TEORIA ERGODICA

PERCORSO DI ECCELLENZA DELLE LAUREE TRIENNALI
FACOLTÀ DI INGEGNERIA - UNIVERSITÀ DI PISA
A.A. 2011-2012

CLAUDIO BONANNO

Indice

1. Definizioni principali - 08/03/2012

Dato un sistema di natura fisica, biomedica, economica, o altro, vogliamo studiarne il comportamento, caratterizzando le sue possibili "orbite" e la misura di osservabili. Cercheremo anche di distinguere i possibili comportamenti dei sistemi di natura casuale e di quelli di natura deterministica, cercando caratteristiche che ci permettano di identificare la natura di un sistema basandoci solo sulla conoscenza delle sue orbite.

Iniziamo introducendo lo *spazio degli stati* $\mathcal{X}$ di un sistema. Si tratta dell'insieme dei possibili stati che può assumere un sistema e che lo caratterizzino completamente.

Esempio 1.1. Vediamo alcuni esempi di sistemi e dei loro spazi degli stati.

- Lo spazio degli stati di un sistema che descrive il moto di un punto materiale nello spazio $\mathbb{R}^3$ è dato dalle sue possibili posizioni e velocità. Quindi $\mathcal{X} = \mathbb{R}^3 \times \mathbb{R}^3$.
- Se vogliamo descrivere gli esiti di un lancio di una moneta, indichiamo con "T" e con "C le due possibilità testa e croce. Quindi $\mathcal{X} = \{T, C\}$. Possiamo, per fini di calcolo, usare anche una convenzione numerica, ponendo ad esempio testa=-1 e croce=+1, usando quindi $\mathcal{X} = \{-1, +1\}$.
- Analogamente, se vogliamo descrivere gli esiti di un lancio di un dado, possiamo usare come spazio degli stati $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$. Vedremo che può essere interessante interpretare il lancio di un dado anche come scelta di una direzione in $\mathbb{R}^3$, ponendo quindi
$$\mathcal{X} = \left\{ \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$
- Infine, consideriamo il caso di un sistema che descriva l'evoluzione del numero di membri di una popolazione. Nel caso per esempio di risorse finite, esiste un limite massimo M alla numerosità di una popolazione, quindi possiamo usare come spazio degli stati l'intervallo $\mathcal{X} = [0, M]$. Normalizzando le rilevazioni possiamo considerare $\mathcal{X} = [0, 1]$.

Lo studio della dinamica di un sistema consiste nello studio dell'evoluzione dello stato di un sistema. Consideriamo solo il caso in cui l'evoluzione di un sistema viene monitorata a intervalli di tempo regolari, quindi registriamo lo stato di un sistema solo in tempi discreti. Il passare del tempo è quindi misurato dall'insieme discreto $\mathbb{N}$.

Dato uno spazio degli stati $\mathcal{X}$, sia $x_0 \in \mathcal{X}$ lo stato iniziale di un sistema. L'insieme degli stati che il sistema assume con il passare del tempo costituisce la sua orbita. Quindi l'*orbita* di un sistema dinamico è un insieme della forma

$$\mathcal{O} = \{x_0, x_1, \dots, x_n, \dots\}, \quad \text{dove } x_n \in \mathcal{X} \text{ per ogni } n \in \mathbb{N}.$$

Un concetto fondamentale nello studio delle proprietà di un sistema dinamico è quello di "probabilità" per un sistema di trovarsi in un determinato stato o in un insieme di stati. Trattiamo separatamente il caso in cui lo spazio degli stati sia un insieme discreto o un intervallo di $\mathbb{R}$.

Sia $\mathcal{X}$ un insieme discreto di cardinalità C, ad esempio scriviamo $\mathcal{X} = \{s_1, s_2, \dots, s_C\}$.

Definizione 1.1 (**Probabilità nel caso discreto**)**.** Una *distribuzione di probabilità* P sull'insieme $\mathcal{X} = \{s_1, \dots, s_C\}$ è un insieme discreto $P = \{p_1, \dots, p_C\}$ di numeri reali con le proprietà:

(i) $p_i \in [0, 1]$ per ogni $i = 1, \dots, C$;
(ii) $p_1 + p_2 + \cdots + p_C = 1$.

La *probabilità di uno stato* $s_i \in \mathcal{X}$ è data da $P(s_i) := p_i$. Dato un sottoinsieme $A \subseteq \mathcal{X}$, la *probabilità di* A è data da

$$P(A) = \sum_{s_i \in A} P(s_i)$$

Da cui segue che $P(\mathcal{X}) = 1$.

Esempio 1.2. Consideriamo il caso del lancio di una moneta con $\mathcal{X} = \{T, C\}$. Una distribuzione di probabilità è allora un insieme $P = \{p_T, p_C\}$ con $p_T + p_C = 1$ e $p_T, p_C \geq 0$. Una moneta si dice *onesta* se $p_T = p_C = \frac{1}{2}$.

♢

Nel caso di intervalli di $\mathbb{R}$, consideriamo solo il caso di probabilità "assolutamente continue".

Definizione 1.2 (**Probabilità nel caso continuo**)**.** Una *distribuzione di probabilità assolutamente continua* P su un insieme $\mathcal{X} \subseteq \mathbb{R}$ è definita da una funzione reale $f(x)$, con $x \in \mathcal{X}$, che sia integrabile[1], non-negativa, tale che per ogni $A \subseteq \mathcal{X}$ con $\int_A 1dx = 0$ si ha $\int_A f(x)dx = 0$, e con la proprietà che

$$\int_{\mathcal{X}} f(x)\, dx = 1$$

La *probabilità di un intervallo* $A \subseteq \mathcal{X}$ è data da

$$P(A) := \int_A f(x)\, dx$$

e soddisfa tutte le proprietà dell'integrale per somme e intersezioni di insiemi. Ne segue che per ogni stato $x \in \mathcal{X}$ si ha $P(x) = 0$.

Esempio 1.3 (**La distribuzione uniforme**)**.** Nel caso dello studio dell'evoluzione di una popolazione, abbiamo considerato $\mathcal{X} = [0, 1]$. Un esempio di distribuzione di probabilità assolutamente continua P è la cosiddetta *distribuzione uniforme* data da $f(x) \equiv 1$. In questo caso per ogni intervallo $A = [a, b] \subseteq \mathcal{X}$ si ha

$$P(A) = \int_a^b 1\, dx = b - a$$

♢

Esempio 1.4 (**La distribuzione normale**)**.** Nel caso $\mathcal{X} = \mathbb{R}$ un esempio di distribuzione di probabilità particolarmente importante è la cosiddetta *distribuzione normale di parametri* m *e* $\sigma > 0$ data da

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

La distribuzione normale con $m = 0$ e $\sigma = 1$ si chiama *standard*.

♢

2. Definizioni principali - 15/03/2012

Il passo successivo nella definizione dei concetti fondamentali per la caratterizzazione di un sistema, è lo studio della relazione che intercorre tra lo stato di un sistema a un tempo n e lo stato del sistema al tempo $n+1$. Diremo che un sistema è *casuale* se dato lo stato di un sistema al tempo n non è possibile determinare in maniera univoca lo stato del sistema al tempo $n + 1$. Diremo che un sistema è *deterministico* se invece è possibile farlo. Nello studio dei sistemi casuali conviene considerare una distribuzione di probabilità sull'insieme delle orbite.

Introduciamo prima la definizione di σ*-algebra* e *misura di probabilità* su uno spazio qualsiasi.

Definizione 2.1 (**Misura di probabilità**)**.** Dato un insieme X, si chiama σ*-algebra* $\mathcal{A}$ una famiglia di sottoinsiemi di X tale che:

(i) $X \in \mathcal{A}$;

[1]Consideriamo il caso dell'integrale di Lebesgue.

(ii) se $Y \in \mathcal{A}$ allora $Y^c \in \mathcal{A}$;
(iii) se $\{Y_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ allora $\cup_n Y_n \in \mathcal{A}$.

La coppia $(X, \mathcal{A})$ definisce un *insieme misurabile*.
Una *misura di probabilità* $\mathbb{P}$ su un insieme misurabile $(X, \mathcal{A})$ è una funzione $\mathbb{P} : \mathcal{A} \to [0, 1]$ tale che

(i) $\mathbb{P}(\emptyset) = 0$ e $\mathbb{P}(X) = 1$;
(ii) se $\{Y_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ sono a due a due disgiunti, allora

$$\mathbb{P}\Big(\bigcup_n Y_n\Big) = \sum_n \mathbb{P}(Y_n).$$

Dato un insieme $Y \in \mathcal{A}$, si chiama *probabilità di* Y il valore $\mathbb{P}(Y)$. Dati due insiemi $Y_1, Y_2 \in \mathcal{A}$ si chiama *probabilità condizionata di* Y_1 *rispetto a* Y_2, indicata con $\mathbb{P}(Y_1|Y_2)$ il valore

$$\mathbb{P}(Y_1|Y_2) := \frac{\mathbb{P}(Y_1 \cap Y_2)}{\mathbb{P}(Y_2)}$$

Le misure di probabilità introdotte nelle Definizioni 1.1 e 1.2 sono esempi di misure di probabilità.

Per i sistemi casuali si farà sempre riferimento a una misura di probabilità $\mathbb{P}$ sul cosiddetto *spazio delle realizzazioni* Ω. Dato uno spazio degli stati $\mathcal{X}$, consideriamo lo spazio prodotto

$$\mathcal{X}^{\mathbb{N}} := \{\omega = (\omega_0, \omega_1, \dots) \ : \ \omega_n \in \mathcal{X} \ \forall\, n \in \mathbb{N}\}$$

Definizione 2.2 (**Spazio delle realizzazioni**)**.** Sia $\mathcal{X}$ lo spazio degli stati discreto di un sistema casuale, si dice *spazio delle realizzazioni* Ω del sistema l'insieme delle sue orbite possibili

$$\Omega := \bigcup_{x_0\in\mathcal{X}} \omega(x_0) \subseteq \mathcal{X}^{\mathbb{N}}$$

dove $\omega(x_0) \in \mathcal{X}^{\mathbb{N}}$ indica l'orbita del sistema con condizione iniziale x_0, ossia $\omega_0 = x_0$.
Particolari sottoinsiemi di Ω sono i *cilindri*. Chiameremo cilindro di lunghezza $n \geq 1$, centro $(x_1, \dots, x_n) \in \mathcal{X}^n$ e posizione $k \geq 0$ l'insieme

$$\mathcal{C}_k(x_1, \dots, x_n) := \Big\{\omega \in \Omega \ : \ \omega_k = x_1, \, \omega_{k+1} = x_2, \dots, \, \omega_{k+n-1} = x_n\Big\}$$

Lo spazio Ω con la σ-algebra generata dai cilindri definisce lo spazio misurabile $(\Omega, \mathcal{C})$.
Una misura di probabilità $\mathbb{P}$ su $(\Omega, \mathcal{C})$ rende il sistema casuale *stazionario* se

$$\mathbb{P}\left(\mathcal{C}_k(x_1, \dots, x_n)\right) = \mathbb{P}\left(\mathcal{C}_{k+r}(x_1, \dots, x_n)\right)$$

per ogni $k, r \in \mathbb{N}$, ogni $n \geq 1$ e ogni $(x_1, \dots, x_n) \in \mathcal{X}^n$.

Esempio 2.1 (**Probabilità prodotto**)**.** Sia $\mathcal{X} = \{0, 1\}$ lo spazio degli stati del lancio di una moneta con $P = \{p_0, p_1\}$. Allora lo spazio delle realizzazioni

$$\Omega = \Big\{\omega = (\omega_0, \omega_1, \dots, \omega_n, \dots) \ : \ \omega_i \in \{0, 1\} \ \ \forall\, i \in \mathbb{N}\Big\} = \{0, 1\}^{\mathbb{N}}$$

è legato all'intervallo $[0, 1]$ tramite la funzione

$$\varphi : \Omega \to [0, 1], \qquad \varphi(\omega) = \sum_{i=0}^{\infty} \omega_i \, 2^{-i}$$

È facile dimostrare che φ è surgettiva ma non iniettiva. Inoltre questa funzione permette di definire su Ω una misura di probabilità, analoga della misura di Lebesgue su $[0, 1]$, data da

$$\mathbb{P}\left(\mathcal{C}_k(x_1, \dots, x_n)\right) = \prod_{i=1}^{n} P(x_i) \tag{2.1}$$

per ogni $k \in \mathbb{N}$, ogni $n \geq 1$ e ogni $(x_1, \ldots, x_n) \in \mathcal{X}^n$. La formula (2.1) definisce una *misura di probabilità prodotto*. Il sistema $(\Omega, \mathbb{P})$ così definito è stazionario. Nel caso di una moneta onesta, quindi con $p_0 = p_1 = \frac{1}{2}$, si ottiene

$$\mathbb{P}\left(\mathcal{C}_k(x_1, \ldots, x_n)\right) = 2^{-n}$$

che dipende quindi solo dalla lunghezza del cilindro e non dal centro.

