.2257
40	16.6091	1.2257
41	16.6091	1.2257
42	16.6090	1.2257
43	16.6090	1.2257
44	16.6090	1.2257
45	16.6090	1.2257
46	16.6089	1.2257
47	16.6089	1.2257
48	16.6089	1.2257
49	16.6089	1.2257
50	16.6089	1.2257
⋮	⋮	⋮
91	16.6083	1.2257
92	16.6083	1.2257
93	16.6083	1.2257
94	16.6083	1.2257
95	16.6083	1.2257
96	16.6083	1.2257
97	16.6082	1.2257
98	16.6082	1.2257
99	16.6082	1.2257
100	16.6082	1.2257

Capitolo 4

Conclusioni e studi futuri

In questo lavoro è stato sviluppato e implementato il metodo Frequency-Based, basato sulla Ziv Entropy, per stimare l'entropia di stringhe binarie generate da sistemi dinamici, con una particolare attenzione alla Arnold's Cat Map. L'approccio adottato ha previsto diverse fasi: inizialmente, per ogni sequenza binaria analizzata, sono state estratte tutte le sottostringhe di lunghezza variabile e costruita la rispettiva distribuzione di probabilità. Dopodichè, è stata calcolata l'entropia di Shannon per ciascun valore di d , fornendo una misura quantitativa della diversità e della complessità delle sottostrutture presenti. Infine, l'andamento di $H(d)$ è stato interpolato tramite regressione lineare per ottenere il coefficiente di Lyapunov, che distingue tra sistemi caotici e regolari.

I risultati hanno confermato che, nei sistemi caotici, l'entropia cresce rapidamente per i primi valori di d e poi si assesta in un plateau, evidenziando la dispersione dell'informazione e la natura ergodica del sistema. Nei sistemi regolari, invece, $H(d)$ non mostra una crescita sostenuta, ma segue un andamento oscillatorio o rimane costante, indicando un comportamento prevedibile. L'applicazione del metodo su un campione più esteso di 50 mappe caotiche e 50 regolari ha permesso di confermare la stabilità dei risultati, con curve medie dell'entropia che mostrano differenze marcate tra i due tipi di sistemi. Inoltre, lo studio della scalabilità ha evidenziato che, per sequenze sufficientemente lunghe ($N \geq 100000$) e un intervallo adeguato di d (da 2 a 100), l'entropia raggiunge il plateau senza richiedere un'elevata complessità computazionale, rendendo il metodo adatto all'analisi di dataset estesi.

Poiché l'entropia calcolata riflette direttamente la ridondanza strutturale delle sequenze, il metodo potrebbe risultare utile per lo studio della comprimibilità dei dati. Infatti nei sistemi caotici, sebbene l'alta variabilità delle sottostringhe possa inizialmente far pensare a una minore comprimibilità, il raggiungimento di un plateau nell'entropia indica che anche queste sequenze potrebbero essere soggette a compressione senza perdita. La relazione tra Ziv Entropy e complessità compressiva potrebbe consentire di valutare in modo quantitativo i limiti teorici della compressione per differenti classi di sequenze, con possibili applicazioni nell'ottimizzazione degli algoritmi di compressione dati e nella caratterizzazione della struttura informativa di sistemi dinamici complessi.

In particolar modo si è mostrato che, per la Arnold's Cat Map, la crescita dell'entropia al variare della dimensione d delle sottostringhe raggiunge un valore di saturazione, il cosiddetto plateau, che implica l'esistenza di una costante α tale che $\lim_{d \rightarrow \infty} \frac{H(d)}{d} = \alpha$, collegata all'esponente di Lyapunov del sistema. Quindi, l'entropia $H(d)$ risulta proporzionale a d quando d è abbastanza grande.

Esistono però mappe dinamiche che non presentano questa linearità. Un esempio significativo è la mappa di Pomeau–Manneville [10], nota anche come mappa di intermittenza, descritta da:

$$x_{n+1} = x_n + x_n^z \pmod{1},$$

dove $x_n \in [0, 1)$ e $z > 1$. In tale evoluzione, il sistema alterna fasi in cui la dinamica sembra quasi regolare a fasi di comportamento molto caotico. Il punto $x = 0$ funziona da zona critica, rallentando il moto e innescando una crescita dell'entropia non lineare.

Nelle Figure 4.1 e 4.2 si confrontano i risultati di una stessa procedura di conteggio delle sottostringhe applicata sia alla Arnold's Cat Map sia alla Pomeau–Manneville, mantenendo identici il numero di iterazioni e l'intervallo di d .