♢

La scelta di $\mathbb{P}$ su Ω determina le proprietà di dipendenza dello stato di un sistema dallo stato precedente.

Indichiamo con $\{\Sigma_n\}_{n\geq 0}$ l'insieme delle funzioni

$$\Sigma_n : \Omega \to \mathcal{X}\,, \qquad \Sigma_n(\omega) := \omega_n \quad \forall\, n \geq 0 \tag{2.2}$$

che indicano lo stato del sistema con orbita ω al tempo n. Un sistema casuale si dice *a variabili a due a due indipendenti* se per ogni $n, m \geq 0$, $n \neq m$, e per ogni coppia di sottoinsiemi $A, B \subset \mathcal{X}$ si ha

$$\mathbb{P}\Big(\, \{\Sigma_n(\omega) \in A\} \,\Big|\, \{\Sigma_m(\omega) \in B\} \,\Big) = \mathbb{P}\Big(\, \{\Sigma_n(\omega) \in A\} \,\Big)$$

Questo corrisponde a

$$\mathbb{P}\Big(\, \{\Sigma_n(\omega) \in A\} \cap \{\Sigma_m(\omega) \in B\} \,\Big) = \mathbb{P}\Big(\, \{\Sigma_n(\omega) \in A\} \,\Big)\, \mathbb{P}\Big(\, \{\Sigma_n(\omega) \in B\} \,\Big) \tag{2.3}$$

Un sistema casuale si dice *a variabili indipendenti* se (2.3) continua a valere per ogni insieme $\{\Sigma_{n_1}, \ldots, \Sigma_{n_k}\}$ di funzioni. Un sistema Ω con misura $\mathbb{P}$ prodotto è a variabili indipendenti. Gli stati di un sistema possono essere legati in base alla distanza temporale tra le misurazioni. Si parla in questo caso di sistemi casuali *a memoria finita*. Un esempio sono le *catene di Markov*, sistemi casuali con memoria uno, ossia per ogni $n \geq 0$ e $1 \leq h \leq n$ si ha

$$\mathbb{P}\Big(\, \{\Sigma_n(\omega) \in A\} \,\Big|\, \{\Sigma_{n-1}(\omega) \in B_1\}, \ldots, \{\Sigma_{n-h}(\omega) \in B_h\} \,\Big) = \mathbb{P}\Big(\, \{\Sigma_n(\omega) \in A\} \,\Big|\, \{\Sigma_{n-1}(\omega) \in B_1\} \,\Big)$$

qualunque siano i sottoinsiemi $A, B_1, \ldots, B_h$ di $\mathcal{X}$.

Esempio 2.2 (**Passeggiata aleatoria**)**.** Un esempio di catena di Markov è il sistema, denominato *passeggiata aleatoria*, che descrive il moto per esempio uni-dimensionale in cui a ogni istante si fa un passo in una direzione o nell'altra a secondo del lancio di una moneta. Con il nostro formalismo si ha $\mathcal{X} = \mathbb{Z}$ e un elemento $\omega \in \mathbb{Z}^{\mathbb{N}}$ è una possibile orbita, quindi $\omega \in \Omega$, se $\omega_n = \omega_{n-1} \pm 1$ per ogni $n \geq 1$. Se $P = \{p_{-1}, p_1\}$ è la distribuzione di probabilità della moneta, si trova che per ogni $n \geq 0$ e $1 \leq h \leq n$, e per ogni $N, M_1, \ldots, M_h \in \mathbb{Z}$,

$$\mathbb{P}\Big(\, \{\Sigma_n(\omega) = N\} \,\Big|\, \{\Sigma_{n-1}(\omega) = M_1\}, \ldots, \{\Sigma_{n-h}(\omega) = M_h\} \,\Big) = \mathbb{P}\Big(\, \{\Sigma_n(\omega) = N\} \,\Big|\, \{\Sigma_{n-1}(\omega) = M_1\} \,\Big) =$$

$$= \begin{cases} p_1 & \text{se } N = M_1 + 1 \\ p_{-1} & \text{se } N = M_1 - 1 \\ 0 & \text{altrimenti} \end{cases}$$

Osserviamo che in quest'esempio stiamo considerando il caso di un sistema con spazio degli stati $\mathcal{X}$ infinito. Tuttavia, ponendo $\mathbb{P}(\mathcal{C}_0(N)) = p_N = 1$ per ogni $N \in \mathcal{X}$, otteniamo un sistema $(\Omega, \mathbb{P})$ stazionario, con $\sum_{N\in\mathcal{X}} p_N = +\infty$, e quindi $\mathbb{P}(\Omega) = +\infty$.

Nel seguito considereremo solo il caso di sistemi casuali a variabili indipendenti.

3. Passeggiate aleatorie per sistemi casuali - 22/03/2012

Sia $\Omega = \{-1, +1\}^{\mathbb{N}}$ lo spazio delle realizzazioni del sistema casuale che descrive gli esiti del lancio di una moneta con distribuzione di probabilità $P = \{p, q\}$, $p = P(+1)$ e $q = 1 - p = P(-1)$. Su Ω consideriamo la misura di probabilità prodotto definita nell'esempio 2.1. Il sistema $(\Omega, \mathbb{P})$ è stazionario e a variabili indipendenti, con

$$\mathbb{P}\Big(\{\Sigma_n(\omega) = \pm 1\} \Big) = \begin{cases} p \\ q \end{cases}, \quad \forall\, n \geq 0$$

Il problema che studiamo è il comportamento delle somme

$$S_n : \Omega \to \mathbb{Z}, \qquad S_n(\omega) := \sum_{k=0}^{n-1} \Sigma_k(\omega) = \sum_{k=0}^{n-1} \omega_k\,, \quad \forall\, n \geq 1 \tag{3.1}$$

Osserviamo innanzitutto che l'insieme $\{S_n(\omega) = r\}$ è non vuoto solo se esistono $s, t \in \mathbb{N}$ per cui $n = s + t$ e $r = s - t$. In questo caso si ottiene usando le nozioni del calcolo combinatorio

$$\mathbb{P}\Big(\{S_n(\omega) = r\} \Big) = \binom{n}{s} p^s\, q^t = \binom{n}{\frac{n+r}{2}} p^s\, q^t \qquad \text{dove } n = s + t,\ r = s - t \tag{3.2}$$

Definizione 3.1 (**Valor medio**)**.** Sia G una funzione su uno spazio misurabile $(X, \mathcal{A})$ con misura di probabilità $\mathbb{P}$, a valori in $\mathbb{Z}$, e tale che $\{G(x) = r\} \in \mathcal{A}$ per ogni $r \in \mathbb{Z}$. Si definisce *valor medio di* G il valore

$$\mathbb{E}_{\mathbb{P}}[G] := \sum_{r \in \mathbb{Z}} r\, \mathbb{P}\Big(\{G(x) = r\} \Big)$$

Si verifica che il valor medio è un operatore lineare, ossia $\mathbb{E}_{\mathbb{P}}[G_1 + G_2] = \mathbb{E}_{\mathbb{P}}[G_1] + \mathbb{E}_{\mathbb{P}}[G_2]$. Nel nostro caso

$$\mathbb{E}_{\mathbb{P}}[\Sigma_n] = p - q, \quad \mathbb{E}_{\mathbb{P}}[S_n] = n(p - q), \qquad \forall\, n \tag{3.3}$$

3.1. **Probabilità di successi.** Supponiamo di puntare sull'uscita di "testa", che corrisponde a $+1$ nel nostro modello. Allora in n lanci, qual è la probabilità di avere k successi? La risposta si ottiene facilmente definendo la funzione

$$T_n : \Omega \to \mathbb{N}, \qquad T_n(\omega) := \frac{1}{2}\,(S_n(\omega) + n) = \frac{1}{2} \sum_{k=0}^{n-1} (\omega_k + 1)$$

Si verifica che $T_n(\omega) \leq n$ e, usando le nozioni del calcolo combinatorio,

$$\mathbb{P}\Big(\{T_n(\omega) = k\} \Big) = \binom{n}{k} p^k\, q^{n-k}\,, \qquad \forall\, k = 0, \ldots, n \tag{3.4}$$

Per le proprietà del valor medio, si trova da (3.3) che

$$\mathbb{E}_{\mathbb{P}}[T_n] = \frac{1}{2}\, \mathbb{E}_{\mathbb{P}}[S_n + n] = \frac{n(p - q) + n(p + q)}{2} = np \tag{3.5}$$

Analogamente, la probabilità di insuccessi si ottiene studiando la funzione

$$C_n : \Omega \to \mathbb{N}, \qquad C_n(\omega) := \frac{1}{2}\,(n - S_n(\omega)) = \frac{1}{2} \sum_{k=0}^{n-1} (1 - \omega_k)$$

per la quale si trova come prima

$$\mathbb{P}\Big(\{C_n(\omega) = h\} \Big) = \binom{n}{h} p^{n-h}\, q^h\,, \qquad \forall\, h = 0, \ldots, n \tag{3.6}$$

$$\mathbb{E}_{\mathbb{P}}[C_n] = \frac{1}{2}\,\mathbb{E}_{\mathbb{P}}[n - S_n] = \frac{n(p+q) - n(p-q)}{2} = nq \tag{3.7}$$

3.2. **Ritorni in $r = 0$ per monete oneste.** Studiamo il comportamento dei ritorni in $r = 0$. Per $r = 0$ chiaramente $\mathbb{P}\Big(\{S_n(\omega) = 0\}\Big)$ è positiva solo per n pari. Poniamo quindi $n = 2m$. Si trova

$$\mathbb{P}\Big(\{S_{2m}(\omega) = 0\}\Big) = \left(\begin{array}{c} 2m \\ m \end{array}\right) (pq)^m$$

che si può approssimare, usando la *formula di Stirling*

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n} \tag{3.8}$$

con

$$\mathbb{P}\Big(\{S_{2m}(\omega) = 0\}\Big) \sim \frac{1}{\sqrt{\pi}}\,\frac{(4p(1-p))^m}{\sqrt{m}}$$

Restringiamoci al caso $p = q = \frac{1}{2}$, per cui si trovano i valori nella tabella

n	2	4	6	8	10	20	50	100
$\mathbb{P}(\{S_n = 0\})$	0.5	0.375	0.312	0.273	0.246	0.176	0.112	0.079
$\sqrt{\frac{2}{\pi\, n}}$	0.564	0.399	0.326	0.282	0.252	0.178	0.113	0.079

Si vede che arrivare precisamente in $r = 0$ dopo un numero fissato di passi è un evento poco probabile, e lo stesso vale per il ritorno in un intervallo fissato di $r = 0$. Tuttavia, la probabilità di tornare "prima o poi" in $r = 0$ è 1, ossia l'evento è certo. Questa e altre interessanti proprietà delle somme si possono stabilire con metodi elementari. Poniamo

$$\mathtt{p}_n := \mathbb{P}\Big(\{S_k(\omega) \neq 0 \ : \ k = 1, \dots, n-1\}\,,\, \{S_n(\omega) = 0\}\Big)$$

la probabilità che il primo ritorno in $r = 0$ avvenga al tempo n. Si verifica che

$$\mathtt{p}_{2m} = \frac{\mathbb{P}(\{S_{2m}(\omega) = 0\})}{2m-1}, \qquad \sum_{m=1}^{\infty} \mathtt{p}_{2m} = 1 \tag{3.9}$$

Nella tabella sotto calcoliamo le somme finite $\sum_{m=1}^{n/2} \mathtt{p}_{2m}$.

n	10	20	40	80	100	200	400	1000
$\sum_{m=1}^{n/2} \mathtt{p}_{2m}$	0.754	0.824	0.875	0.911	0.920	0.944	0.960	0.975

Esempio 3.1 (**Ritorni in più dimensioni**)**.** Consideriamo ora una passeggiata aleatoria in spazi a più dimensioni. Per esempio, abbiamo visto come con un dado è possibile costruire una passeggiata aleatoria in $\mathbb{R}^3$. Si trova (Polya, 1921) che la probabilità di ritorno nell'origine $\mathtt{p}(d)$ per una passeggiata aleatoria in dimensione d dipende dalla dimensione, in particolare

d	2	3	4	5	6	7	8
$\mathtt{p}(d)$	1	0.3405	0.1932	0.1351	0.1047	0.0858	0.0729

♢

Concludiamo lo studio dei ritorni in $r = 0$ con un risultato interessante sulle permanenze nel lato positivo o negativo.

Teorema 3.2 (**Legge dell'arcoseno**)**.** *Per ogni $0 \leq \theta_1 < \theta_2 \leq 1$ si ha*

$$\mathbb{P}\left(\frac{Card\,\{k \in [1, 2n] \ : \ S_k(\omega) \geq 0\}}{2n} \in (\theta_1, \theta_2)\right) \sim \int_{\theta_1}^{\theta_2} \frac{1}{\pi\sqrt{x(1-x)}}\, dx$$

In particolare per ogni $\theta \in (0,1)$ si ha

$$\mathbb{P}\left(\frac{Card\{k \in [1,2n] \ : \ S_k(\omega) \geq 0\}}{2n} \in [0,\theta)\right) \sim \int_0^\theta \frac{1}{\pi\sqrt{x(1-x)}}\,dx = \frac{2}{\pi}\,\arcsin\sqrt{\theta}$$

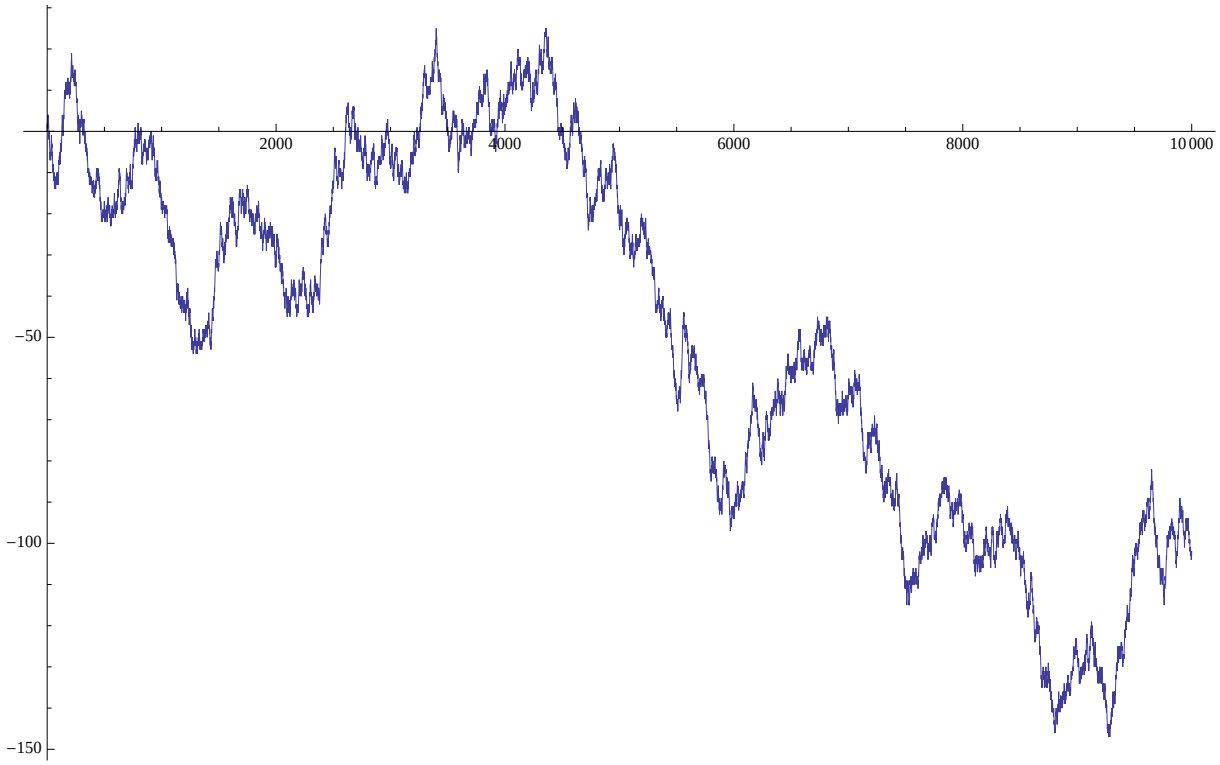

FIGURA 1. Una passeggiata aleatoria nel caso $p = q = \frac{1}{2}$.

4. Leggi limite per sistemi casuali - 29/03/2012

Vediamo adesso alcune proprietà delle funzioni S_n, T_n e C_n. Abbiamo visto che la probabilità di avere un ugual numero di teste e croci in un numero n fissato di lanci è asintoticamente (per n che tende a infinito) nulla, e di più

$$\lim_{n\to\infty} \mathbb{P}\Big(\{S_n(\omega) \in (-k,k)\}\Big) = 0\,, \qquad \forall\, k \in \mathbb{Z}$$

Ma abbiamo anche visto che la probabilità che i due numeri siano uguali "prima o poi" è uno. Allora quello che ci possiamo aspettare è che asintoticamente metà dei lanci avrà testa come esito e metà croce. Questo si traduce nello studio dei rapporti

$$\frac{S_n(\omega)}{n}, \quad \frac{T_n(\omega)}{n} \quad \text{e} \quad \frac{C_n(\omega)}{n}$$

e ci aspettiamo che per ogni $\varepsilon > 0$ abbiano grandi probabilità gli insiemi

$$\left\{\frac{S_n(\omega)}{n} \in (-\varepsilon,\varepsilon)\right\} = \left\{\frac{T_n(\omega)}{n} \in \left(\frac{1}{2}(1-\varepsilon),\, \frac{1}{2}(1+\varepsilon)\right)\right\} = \left\{\frac{C_n(\omega)}{n} \in \left(\frac{1}{2}(1-\varepsilon),\, \frac{1}{2}(1+\varepsilon)\right)\right\}$$

Il primo risultato che si può ottenere è il seguente

Teorema 4.1 (**Legge debole dei grandi numeri**). *Per ogni $\varepsilon > 0$ si ha*

$$\mathbb{P}\left(\left\{\frac{T_n(\omega)}{n} \leq \left(\frac{1}{2} - \varepsilon\right)\right\}\right) \leq \frac{1}{2}\, e^{-2n\varepsilon^2}$$

per n abbastanza grande. Quindi, per simmetria, si ha

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{\frac{T_n(\omega)}{n} \in \left(\frac{1}{2} - \varepsilon,\, \frac{1}{2} + \varepsilon\right)\right\}\right) = 1$$

Il risultato si può rendere più forte e vale in generale per sistemi casuali a variabili indipendenti.

Teorema 4.2 (**Legge forte dei grandi numeri**). *Sia $\{\Sigma_k\}_k$ una successione di funzioni su uno spazio di probabilità $(\Omega, \mathbb{P})$ per cui possiamo definire il valor medio. Se le Σ_k sono indipendenti e hanno la stessa legge, in particolare si ha $\mathbb{E}_{\mathbb{P}}[\Sigma_k] = m$ per ogni k, allora*

$$\mathbb{P}\left(\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \Sigma_k(\omega) = m\right\}\right) = 1$$

Equivalentemente, per ogni $\varepsilon > 0$ si ha

$$\mathbb{P}\left(\left\{\left|\frac{1}{n}\sum_{k=0}^{n-1} \Sigma_k(\omega) - m\right| > \varepsilon \text{ infinite volte}\right\}\right) = 0$$

Applicato al nostro caso, si trova

$$\mathbb{P}\left(\left\{\lim_{n\to\infty} \frac{S_n}{n} = 0\right\}\right) = \mathbb{P}\left(\left\{\lim_{n\to\infty} \frac{T_n}{n} = \frac{1}{2}\right\}\right) = \mathbb{P}\left(\left\{\lim_{n\to\infty} \frac{C_n}{n} = \frac{1}{2}\right\}\right) = 1 \tag{4.1}$$

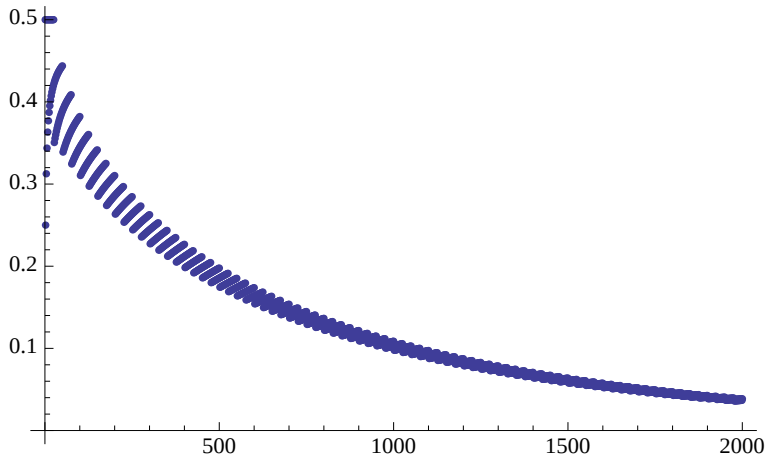

FIGURA 2. La probabilità di $\left\{\frac{T_n}{n} < n\left(\frac{1}{2} - \frac{1}{50}\right)\right\}$.

Studiamo adesso la "diffusione" tipica di una passeggiata aleatoria, ossia quanto ci dobbiamo aspettare che i valori di S_n si allontanino tipicamente dal valor medio. Strumenti utili sono la "varianza" e la "deviazione standard" di una funzione.

Definizione 4.3 (**Varianza e deviazione standard**). Sia G una funzione su uno spazio misurabile $(X, \mathcal{A})$ con misura di probabilità $\mathbb{P}$, a valori in $\mathbb{Z}$, e tale che $\{G(x) = r\} \in \mathcal{A}$ per ogni $r \in \mathbb{Z}$. Si definisce *varianza di G* il valore

$$\mathbb{V}_{\mathbb{P}}[G] := \mathbb{E}_{\mathbb{P}}\left[(G - \mathbb{E}_{\mathbb{P}}[G])^2\right] = \mathbb{E}_{\mathbb{P}}[G^2] - (\mathbb{E}_{\mathbb{P}}[G])^2 = \sum_{r\in\mathbb{Z}} r^2\, \mathbb{P}\Big(\{G(x) = r\}\Big) - \left(\sum_{r\in\mathbb{Z}} r\, \mathbb{P}\Big(\{G(x) = r\}\right)^2$$

e *deviazione standard di* G il valore $\sigma(G) = \sqrt{\mathbb{V}_{\mathbb{P}}[G]}$.

Osserviamo che la varianza verifica

$$\mathbb{V}_{\mathbb{P}}[G_1 + G_2] = \mathbb{V}_{\mathbb{P}}[G_1] + \mathbb{V}_{\mathbb{P}}[G_2]$$

solo per funzioni indipendenti.

Nel nostro caso troviamo in generale

$$\mathbb{V}_{\mathbb{P}}[\Sigma_n] = p + q - (p-q)^2 = 4pq, \qquad \sigma(\Sigma) = 2\sqrt{pq}, \quad \forall\, n \tag{4.2}$$

$$\mathbb{V}_{\mathbb{P}}[T_n] = \sum_{k=0}^{n} k^2 \begin{pmatrix} n \\ k \end{pmatrix} p^k\, q^{n-k} - (np)^2 = npq, \qquad \sigma(T_n) = \sqrt{npq} \tag{4.3}$$

e analogamente

$$\mathbb{V}_{\mathbb{P}}[S_n] = \mathbb{E}_{\mathbb{P}}[(2T_n - n)^2] - (n(p-q))^2 = 4npq, \qquad \mathbb{V}_{\mathbb{P}}[C_n] = npq \tag{4.4}$$

Teorema 4.4 (**Limite centrale**). *Sia* $\{\Sigma_k\}_k$ *una successione di funzioni su uno spazio di probabilità* $(\Omega, \mathbb{P})$ *per cui possiamo definire il valor medio e varianza. Se le* Σ_k *sono indipendenti e hanno la stessa legge, in particolare si ha* $\mathbb{E}_{\mathbb{P}}[\Sigma_k] = m$ *e* $\mathbb{V}_{\mathbb{P}}[\Sigma_k] = \sigma^2$ *per ogni* k*, allora per ogni* $x \in \mathbb{R}$ *si ha*

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{ \frac{\sum_{k=0}^{n-1} \Sigma_k(\omega) - nm}{\sigma\sqrt{n}} < x \right\}\right) = \int_{-\infty}^{x} f_{0,1}(t)\, dt$$

dove $f_{0,1}$ *è la distribuzione normale standard dell'esempio 1.4.*

Nel nostro caso troviamo quindi

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{ \frac{S_n(\omega) - n(p-q)}{2\sqrt{npq}} < x \right\}\right) = \int_{-\infty}^{x} f_{0,1}(t)\, dt \qquad \forall\, x \in \mathbb{R}$$

che con $p = q = \frac{1}{2}$, ci dice per esempio che

$$\mathbb{P}\left(\left\{ \frac{S_n(\omega)}{\sqrt{n}} \in (-3, 3) \right\}\right) \sim 0.9973$$

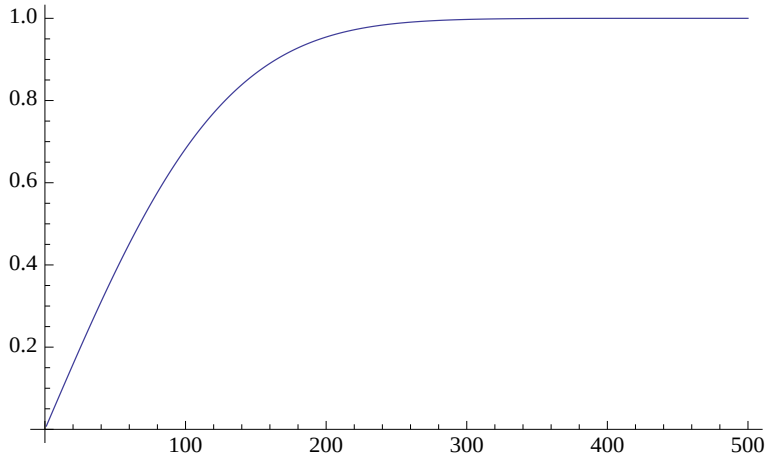

FIGURA 3. La probabilità di $\frac{S_n(\omega)}{\sqrt{n}} \in \left(-\frac{n}{100}, \frac{n}{100}\right)$.