In particolare:

- Arnold's Cat Map: $H(d)$ cresce inizialmente e poi si assesta su un plateau, coerente con

$$\lim_{d \rightarrow \infty} \frac{H(d)}{d} = \alpha.$$

- Pomeau–Manneville: non compare un plateau stabile e la crescita non risulta semplicemente lineare in d ; in alcuni casi si osserva

$$\lim_{d \rightarrow \infty} \frac{H(d)}{d^\beta} = \alpha, \quad \text{con } \beta \neq 1.$$

Per i seguenti grafici vengono utilizzati ancora una volta i parametri: $N = 100000$, $min_dim = 2$, $max_dim = 100$.

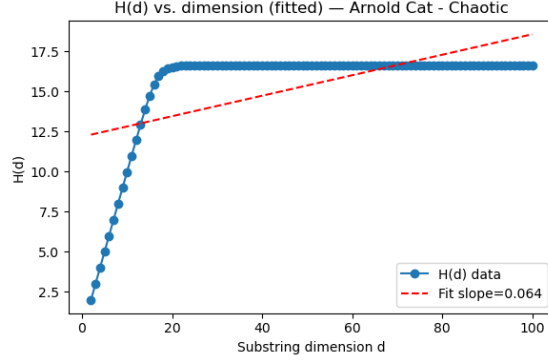


Figura 4.1: Andamento dell'entropia $H(d)$ per la Arnold's Cat Map: la curva mostra un plateau per valori elevati di d .

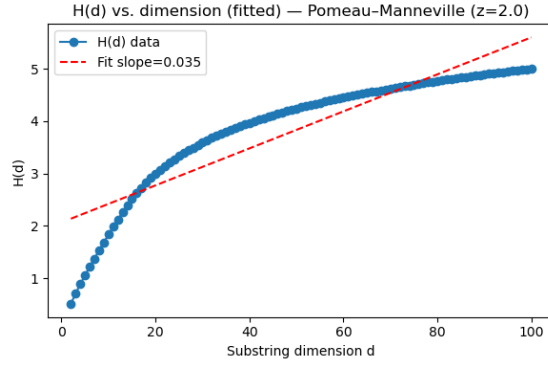


Figura 4.2: Esempio per la mappa di Pomeau–Manneville: la curva non rivela un plateau nello stesso intervallo di d .

Lo studio di altre mappe che mostrano crescita non lineare dell'entropia, o addirittura leggi di potenza $H(d) \sim d^\beta$ con $\beta \neq 1$, rappresenta un percorso interessante per approfondire in futuro, in generale, tutti i sistemi caratterizzati da esponenti diversi da 1, e richiede analisi più approfondite allo scopo di determinare con precisione l'esponente β che ne governa la crescita.

Bibliografia

- [1] G. Benettin et al. «Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory». In: *Meccanica* 15 (1980), pp. 9–20.
- [2] T. M. Cover. *Elements of information theory*. John Wiley & Sons, 1999.
- [3] J. M. Franks. «Invariant sets of hyperbolic toral automorphisms». In: *American Journal of Mathematics* 99.5 (1977), pp. 1089–1095.
- [4] T. Goldfriend e J. Kurchan. «Quantum kolmogorov-sinai entropy and pesin relation». In: *Physical Review Research* 3.2 (2021), p. 023234.
- [5] A. V. Igorevich. *Mathematical methods of classical mechanics*. Vol. 60. Springer Science & Business Media, 2013.
- [6] E. N. Lorenz. «Deterministic nonperiodic flow». In: *Journal of atmospheric sciences* 20.2 (1963), pp. 130–141.
- [7] D. J. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
- [8] R. Mañé. «A proof of Pesin’s formula». In: *Ergodic Theory and Dynamical Systems* 1.1 (1981), pp. 95–102.
- [9] S. E. Newhouse. «Lectures on dynamical systems». In: *Dynamical Systems: Lectures given at a Summer School of the Centro Internazionale Matematico Estivo (CIME), held in Bressanone (Bolzano), Italy, June 19–27, 1978* (1980), pp. 209–312.
- [10] Y. Pomeau e P. Manneville. «Intermittent transition to turbulence in dissipative dynamical systems». In: *Communications in Mathematical Physics* 74 (1980), pp. 189–197.
- [11] P. C. Shields. *The ergodic theory of discrete sample paths*. Vol. 13. American Mathematical Soc., 1996.
- [12] S. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. 2018.
- [13] J. Ziv e A. Lempel. «Compression of individual sequences via variable-rate coding». In: *IEEE transactions on Information Theory* 24.5 (1978), pp. 530–536.