Infine citiamo nel nostro caso, con $p = q = \frac{1}{2}$, la *Legge del Logaritmo Iterato*, che ci dice

$$\mathbb{P}\left(\left\{\limsup_{n\to\infty} \frac{S_n(\omega)}{\sqrt{2n\,\log(\log n)}} = 1\right\}\right) = 1 \tag{4.5}$$

o, equivalentemente, per ogni $\lambda > 1$

$$\mathbb{P}\left(\left\{S_n(\omega) > \lambda\sqrt{2n\,\log(\log n)} \ \text{ infinite volte}\right\}\right) = 0$$

Analogamente per il lim inf e le stime dal basso.

5. Qualche esempio di rovina del giocatore - 29/03/2012

Esempio 5.1 (**Il raddoppio a ogni giocata**)**.** Scommettiamo sull'uscita di "testa" su infiniti lanci di monete. La moneta abbia distribuzione di probabilità $P(T) = p$ e $P(C) = q$ e la vincita sia uguale alla somma scommessa. Seguiamo la strategia del raddoppio a ogni giocata fino alla vittoria: "a ogni giocata scommettiamo quello che abbiamo perso fino a quel momento più 1 euro e ci fermiamo se vinciamo". Introduciamo le notazioni:

$$\mathtt{p}_k = \text{probabilità di aver perso nei primi } (k-1) \text{ lanci e di vincere al } k\text{-esimo lancio} = q^{k-1}\,p$$

$$\mathtt{P}_n = \sum_{k=1}^{n} \mathtt{p}_k = \text{probabilità di vincere in al più } n \text{ lanci} = \sum_{k=1}^{n} q^{k-1}\,p = p\,\frac{1-q^n}{1-q} = 1 - q^n$$

Si vede che $\mathtt{P}_n \to 1$, quindi la vincita è un evento certo in infiniti lanci, ma vediamo qual è la situazione se consideriamo il denaro. Sia D_n la funzione denaro definita per ogni $n \geq 1$ come

$$D_n : \{T, C\}^n \to \mathbb{Z}, \qquad D_n(\omega_0^{n-1}) = \begin{cases} 1 & \text{se } \omega_k = T \text{ per qualche } k = 0, \dots, n-1 \\ -(2^n - 1) & \text{se } \omega_k = C \text{ per ogni } k = 0, \dots, n-1 \end{cases}$$

che rappresenta la variazione di denaro dopo n lanci. Per definizione si trova

$$\mathbb{P}\Big(D_n = 1\Big) = \mathtt{P}_n, \qquad \mathbb{P}\Big(D_n = -(2^n - 1)\Big) = 1 - \mathtt{P}_n$$

Quindi per il valor medio si trova

$$\mathbb{E}_{\mathbb{P}}[D_n] = (1 - q^n) - q^n(2^n - 1) = 1 - (2q)^n = \begin{cases} 0 & \text{se } q = \frac{1}{2} \\ < 0 & \text{se } q > \frac{1}{2} \end{cases}$$

Se ne deduce che ripetere la strategia del raddoppio non è mai vincente.

♢

Esempio 5.2 (**Il gioco con scopo fissato**)**.** Supponiamo che le funzioni Σ_n rappresentino adesso la variazione di denaro del giocatore a ogni giocata. Poniamo $\Omega = \{T, C\}^{\mathbb{N}}$ e $\Sigma_n(\omega) = +1$ indica la vincita, che avviene con probabilità p, e $\Sigma_n(\omega) = -1$ indica la perdita, che avviene con probabilità q. La somma $S_n(\omega)$ indica quindi la variazione totale di denaro dopo n giocate.

Vogliamo studiare le probabilità

$$\mathtt{P}(r, t) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} \{r + S_k(\omega) = t, \qquad 0 < r + S_h(\omega) < t \quad \forall\, h = 1, \dots, k-1\}\right), \qquad 0 < r < t$$

che rappresenta la probabilità di riuscire a raggiungere prima o poi t partendo da r, senza mai arrivare a 0. È intuitivo che

$$\mathtt{P}(r, t) = p\,\mathtt{P}(r+1, t) + q\,\mathtt{P}(r-1, t)\,, \qquad 0 < r < t$$

con le condizioni $\mathtt{P}(0,t)=0$ e $\mathtt{P}(t,t)=1$. Dalla teoria delle successioni per ricorrenza, si trova che le soluzioni sono della forma

$$\mathtt{P}(r,t)=\begin{cases} c_1+c_2\left(\frac{q}{p}\right)^r & \text{se } p\neq q \\ c_1+c_2 r & \text{se } p=q \end{cases}$$

con le costanti c_1 e c_2 determinate dalle condizioni al bordo. Si trova

$$\mathtt{P}(r,t)=\begin{cases} \frac{\left(\frac{q}{p}\right)^r-1}{\left(\frac{q}{p}\right)^t-1} & \text{se } p\neq q \\ \frac{r}{t} & \text{se } p=q \end{cases}$$

Consideriamo per esempio il caso $p=\frac{18}{37}$ e $q=\frac{19}{37}$, allora ponendo $t=r+N$ e otteniamo

N	1	5	10	25	50	100
$\mathtt{P}(100,100+N)$	0.947	0.762	0.581	0.258	0.066	0.004
$\mathtt{P}(500,500+N)$	0.947	0.763	0.582	0.259	0.067	0.004

♢

6. Sistemi deterministici - 12/04/2012

I sistemi deterministici possono essere definiti dallo spazio degli stati $\mathcal{X}$ e da una funzione $T:\mathcal{X}\to\mathcal{X}$ che determina in maniera univoca la relazione tra gli stati di un sistema a intervalli di tempo successivi. In questo caso lo spazio delle realizzazioni è in realtà in bigezione con $\mathcal{X}$ stesso, e non risulta quindi particolarmente interessante nella descrizione di un sistema deterministico. Tuttavia, osserviamo che se T non è una funzione iniettiva, è possibile che orbite di due condizioni iniziali diverse coincidano da un certo istante in poi.

Un *sistema deterministico* è quindi definito da uno spazio misurabile $(\mathcal{X},\mathcal{A})$ e da una funzione $T:\mathcal{X}\to\mathcal{X}$ *misurabile*, ossia tale che per ogni $A\in\mathcal{A}$ si ha $T^{-1}(A)\in\mathcal{A}$.

Richiamiamo alcune semplici definizioni di base nel caso generale di una funzione T non invertibile. Dato un punto $x\in\mathcal{X}$, si definisce *orbita* di x rispetto al sistema dinamico $(\mathcal{X},T)$, l'insieme

$$\mathcal{O}(x)=\{T^n(x)\ :\ n\geq 0\}$$

Definizione 6.1 (**Punto periodico**)**.** Un punto x si dice *periodico* se esiste un intero $p(x)>0$ tale che $T^{p(x)}(x)=x$. L'intero $p(x)$ si chiama *periodo* se vale anche che per ogni $1\leq k<p(x)$ si ha $T^k(x)\neq x$. Un punto periodico di periodo $p(x)=1$ si dice *fisso*, verificando $T(x)=x$.

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

Definizione 6.2 (**Insieme invariante**)**.** Un insieme $A\in\mathcal{A}$ si dice *invariante* per il sistema $(\mathcal{X},T)$ se $T^{-1}(A)=A$.

Definizione 6.3 (**Misura invariante**)**.** Una misura di probabilità $\mathbb{P}$ su $(\mathcal{X},\mathcal{A})$ si dice *invariante per* T se per ogni $A\in\mathcal{A}$ si ha $\mathbb{P}(T^{-1}(A))=\mathbb{P}(A)$

Esempio 6.1 (**Mappa di Bernoulli**)**.** Sia $\mathcal{X}=[0,1]$ con σ-algebra generata dagli intervalli aperti di $[0,1]$. Consideriamo la funzione

$$T(x)=2x\ (\text{mod } 1)\ =\begin{cases} 2x & \text{se } x<\frac{1}{2} \\ 2x-1 & \text{se } x\geq\frac{1}{2} \end{cases}$$

Una misura di probabilità invariante è la distribuzione uniforme dell'esempio 1.3.

♢

Esempio 6.2 (**Rotazioni del cerchio**)**.** Sia $\mathcal{X} = [0,1]$ con σ-algebra generata dagli intervalli aperti di $[0,1]$. Consideriamo la funzione con $\alpha \in [0,1] \setminus \mathbb{Q}$

$$T_\alpha(x) = x + \alpha \pmod 1 = \begin{cases} x+\alpha & \text{se } x < 1-\alpha \\ x+\alpha-1 & \text{se } x \geq 1-\alpha \end{cases}$$

In questo caso la distribuzione uniforme è l'unica misura di probabilità invariante.

♢

Esempio 6.3 (**Mappa logistica**)**.** Sia $\mathcal{X} = [0,1]$ con σ-algebra generata dagli intervalli aperti di $[0,1]$. Consideriamo la funzione $T(x) = 4x(1-x)$. In questo caso una misura di probabilità invariante è data dalla funzione

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$$

♢

Iniziamo lo studio delle proprietà dei sistemi determinisitici, con particolare attenzione a quali sono i comportamenti delle orbite di un sistema.

Teorema 6.4 (**Ricorrenza di Poincaré**)**.** *Sia $(\mathcal{X}, T)$ un sistema deterministico con misura di probabilità $\mathbb{P}$ invariante, allora per ogni $A \in \mathcal{A}$ si ha*

$$\mathbb{P}\left(\left\{x \in A \ : \ T^n(x) \in A \ \text{ per infiniti } n \in \mathbb{N}\right\}\right) = \mathbb{P}(A)$$

In particolare se definiamo il *tempo di primo ritorno* di un punto x di un insieme $A \in \mathcal{A}$ in sé stesso come

$$\varphi_A : A \to \mathbb{N} \cup \{\infty\}, \qquad \varphi_A(x) := \min\{n \geq 1 \ : \ T^n(x) \in A\} \tag{6.1}$$

allora si ha $\varphi_A(x) < \infty$ per $\mathbb{P}$-q.o. $x \in A$.

7. Sistemi ergodici - 19/04/12

Come nel caso delle passeggiate aleatorie, consideriamo il comportamento delle somme di funzioni definite su $\mathcal{X}$. Data una funzione $f : \mathcal{X} \to \mathbb{R}$ definiamo le *somme di Birkhoff di f* come

$$S_n f(x) := \sum_{k=0}^{n-1} f(T^k(x)) \tag{7.1}$$

Ricordando la definizione degli spazi

$$L^p(\mathcal{X}, \mathbb{P}) := \left\{ f : \mathcal{X} \to \mathbb{R} \ : \ \int_X |f(x)|^p \, d\mathbb{P}(x) < \infty \right\}, \qquad p \in [1, \infty)$$

si ha

Teorema 7.1 (**di Birkhoff**)**.** *Sia $(\mathcal{X}, T)$ un sistema deterministico con misura di probabilità $\mathbb{P}$ invariante, per ogni $f \in L^1(\mathcal{X}, \mathbb{P})$ vale*

a) *per $\mathbb{P}$-q.o. $x \in \mathcal{X}$ il limite $\tilde{f}(x) := \lim_{n\to\infty} \frac{1}{n} S_n f(x)$ esiste finito;*
b) *$\tilde{f}$ è una funzione invariante su $\mathcal{X}$, ossia $\tilde{f}(T(x)) = \tilde{f}(x)$ quando è definita;*
c) *$\tilde{f} \in L^1(\mathcal{X}, \mathbb{P})$;*
d) *per ogni insieme invariante $A \in \mathcal{A}$ si ha $\int_A f(x)\, d\mathbb{P}(x) = \int_A \tilde{f}(x)\, d\mathbb{P}(x)$.*

Scegliendo la funzione $f : \mathcal{X} \to \{-1, +1\}$ con $f|_A \equiv +1$ e $f|_{A^c} \equiv -1$, per un insieme $A \in \mathcal{A}$, si ottengono le passeggiate aleatorie in una dimensione che abbiamo visto nel caso dei sistemi casuali.

Definizione 7.2 (**Sistema ergodico**)**.** Una misura di probabilità $\mathbb{P}$ si dice *ergodica* per un sistema dinamico $(\mathcal{X}, T)$ se per ogni insieme T-invariante $A \in \mathcal{A}$ si ha $\mathbb{P}(A) = 0$ o $\mathbb{P}(A^c) = 0$.

Proposizione 7.3. *Una misura di probabilità $\mathbb{P}$ è ergodica per un sistema dinamico $(\mathcal{X}, T)$ se e solo se ogni $f \in L^1(\mathcal{X}, \mathbb{P})$ che soddisfa $f(T(x)) = f(x)$ per $\mathbb{P}$-q.o. $x \in \mathcal{X}$ è costante $\mathbb{P}$-q.o., ossia esiste $c \in \mathbb{R}$ tale che $f(x) = c$ per $\mathbb{P}$-q.o. $x \in \mathcal{X}$.*

Teorema 7.4 (**Ergodico di Birkhoff**)**.** *Sia $(\mathcal{X}, T)$ un sistema deterministico con misura di probabilità $\mathbb{P}$ invariante ed ergodica, allora per ogni $f \in L^1(\mathcal{X}, \mathbb{P})$ si ha*

$$\lim_{n\to\infty} \frac{1}{n} S_n f(x) = \int_{\mathcal{X}} f(x)\, d\mathbb{P}(x) \tag{7.2}$$

per $\mathbb{P}$-q.o. $x \in \mathcal{X}$.

Il Teorema Ergodico di Birkhoff è l'analogo della Legge Forte dei Grandi Numeri. Infatti (7.2) si può anche scrivere nella forma

$$\mathbb{P}\left(\left\{x \in \mathcal{X} \ : \ \lim_{n\to\infty} \frac{1}{n} S_n f(x) = \mathbb{E}_{\mathbb{P}}[f]\right\}\right) = 1$$

Corollario 7.5. *Sia $(\mathcal{X}, T)$ un sistema deterministico con misura di probabilità $\mathbb{P}$ invariante. La misura $\mathbb{P}$ è ergodica se e solo se per ogni $A, B \in \mathcal{A}$ vale*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}(T^{-k}(A) \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{7.3}$$

Osservazione 7.6. Notiamo che per un sistema ergodico si ottiene un'informazione aggiuntiva sul comportamento dei tempi di primo ritorno. Vale infatti la "Formula di Kac": sia $(\mathcal{X}, T)$ un sistema deterministico con misura di probabilità $\mathbb{P}$ invariante ed ergodica, allora per ogni $A \in \mathcal{A}$ si ha

$$\int_A \varphi_A(x)\, d\mathbb{P}(x) = 1 \tag{7.4}$$

dove φ_A è la funzione definita in (6.1).

La formula (7.4) si può interpretare dicendo che il valor medio del tempo di primo ritorno in A è uguale a $\frac{1}{\mathbb{P}(A)}$, ossia

$$\mathbb{E}_{\mathbb{P}|_A}[\varphi_A] = \frac{1}{\mathbb{P}(A)}$$

Esempio 7.1. Le misure invarianti descritte negli Esempi 6.1, 6.2 e 6.3 sono ergodiche.

Una nozione più forte dell'ergodicità è quella di *sistema miscelante*, un rafforzamento della proprietà (7.3).

Definizione 7.7 (**Sistema miscelante**)**.** Una misura di probabilità $\mathbb{P}$ si dice *miscelante* per un sistema dinamico $(\mathcal{X}, T)$ se per ogni $A, B \in \mathcal{A}$ vale

$$\lim_{k\to\infty} \mathbb{P}(T^{-k}(A) \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Esempio 7.2. Le misure invarianti descritte negli Esempi 6.1 e 6.3 sono miscelanti, quella dell'Esempio 6.2 non lo è.

Per alcuni sistemi miscelanti si può enunciare un analogo del Teorema del Limite Centrale.

Definizione 7.8 (**Proprietà di Limite Centrale**)**.** Sia $(\mathcal{X}, T)$ un sistema deterministico con misura di probabilità $\mathbb{P}$ invariante. Data una funzione $f : \mathcal{X} \to \mathbb{R}$ con $f \in L^2(\mathcal{X}, \mathbb{P})$, diciamo che f *soddisfa il Teorema del Limite Centrale rispetto a* $(\mathcal{X}, T)$ se per ogni $t \in \mathbb{R}$ si ha

$$\lim_{n\to\infty} \mathbb{P}\left(\left\{x \in \mathcal{X} \ : \ \frac{S_n f(x) - n \int_{\mathcal{X}} f(x)\, d\mathbb{P}(x)}{\sigma_f \sqrt{n}} \le t\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{t^2}{2}}\, dt \tag{7.5}$$

con

$$\sigma_f^2 := \lim_{n\to\infty} \frac{1}{n} \int_{\mathcal{X}} \Big(S_n f(x) - n \int_{\mathcal{X}} f(x)\, d\mathbb{P}(x)\Big)^2 d\mathbb{P}(x) \neq 0$$

Esempio 7.3. Osservabili $f \in L^2(\mathcal{X}, \mathbb{P})$ soddisfano il Teorema del Limite Centrale rispetto ai sistemi degli Esempi 6.1 e 6.3.

8. Dinamica e rappresentazione simbolica - 26/04/12

Descriviamo un esempio di sistema deterministico "astratto", che si può considerare l'esempio base di sistema caotico e introduce l'approccio ai sistemi deterministici tramite la Teoria dell'Informazione.

Esempio 8.1 (**Dinamica simbolica**)**.** Dato un insieme finito $S = \{1, \dots, N\}$, che consideriamo l'alfabeto, indichiamo con S^n l'insieme delle stringhe s di lunghezza n i cui simboli sono lettere di S. Indichiamo poi con S^* l'insieme di tutte le stringhe finite su S, ossia

$$S^* := \bigcup_{n\in\mathbb{N}} S^n \tag{8.1}$$

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

Sull'alfabeto S si possono considerare anche le stringhe infinite, che indicheremo con ω. Poniamo

$$\Omega := S^{\mathbb{N}} = \{\omega = (\omega_i)_{i\in\mathbb{N}} \ : \ \omega_i \in S \ \forall\ i \in \mathbb{N}\} \tag{8.2}$$

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

L'insieme Ω può essere munito della seguente metrica

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

dove δ è definita sulle lettere di S tramite $\delta(a, b) = 0$ se $a = b$, e $\delta(a, b) = 1$ se $a \neq b$, per cui due stringhe sono "vicine" se hanno "abbastanza" simboli iniziali uguali. Lo spazio Ω risulta essere uno spazio metrico compatto e un ruolo importante hanno gli insiemi

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

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

Consideriamo su Ω l'azione di una funzione τ che faccia "scorrere" la stringa, simulando in questo modo ad esempio la successiva produzione dei simboli della stringa. Questo definisce il cosiddetto *shift* $\tau : \Omega \to \Omega$, per cui la stringa $\tau(\omega)$ ha simboli dati da

$$\big(\tau(\omega)\big)_i = \omega_{i+i} \qquad \forall\ i \ge 0 \tag{8.3}$$

Proposizione 8.1. *Il sistema* (Ω, τ) *verifica le seguenti proprietà:*

(i) *la funzione* τ *è continua su* Ω*;*
(ii) *ha un'infinità numerabile di orbite periodiche;*

(iii) *ha un'infinità più che numerabile di orbite non periodiche;*
(iv) *ha un'orbita densa;*
(v) *ha dipendenza sensibile dalle condizioni iniziali, ossia esiste $\eta > 0$ tale che per ogni stringa ω e per ogni ε, esistono una stringa $\bar\omega$ e un intero n tale che $d(\omega, \bar\omega) < \varepsilon$ e $d(\tau^n\omega, \tau^n\bar\omega) > \eta$.*

Dimostrazione. (i) La continuità di τ segue facilmente dalla forma dei cilindri. Infatti basta verificare che $\tau^{-1}(C_k(\omega, n))$ è un aperto per ogni $\omega \in \Omega$ e ogni k, n. Basta quindi osservare che $\tau^{-1}(C_k(\omega, n)) = \cup_{i\in S} C_{k+1}(i\omega, n)$, dove $i\omega$ indica la concatenazione della lettera i con ω, e quindi è un aperto.

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

(iii) Le stringhe periodiche sono un'infinità numerabile, quindi le stringhe non periodiche sono un'infinità più che numerabile, avendo Ω la cardinalità del continuo. Infatti ogni orbita comprende una quantità numerabile di stringhe di Ω.

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

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

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

Per ogni k, vale $\#(S^k) = N^k$, e supponiamo di aver scelto una relazione d'ordine in S^k in modo da numerare le stringhe finite. Indichiamo con $s^k_1, \dots, s^k_{N^k}$ le stringhe di S^k. La stringa $\bar\omega$ che cerchiamo si ottiene allora come la concatenazione di tutte le stringhe di S^1 in ordine, concatenate con tutte le stringhe di S^2 scelte in ordine, e così via.

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

Le proprietà della Proposizione 8.1 fanno del sistema (Ω, τ) di shift un esempio di sistema dinamico "caotico".

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

$$\Omega_M := \big\{\omega \in \Omega\ :\ m_{\omega_i\ \omega_{i+1}} = 1 \text{ per ogni } i \geq 0\big\} \tag{8.4}$$

quindi gli elementi di M uguali a 1 dicono quali sono le coppie di simboli di S che possono essere prodotte.

Si dimostra che Ω_M è chiuso rispetto all'azione della funzione τ, che con la metrica indotta Ω_M è metrico compatto, e la σ-algebra corrisponde alla restrizione della σ-algebra dei cilindri. Se indichiamo con τ_M la restrizione di τ a Ω_M, otteniamo quindi il sistema deterministico (Ω_M, τ_M). Tale sistema dinamico si chiama comunemente *subshift di tipo finito.*

Le proprietà dinamiche di un subshift di tipo finito dipendono dalle proprietà della matrice di transizione, che stabilisce le dipendenze tra i simboli di S.

Definizione 8.2. *Una matrice di transizione M si dice* irriducibile *se per ogni coppia di simboli (i,j) esiste un intero k tale che $m_{ij}^k \neq 0$, dove $(m_{ij}^k)_{ij}$ sono gli elementi della matrice M^k, la k-esima potenza di M.*

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

Definizione 8.3. *Una matrice di transizione M si dice* irriducibile e aperiodica *se esiste un intero k tale che $m_{ij}^k \neq 0$ per ogni coppia di simboli (i,j).*

Proposizione 8.4. *Un subshift di tipo finito con matrice di transizione irriducibile e aperiodica verifica le proprietà (i)-(v) della Proposizione 8.1.*

♢

9. Rappresentazione simbolica e le sorgenti di informazione - 03/05/12

Due sistemi dinamici diversi possono avere le stesse caratteristiche dal punto di vista topologico.

Definizione 9.1. Siano $(\mathcal{X}_1, \mathcal{C}_1, T_1)$ e $(\mathcal{X}_2, \mathcal{C}_2, T_2)$ due sistemi deterministici. Se esiste $h : \mathcal{X}_1 \to \mathcal{X}_2$ surgettiva e misurabile tale che $h \circ T_1 = T_2 \circ h$ si dice che T_2 è un *fattore* di T_1 e che h è un *semi-coniugio*. Se h è anche iniettiva con inversa misurabile allora h si chiama *coniugio* e T_1 e T_2 si dicono *coniugati*.

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

Nel caso di un semi-coniugio si dice che T_2 è un fattore di T_1 perché in generale la dinamica di T_2 è meno "complicata". Infatti, mancando l'iniettività, h potrebbe mandare un'intero insieme invariante di $\mathcal{X}_1$ in un punto di $\mathcal{X}_2$, che quindi perderebbe la ricchezza della dinamica dell'insieme invariante. Si può addirittura pensare il caso estremo in cui $\mathcal{X}_2$ sia un unico punto e h manda tutto $\mathcal{X}_1$ in un punto.

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

Un coniugio topologico può essere visto come una rappresentazione topologica di un sistema dinamico tramite un'altro sistema di più facile interpretazione. Un caso particolare di rappresentazione di un sistema è la *rappresentazione simbolica*. Sia $(\mathcal{X}, \mathcal{A}, T)$ un sistema deterministico e sia $Z = \{I_1, \dots, I_N\}$ una partizione finita e misurabile dello spazio $\mathcal{X}$. Ossia gli insiemi $I_j \subset \mathcal{X}$ stanno in $\mathcal{A}$ e sono tali che $\cup_{j=1}^N I_j = \mathcal{X}$ e $I_i \cap I_j = \emptyset$ per $i \neq j$. Alla partizione Z viene associato l'alfabeto $S = \{1, \dots, N\}$ e sia $\Omega = S^{\mathbb{N}}$.

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

$$\varphi_Z(x) = \omega \quad \Longleftrightarrow \quad T^j(x) \in I_{\omega_j} \ \ \forall\, j \in \mathbb{N} \tag{9.1}$$

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

L'applicazione φ_Z risulta misurabile per definizione. Infatti se $C_k(\omega, n)$ è un cilindro di Ω si ha

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

che è intersezione di insiemi in $\mathcal{A}$.

Otteniamo quindi che la rappresentazione simbolica è un coniugio o un semi-coniugio del sistema dinamico, e l'iniettività dell'applicazione φ_Z dipende dalla scelta della partizione. Le partizioni buone sono chiamate "generanti" per il sistema dinamico $(\mathcal{X}, T)$.

Esempio 9.1. Sia $\mathcal{X} = [0,1]$ con la partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$ e l'azione della mappa T di Bernoulli dell'Esempio 6.1. L'alfabeto associato alla partizione Z è l'alfabeto binario $S = \{0,1\}$ e la rappresentazione simbolica ha una facile interpretazione. Sia infatti

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

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

Lo studio di questa rappresentazione simbolica diventa quindi molto più semplice usando l'espansione binaria dei punti x. Si ottiene facilmente che $\varphi_Z(X) = \{0,1\}^{\mathbb{N}}$ e l'iniettività di φ_Z. Inoltre φ_Z è continua tranne che nei punti diadici, che sono un insieme numerabile e quindi trascurabile rispetto alla distribuzione uniforme.

Si ottiene che la rappresentazione simbolica relativa alla partizione Z data è un coniugio topologico (a meno di un insieme di misura nulla e dalla dinamica trascurabile) tra la mappa di Bernoulli e la mappa di shift su $\{0,1\}^{\mathbb{N}}$.

♢

La rappresentazione simbolica φ_Z di un sistema deterministico $(\mathcal{X}, T)$ permette di interpretare i sistemi dinamici come *sorgenti di informazione* rispetto a una partizione.

Definizione 9.2 (**Sorgente di informazione**)**.** Dato un alfabeto finito $S = \{1, \dots, N\}$, una *sorgente di informazione* è determinata da una terna $(\Omega, \mathcal{C}, \mathbb{P})$, dove $\Omega \subseteq S^{\mathbb{N}}$ è lo spazio di tutte le possibili stringhe infinite prodotte dalla sorgente, $\mathcal{C}$ è la σ-algebra dei cilindri su Ω, e $\mathbb{P}$ è una misura di probabilità stazionaria su $(\Omega, \mathcal{C})$.

Sia $(\mathcal{X}, \mathcal{A}, T)$ un sistema deterministico con misura di probabilità invariante $\mathbb{P}_T$. Sia $Z = \{I_1, \dots, I_N\}$ una partizione finita e misurabile di $\mathcal{X}$, con associato l'alfabeto $S = \{1, \dots, N\}$. La rappresentazione φ_Z genera quindi una sorgente di informazione data dalla terna $(\varphi_Z(\mathcal{X}), \mathcal{C}, \mathbb{P}_Z)$, dove

$$\mathbb{P}_Z(C) := \mathbb{P}_T\left(\varphi_Z^{-1}(C)\right), \qquad \forall\, C \in \mathcal{C} \tag{9.2}$$

La misura $\mathbb{P}_Z$ è stazionaria su $(\varphi_Z(\mathcal{X}), \mathcal{C})$ e risulta anche invariante per l'azione dello shift τ su $\varphi_Z(\mathcal{X})$.

10. L'entropia di Shannon - 10/05/12

Seguiamo adesso l'approccio originale di Shannon alla teoria delle comunicazioni, il cui inizio si può far risalire all'articolo apparso nel 1948 sul Bell System Technology Journal.

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

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

Il significato del messaggio diventa quindi irrilevante per la trasmissione dello stesso, ma ad avere importanza sono le caratteristiche "strutturali" del messaggio, quindi ad esempio la sua lunghezza, il tipo di simboli che contiene, e così via. Un messaggio viene spesso scelto in un insieme finito di possibili messaggi, e questo permette un'interpretazione del contenuto di informazione del messaggio. Citando ancora Shannon

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

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

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

Un sistema di comunicazione è formato da 5 parti:

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

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

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

Proposizione 10.1. *Sia $(a_n)_n$ una successione di numeri reali tale che $a_{m+n} \leq a_m + a_n$ per ogni m, n interi. Allora*

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

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

$$a_n \leq a_1 + a_{n-1} \leq \cdots \leq na_1$$

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

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

da cui

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

e il teorema segue dall'arbitrarietà di ε. □

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

Definizione 10.2. Si definisce *capacità* C della sorgente di informazione il limite (quando esiste)

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

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

L'esistenza del limite nella definizione di capacità si può ottenere applicando alla successione $(\log M(n))_n$ la Proposizione 10.1. Basta infatti notare che basta supporre che $M(m+n) \leq M(m)M(n)$ per ogni m, n interi (infatti è ragionevole pensare che per una sorgente di informazione, parti di un possibile messaggio siano singoli possibili messaggi).

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

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

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

Introduciamo ora il concetto di *entropia*. Supponiamo di ricevere messaggi da una sorgente di informazione con le proprietà descritte sopra. Ci possiamo chiedere quale sia l'informazione prodotta dalla sorgente, oppure equivalentemente quale sia la nostra incertezza sul messaggio che riceveremo.

Una prima risposta parziale si ottiene usando il concetto di capacità. Se infatti supponiamo che la nostra sorgente produca $M(n) \sim 2^{nC}$ messaggi equivalenti di lunghezza n, possiamo dire, usando

il concetto di informazione contenuto in un messaggio introdotto da Shannon, che un messaggio lungo n ha un contenuto di informazione $\log M(n) \sim nC$. Quindi la sorgente produce C bits di informazione per unità di tempo.

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

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

(1) H sia continua nelle variabili;
(2) se tutte le probabilità sono uguali tra loro e quindi $p_i = \frac{1}{N}$ per ogni $i = 1, \dots, N$, allora H deve essere una funzione crescente in N;
(3) se la scelta del messaggio si spezza in due scelte successive, allora H deve essere la media pesata delle H_i, le funzioni legate alla seconda scelta, più la funzione legata alla prima scelta.

Esempio 10.1. Consideriamo il caso di una stringa binaria $s = (s_1 s_2)$ (sull'alfabeto $\{0, 1\}$) di lunghezza 2, prodotta da una sorgente che abbia le seguenti probabilità sulle stringhe lunghe 2, che costituiscono i messaggi:

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

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

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

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

♢

Teorema 10.3. *Sia $H(p_1, \dots, p_N)$ una funzione che soddisfa le ipotesi (1)-(3). Allora deve essere*

$$H(p_1, \dots, p_N) = -k \sum_{i=1}^{N} p_i \log p_i \tag{10.1}$$

con k costante positiva arbitraria.

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

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

Dati due numeri reali positivi s e t, per ogni intero n esiste un intero m tale che $s^m \leq t^n < s^{m+1}$. Allora

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

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

$$\left| \frac{m}{n} - \frac{\log t}{\log s} \right| \leq \frac{1}{n}$$

$$\left| \frac{m}{n} - \frac{h(t)}{h(s)} \right| \leq \frac{1}{n}$$

Quindi, mettendo insieme le due equazioni si ottiene

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

con $\varepsilon = \frac{2}{n}$, che diventa piccola a piacere con n abbastanza grande. Quindi $h(t) = k \log t$ con k costante positiva, per rispettare l'ipotesi (2).

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

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

Da questa uguaglianza si ricava

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

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

La funzione H con $k = 1$ è chiamata *entropia* di una sorgente di informazione che produce messaggi in cui ogni simbolo non è correlato al precedente.

Generalizzando il concetto di contenuto di informazione di un messaggio dato nel caso di messaggi equivalenti, possiamo dire che nel caso precedente, ogni lettera dell'alfabeto S ha un contenuto di informazione pari a $\log \frac{1}{p}$, dove p è la probabilità che la lettera sia prodotta dalla sorgente. Quindi maggiore è la probabilità che una lettera sia prodotta, minore è l'informazione che contiene, quindi minore è la sorpresa del destinatario del messaggio nel riceverla, e minore è l'incertezza sul messaggio che viene rimossa. Osserviamo che nel caso in cui le lettere siano equi-probabili si ricade nel concetto introdotto da Shannon. L'entropia H dell'equazione (10.1) si può quindi interpretare come l'informazione media di una lettera dell'alfabeto S. Di nuovo nel caso di lettere equi-probabili, l'entropia è uguale alla capacità della sorgente.

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

Definizione 10.4. Data una sorgente di informazione, si chiama *entropia di Shannon* della sorgente, il limite

$$h = \lim_{n\to\infty} \; -\frac{1}{n} \sum_{i=1}^{M(n)} p(m_i^n) \log p(m_i^n) \tag{10.2}$$

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

Notiamo che le funzioni

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

possono considerarsi le funzioni entropie di una sorgente che produca messaggi di lunghezza n con le probabilità $p(m_i^n)$ assegnate. Infatti tutte le H_n sono uguali alla funzione H dell'equazione (10.1), in cui i messaggi di lunghezza n sostituiscono i simboli. Se quindi H_n è l'informazione media contenuta in un messaggio lungo n, allora $\frac{1}{n}H_n$ è l'informazione media contenuta in uno dei simboli di un messaggio lungo n. *L'entropia di Shannon è quindi il contenuto medio di informazione di un simbolo prodotto da una sorgente di informazione.*

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

11. L'entropia nei sistemi dinamici - 10/05/12

Sia $(\mathcal{X}, \mathcal{A}, T)$ un sistema dinamico e sia $Z = \{I_1, \dots, I_N\}$ una partizione finita e misurabile di X. Indichiamo con $T^{-1}Z$ la partizione di X data dalle contro-immagini degli insiemi di Z, quindi $T^{-1}Z = \{T^{-1}I_1, \dots, T^{-1}I_N\}$. Così per ricorrenza poniamo $T^{-n}Z := T^{-1}(T^{-(n-1)}Z)$ per ogni $n \geq 1$, dove $T^0 Z = Z$.

Definizione 11.1. Date due partizioni P e Q, si definisce *unione* di P e Q, e si indica con $P \vee Q$, la partizione data da tutte le possibili intersezioni degli insiemi di P e Q, ossia

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

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

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

$$Z_n := \bigvee_{i=0}^{n-1} T^{-i} Z \tag{11.1}$$

Gli insiemi della partizione Z_n sono della forma

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

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

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

Definizione 11.2. Data una partizione finita Z di $\mathcal{X}$, indichiamo con $d(Z)$ la cardinalità di Z, ossia il numero di insiemi di Z. Si definisce *entropia topologica di* Z il numero

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

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

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

Si verifica che

$$H_{top}(Z_{n+m}) \leq H_{top}(Z_n) + H_{top}(Z_m)$$

per ogni $n, m \in \mathbb{N}$. Quindi per la Proposizione 10.1 il limite nella definizione di entropia topologica esiste.

Esempio 11.1 (**Mappa di Bernoulli, Esempio 6.1**)**.** Consideriamo la partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$. Si verifica che $d(Z_n) = 2^n$, quindi

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

La sorgente di informazione associata ha quindi capacità $C = 1$.

♢

Esempio 11.2 (**Rotazioni irrazionali**)**.** Consideriamo la rotazione $T_\alpha(x) = x + \alpha (\mathrm{mod}\, 1)$ su $\mathcal{X} = [0, 1]$ con $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Proposizione 11.3. *Data la partizione* $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$ *vale* $d(Z_n) = 2n$.

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

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

♢

Esempio 11.3 (**Mappa logistica**)**.** Studiare il comportamento di $h_{top}(T_\lambda, Z)$ con $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1]\right\}$ al variare di $\lambda \in [0, 4]$ comporta notevoli difficoltà, e usa tecniche che sono al di fuori dello scopo di questo corso.

Osserviamo però che per definizione di entropia topologica, si ha che due sistemi coniugati hanno la stessa entropia topologica. Quindi, usiamo il coniugio tra la mappa logistica e lo shift su due simboli (vedi esempio sotto) per determinare che $h_{top}(T_4, Z) = 1$.

♢

Esempio 11.4 (**Dinamica simbolica**)**.** Sia $S = \{1, \ldots, N\}$ l'alfabeto e M la matrice di transizione. Nel caso in cui $m_{ij} = 1$ per ogni coppia di simboli (i, j) (ossia consideriamo la mappa shift), si verifica facilmente che

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

considerando i cilindri $C(j, 0, 1)$ al variare di j in S come partizione Z finita e misurabile. Infatti basta contare tutte le stringhe possibili di lunghezza n per verificare che $d(Z_n) = N^n$. Un po' più complesso risulta il conto nel caso di una matrice di transizione con alcuni zeri (ossia per un subshift di tipo finito). In questo caso bisogna usare

Teorema 11.4 (Frobenius-Perron)**.** *Sia M matrice di transizione irriducibile e aperiodica. Allora esiste $\lambda_1 > 0$, tale che λ_1 è autovalore di M con autospazio di dimensione 1, e tutti gli altri autovalori verificano $|\lambda_i| < \lambda_1$.*

Proposizione 11.5. *Sia M matrice di transizione irriducibile e aperiodica, allora se λ_1 è l'autovalore massimale di M dato dal Teorema di Frobenius-Perron, si ha*

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

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

♢

La definizione di entropia topologica però considera le orbite di un sistema dinamico tutte con lo stesso "peso". Invece, ricordiamo che la scelta di una misura invariante $\mathbb{P}$ riduce lo studio del sistema dinamico allo studio delle proprietà statistiche delle orbite che la misura "vede" e che sono "pesate" in maniera diversa. Dobbiamo quindi introdurre un concetto di *entropia* per i sistemi dinamici che dipenda dalla misura invariante.

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

Definizione 11.6. Si definisce *entropia metrica della partizione Z* il valore

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

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

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

$$Inf_Z(x) := -\log(\mathbb{P}(Z(x))) \tag{11.2}$$

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

$$H_{\mathbb{P}}(Z) = \int_X Inf_Z(x)\, d\mathbb{P}(x)$$

Proposizione 11.7. *Siano P e Q partizioni finite e misurabili di X. Allora*

(i) $0 \le H_{\mathbb{P}}(P) \le \log(d(P))$*;*
(ii) $H_{\mathbb{P}}(P \vee Q) \le H_{\mathbb{P}}(P) + H_{\mathbb{P}}(Q)$*.*

Sia come sopra φ_Z la rappresentazione simbolica associata alla partizione $Z = \{I_1, \dots, I_N\}$ di $\mathcal{X}$ e a valori nello spazio $\Omega \subseteq S^{\mathbb{N}}$, dove $S = \{1, \dots, N\}$. Allora è immediato verificare che se $p_i := \mathbb{P}(I_i)$ per ogni $i = 1, \dots, N$, allora l'entropia metrica di Z, $H_{\mathbb{P}}(Z)$, è uguale alla funzione entropia $H(p_1, \dots, p_n)$ (vedi equazione (10.1)) della sorgente di informazione associata. È quindi evidente che il prossimo passo sarà lo studio della crescita dell'entropia delle partizioni iterate.

Definizione 11.8. Sia $\{Z_n\}_{n\geq 1}$ la successione di partizioni iterate generate da una partizione Z finita e misurabile di $\mathcal{X}$. L'*entropia metrica* $h_{\mathbb{P}}(T, Z)$ del sistema dinamico $(\mathcal{X}, T, \mathbb{P})$ *relativa a* Z e dipendente dalla misura di probabilità $\mathbb{P}$ T-invariante è data da

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

L'esistenza del limite è garantita dalla Proposizione 11.7,(ii) (da cui otteniamo $H_{\mathbb{P}}(Z_{n+m}) \leq H_{\mathbb{P}}(Z_n) + H_{\mathbb{P}}(Z_m)$) e dalla Proposizione 10.1.

Notiamo che questa definizione coincide con la Definizione 10.4, data per la sorgente di informazione generata dalla rappresentazione simbolica φ_Z.

È importante avere un concetto di entropia metrica di un sistema dinamico indipendente dalla scelta di una partizione.

Definizione 11.9. Sia $\mathbb{P}$ una misura di probabilità T-invariante sullo spazio $\mathcal{X}$. L'*entropia metrica* $h_{\mathbb{P}}(T)$ del sistema dinamico $(\mathcal{X}, T)$ dipendente da $\mathbb{P}$ è definita come

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

Per il calcolo effettivo dell'entropia metrica di un sistema serve avere però delle partizioni "buone".

Definizione 11.10. Diciamo che un partizione finita e misurabile Z è *generante* per un sistema dinamico $(\mathcal{X}, T, \mathbb{P})$ se $\vee_{i=0}^{\infty} T^{-i} Z$ genera la σ-algebra di Borel di X.

Possiamo allora enunciare

Teorema 11.11 (Kolmogorov-Sinai)**.** *Se Z è una partizione generante per il sistema dinamico $(\mathcal{X}, T, \mathbb{P})$, allora $h_{\mathbb{P}}(T) = h_{\mathbb{P}}(T, Z)$.*

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

Definizione 11.12. Due sistemi dinamici $(X_1, T_1, \mathbb{P}_1)$ e $(X_2, T_2, \mathbb{P}_2)$ si dicono *isomorfi* se esiste un coniugio misurabile $h : X_1 \to X_2$ che sia invertibile con inversa misurabile, e tale che $\mathbb{P}_2(B) = \mathbb{P}_1(h^{-1}(B))$ per ogni insieme $B \subset X_2$ misurabile.

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

Proposizione 11.13. *L'entropia metrica di un sistema dinamico è invariante per isomorfismo.*

Esempio 11.5 (**Rotazioni del cerchio**)**.** Ricordiamo che la misura di probabilità invariante di T_α è data dalla distribuzione uniforme $d\mathbb{P}(x) = dx$. Si ha $h_{\mathbb{P}}(T_\alpha) = 0$ per ogni $\alpha \in \mathbb{R}$.

♢

Esempio 11.6 (**Mappa di Bernoulli**)**.** Abbiamo dimostrato che la mappa di Bernoulli T è coniugata con lo shift τ su $\Omega = \{0, 1\}^{\mathbb{N}}$ usando la rappresentazione simbolica φ_Z, indotta dalla partizione $Z = \left\{[0, \frac{1}{2}), [\frac{1}{2}, 1)\right\}$. Data su $X = [0, 1]$ la misura di probabilità $\mathbb{P}$ con distribuzione uniforme, invariante per la mappa di Bernoulli, consideriamo la misura ν su Ω indotta da φ_Z tramite $\nu(C) = \mathbb{P}(\varphi_Z^{-1}(C))$ per ogni cilindro C di Ω. Notiamo che ν è τ-invariante. Quindi la mappa di Bernoulli è isomorfa a (Ω, τ, ν). Per il calcolo dell'entropia metrica possiamo quindi rifarci alla Proposizione 11.13 e usare l'esempio 11.8 sotto.

Analogamente si può usare il Teorema di Kolmogorov-Sinai e la definizione di entropia metrica relativa a una partizione.

♢

Esempio 11.7 (**Mappa logistica**)**.** Consideriamo la mappa T_4. Abbiamo dimostrato che la misura di probabilità invariante à data da $d\mathbb{P}(x) = \frac{1}{\pi\sqrt{x(1-x)}}dx$. Enunciamo il seguente risultato senza dimostrazione,

$$h_{\mathbb{P}}(T_4) = \int_0^1 \log(|(T_4)'(x)|)\ d\mathbb{P}(x) = 1$$

♢.

Esempio 11.8 (**Dinamica simbolica**)**.** Per calcolare l'entropia metrica dello shift e dei subshift di tipo finito su un alfabeto $S = \{1, \dots, N\}$ dobbiamo innanzitutto introdurre misure invarianti.

Sia $\{p_1, \dots, p_N\}$ una distribuzione di probabilità su S. Definiamo una misura di probabilità $\mathbb{P}$ su $\Omega := S^{\mathbb{N}}$ che sia τ-invariante, ponendo

$$\mathbb{P}(C_k(\omega, n)) := \prod_{i=0}^{n-1} p_{\omega_{k+i}} \tag{11.3}$$

per ogni cilindro. La misura invariante su un insieme misurabile qualsiasi si definisce usando la proprietà dei cilindri di generare la σ-algebra di Borel di Ω.

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

Proposizione 11.14. *Se M è matrice di transizione irriducibile e aperiodica, allora per ogni matrice stocastica Π associata a M esiste un unico vettore $p = (p_1, \dots, p_N)$, detto distribuzione stazionaria, tale che: $p_i \geq 0$ per ogni $i = 1, \dots, N$; $\sum_{i=1}^N p_i = 1$; $p\Pi = p$.*

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

$$\mathbb{P}_\Pi(C_k(\omega, n)) = \left(\prod_{i=0}^{n-2} \pi_{\omega_{k+i}\omega_{k+i+1}}\right) p_{\omega_k} \tag{11.4}$$

e poi estesa a tutta la σ-algebra dei cilindri. Si verifica che $\mathbb{P}_\Pi$ è τ-invariante.

Calcoliamo ora l'entropia di un subshift di tipo finito rispetto a una misura di Markov usando il Teorema di Kolmogorov-Sinai applicato alla partizione generante $Z = \{C_1, \dots, C_N\}$, dove $C_i := C_0(i, 1)$.

Proposizione 11.15. *Sia $\mathbb{P}_\Pi$ una misura di Markov associata a una matrice di transizione M irriducibile e aperiodica. Allora*

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

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

Notiamo che nel caso particolare dello shift, l'entropia metrica delle misure di probabilità invarianti $\mathbb{P}$ del tipo dato nell'equazione (11.3), si ottiene dalla proposizione ponendo $\pi_{ji} = p_i$ per ogni j, per cui

$$h_{\mathbb{P}}(\tau) = -\sum_{i=1}^N p_i \log(p_i)$$

♢

Esempio 11.9 (**Rappresentazioni simboliche**)**.** Sia $(\mathcal{X}, T, \mathbb{P}_T)$ un sistema dinamico e sia data una partizione finita e misurabile $Z = \{I_1, \dots, I_N\}$, la rappresentazione simbolica φ_Z ha come immagine un sottospazio di $\Omega = S^{\mathbb{N}}$, dove S è l'alfabeto associato a Z. Tale sottospazio è invariante per l'azione dello shift, e quindi definisce un sistema dinamico. Solitamente questo sistema dinamico non è un subshift di tipo finito, ma è comunque possibile definire su $\varphi_Z(\mathcal{X})$ una misura $\mathbb{P}_Z$ invariante indotta da φ_Z. Dati i cilindri della forma

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

si definisce $\mathbb{P}_Z$ sui cilindri e poi si estende a tutta la σ-algebra. Si pone

$$\mathbb{P}_Z(C_k(\omega, n)) = \mathbb{P}_T\left(\varphi_Z^{-1}(C_k(\omega, n))\right) = \mathbb{P}_T\left(\bigcap_{i=k}^{k+n-1} T^{-i} I_{\omega_i}\right)$$

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

♢

Esempio 11.10 (**Sistemi casuali**)**.** Dato un sistema dinamico $(\mathcal{X}, T, \mathbb{P})$, le orbite $(T^n(x))_{n\in\mathbb{N}}$ del sistema associate ai punti $x \in \mathcal{X}$, si possono considerare come realizzazioni di un sistema casuale a valori in $\mathcal{X}$. Si può quindi studiare l'entropia metrica di un processo stocastico in generale. Facciamo adesso l'esempio di un caso particolare. Sia $(Y_n)_n$ un sistema casuale definito su $(\Omega, \mathcal{C}, \nu)$ e a valori nello spazio $(X, \mathcal{A}, \mathbb{P})$. Supponiamo in particolare che $\mathcal{X} = [0, 1]$, che $\mathbb{P}$ sia la misura con distribuzione uniforme, e che il sistema casuale sia costituito da variabili aleatorie indipendenti. Tale sistema viene chiamato *rumore bianco*. Dimostriamo che l'entropia metrica $h_{\mathbb{P}}$ del rumore bianco è infinita. Consideriamo infatti la partizione

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

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

Osserviamo infine che per sistemi casuali a valori in uno spazio $\mathcal{X}$ finito, vale invece $h_{\mathbb{P}} \leq \log(d(X))$.

♢

12. Contenuto di Informazione Algoritmico (AIC) - 17/05/12

Vogliamo adesso introdurre una nozione di contenuto di informazione di una stringa infinita, prodotta da una sorgente di informazione, che non dipenda dalla misura di probabilità stazionaria per la sorgente. Quest'approccio è fondamentale nello studio di serie temporali di origine sperimentale (serie di dati), in cui la misura di probabilità è ignota.

Volendo analizzare una singola serie di dati, dobbiamo innanzitutto specificare cosa intendiamo per *casualità* per una singola stringa. Supponiamo di considerare le possibili realizzazioni di un semplice esperimento: il lancio di una moneta in cui lo spazio degli stati $\mathcal{X} = \{T, C\}$ abbia la distribuzione di probabilità $p_T = p_C = \frac{1}{2}$. Consideriamo allora una serie temporale ottenuta con N lanci della moneta. La teoria della probabilità ci dice che ogni possibile successione di N simboli

dall'alfabeto $S = \{T, C\}$ ha esattamente probabilità $\frac{1}{2^N}$. Se quindi $N = 10$, avremo per esempio

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

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

Se guardiamo però attentamente le tre stringhe dell'esempio, appare evidente che qualche differenza tra le tre possibili realizzazioni del nostro esperimento c'è. Anzi, questa differenza è talmente evidente, che sembra sorprendente non riuscire a coglierla utilizzando il contenuto di informazione definito da Shannon. La prima stringa è infatti semplicemente "la ripetizione di 10 simboli C", la seconda è "la ripetizione 5 volte della coppia CT", mentre la terza stringa è "CCTTCTTCTC"! La differenza sta dunque nella facilità di "descrizione" della stringa. Vediamo qualche altro esempio. Consideriamo due numeri naturali, 10000 e 19654. I due numeri sono dello stesso ordine di grandezza, eppure il primo si può esprimere in maniera più compatta in varie forme. Ad esempio possiamo dire: "diecimila" e "diciannovemila, seicento cinquantaquattro"; oppure scriverli come: 10^5 e 1.9654×10^5. Notiamo che se avessimo scelto numeri di un ordine di grandezza molto maggiore, per esempio dell'ordine di 10^{100}, allora la differenza tra le due descrizioni sarebbe stata molto più accentuata. Così accadrebbe lo stesso per i risultati del nostro esperimento di lancio di moneta.

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

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

$$\pi = 3.14159265358979323846264338328\ldots$$

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

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

Quest'uguaglianza ci fornisce uno degli algoritmi per il calcolo delle cifre decimali di π. Altri algoritmi possono essere ottenuti usando le altre mirabili espressioni in serie di π. La risposta alla domanda quanto sia lunga la descrizione della stringa delle cifre decimali di π, è quindi che tale descrizione è lunga quanto la descrizione dell'algoritmo che voglio usare, più la descrizione del numero naturale che esprime quante cifre decimali voglio conoscere. La prima parte è una costante,

e la dipendenza dalla lunghezza N della stringa numerica che voglio ottenere è solo nella seconda parte della risposta. Dimostreremo che tale dipendenza è dell'ordine $\log N$. Sorprendentemente poco, se si considera che per la maggior parte dei numeri reali si ha una dipendenza dell'ordine N.

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

12.1. **Le funzioni computabili.** Dobbiamo allora innanzitutto formalizzare il concetto di *descrizione*. Introduciamo allo scopo alcuni concetti base della teoria della computabilità.

Il primo passo fondamentale sono i lavori di Alan Turing in cui descrive una *macchina* che possa simulare ogni calcolo possibile. Alcuni brani liberamente tradotti dal suo articolo descrivono in maniera molto intuitiva la sua idea.

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

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

Definizione 12.1 (**Macchina di Turing**)**.** Una *macchina di Turing* consiste di un *programma finito* che può manipolare una lista di cellette, il *nastro*, con un *pointer*. Il programma ha un numero finito di stati Q e ogni celletta contiene uno dei simboli dell'insieme $S = \{\lambda, 0, 1\}$. Il tempo è discreto, e ad ogni istante di tempo il pointer è su una particolare celletta, che è quella su cui operare.

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

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

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

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

La macchina di Turing permette di definire in maniera formale le funzioni computabili sugli interi, usando la rappresentazione degli interi tramite stringhe binarie finite.

Definizione 12.2 (**Funzione computabile**)**.** Una funzione *ricorsiva parziale* o *computabile* è un'operazione, associata a una macchina di Turing, che trasforma una n-upla di interi ($n \geq 1$), che rappresentano l'input, in un intero, l'output della macchina. Se la macchina di Turing associata a una funzione si ferma su ogni input, la funzione si dice *ricorsiva totale*.

Esempio 12.1. Date due stringhe $s, t \in \{0,1\}^*$, esempi di funzioni ricorsive sono:

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

Le funzioni f e g sono ricorsive totali, mentre la funzione h è ricorsiva parziale, non essendo definita per esempio sulla stringa binaria (1111).

♢

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

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

Mostriamo ora che è possibile enumerare le macchine di Turing, fornendo una bigezione dell'insieme delle macchine con i numeri naturali.

Dalla definizione, abbiamo visto che ciò che identifica in maniera univoca una macchina di Turing è l'insieme delle regole che il programma esegue. L'insieme $Q \cup S \cup Op$ contiene $(d(Q)+5)$ elementi, quindi ogni suo elemento può essere descritto con $s \leq \lceil 2\log(d(Q)+5) \rceil$ bits, ad esempio tramite un codice binario ϵ. Data una macchina di Turing T, siano k le sue regole, e siano date dalle quadruple (p_i, t_i, s_i, q_i), per $i = 1, \dots, k$. Un codice $E(T)$ per le macchine di Turing è allora definito da

$$E(T) := \epsilon(s)\ \epsilon(k)\ \epsilon(p_1)\epsilon(t_1)\epsilon(s_1)\epsilon(q_1) \dots \epsilon(p_k)\epsilon(t_k)\epsilon(s_k)\epsilon(q_k) \tag{12.1}$$

Ordinando in maniera lessicografica le stringhe binarie $E(T)$ è possibile stabilire un ordinamento delle rispettive macchine di Turing. Il numero d'ordine della stringa $E(T)$ nell'ordinamento delle stringhe codice che rappresentano le macchine di Turing si chiama il *numero di Gödel* $n(T)$ di una macchina di Turing T.

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

La numerazione delle macchine di Turing permette la costruzione di macchine di Turing *universali*, nel senso che possono simulare il comportamento di una qualsiasi altra macchina di Turing.

Esempio 12.2 (**Macchina di Turing universale**)**.** Sia U una macchina di Turing che accetta, come input, stringhe della forma $(1^i 0 p)$, dove $p \in \{0,1\}^*$. La macchina U funzioni nel seguente modo: sul lato sinistro del nastro rispetto all'input, genera in ordine lessicografico, secondo la lunghezza, tutte le stringhe di $\{0,1\}^*$; per ogni stringa generata decide se è della forma $E(T)$ per qualche T; ogni volta che la stringa generata descrive una macchina di Turing, sostituisce uno dei simboli 1 dell'input con il simbolo λ e legge il simbolo a destra; quando legge il simbolo 0, esegue sulla stringa p le regole specificate dalla stringa binaria $E(T)$ che ha sul lato sinistro del nastro. Quindi, possiamo scrivere $U(1^i 0 p) = T_i(p)$.

♢

Resta aperta una questione importante legata alla possibilità di una macchina di Turing di non fermarsi su alcuni input. È quindi naturale chiedersi quali macchine di Turing si fermano e su quali input. Questo problema è noto con il nome di *halting problem*. La soluzione è contenuta nel seguente risultato.

Teorema 12.3 (Turing)**.** *Sia $(\varphi_i)_i$ la numerazione delle funzioni ricorsive parziali. Non esiste una funzione ricorsiva totale g definita su $(\{0,1\}^*)^2$ tale che per ogni coppia (s,t) valga $g(s,t)=1$ se $\varphi_s(t)$ è definita (ossia T_s si ferma su t), e $g(s,t)=0$ altrimenti.*

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

12.2. **AIC.** Per formalizzare l'approccio intuitivo al concetto di casualità di una stringa discusso sopra risulta fondamentale la macchina di Turing come strumento per la computazione e la formalizzazione del concetto di *descrizione*.

Definizione 12.4 (**Contenuto di Informazione Algoritmico**)**.** Sia U una macchina di Turing universale. Date due stringhe $s,t\in\{0,1\}^*$, si chiama *Algorithmic Information Content* (o *complessità di Kolmogorov*) di s relativa a t il numero

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

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

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

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

$$AIC(s|t)\leq AIC_T(s|t)+c(T,U) \tag{12.2}$$

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

Proposizione 12.5. *Date due stringhe $s,t\in\{0,1\}^*$ vale*

(i) $AIC(s)\leq |s|+cost$
(ii) $AIC(s|t)\leq AIC(s)+cost$
(iii) $AIC(ss)\leq AIC(s)+cost$
(iv) $AIC(n)\leq \log n+cost$ *per ogni* $n\in\mathbb{N}$

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

(*ii*) Costruiamo una macchina T tale che $T(t,p)=s$ se e solo se $U(p)=s$. La tesi segue da $AIC_T(s|t)=AIC(s)$ e dalla disuguaglianza (12.2).

(*iii*) Costruiamo una macchina T che raddoppi l'output di U. Ossia valga $T(p)=U(p)U(p)$ per ogni input p. Sia m il numero di Gödel della macchina T, allora $U(1^m0p)=U(p)U(p)$ per ogni input p. Se in particolare p è il programma di s si ha $AIC(ss)\leq 1+m+AIC(s)$.

(*iv*) Costruiamo una macchina T tale che $T(\sigma)=n$, se σ è la stringa binaria associata ad n. Sia poi T' una macchina tale che $T(p)=\sigma$ se $U(p)=\sigma$ e $AIC(\sigma)=|p|$. Sia m il numero di

Gödel della macchina $T \circ T'$, allora $U(1^m 0p) = T(T'(p)) = T(\sigma) = n$. Quindi $AIC(n) \leq m + |p| = m + AIC(\sigma) \leq m + \log n$ per (i). □

Viene naturale pensare che il AIC sia una funzione sub-additiva, ossia che valga $AIC(s,t) \leq AIC(s) + AIC(t) + cost$ per ogni coppia di stringhe s, t. Se infatti p e q sono programmi per s e t, rispettivamente, il programma per (s,t) sarà la concatenazione pq. Tuttavia, la macchina U deve sapere interpretare la stringa pq come due programmi diversi. Si trova allora

$$AIC(s,t) \leq AIC(s) + AIC(t) + 2\log(\min\{AIC(s), AIC(t)\}) + cost \tag{12.3}$$

Lo stesso accade per $AIC(st)$, diversamente da $AIC(ss)$.

Applichiamo per esempio il AIC alla stringa data dall'espansione decimale di π. Abbiamo osservato che l'espansione decimale di π, nonostante sia intuitivamente casuale, può essere riprodotta usando algoritmi basati sulle uguaglianze tra π e alcune serie numeriche convergenti. Ne segue che, se π_1^n indica la stringa in $\{0,1,2,\ldots,9\}^n$ data dai primi n elementi nell'espansione decimale di π, vale $AIC(\pi_1^n|n) \leq cost$ e $AIC(\pi_1^n) = O(\log n)$. L'espansione decimale di π è quindi una stringa la cui rappresentazione, attraverso una macchina di Turing, è molto più corta della stringa stessa.

Definizione 12.6. Data una costante $c \in \mathbb{N}$, si dice che una stringa $s \in \{0,1\}^*$ è *c-random* se $AIC(s) \geq |s| - c$.

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

Teorema 12.7 (**Incompressibility Theorem**). *Data una costante $c \in \mathbb{N}$, per ogni stringa t fissata, ogni insieme A con $d(A) = m < \infty$ è tale che la disuguaglianza*

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

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

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

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

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

Corollario 12.8. *Data una costante $c \in \mathbb{N}$, almeno $2^n - 2^{n-c} + 1$ stringhe di $\{0,1\}^n$ sono c-random.*

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

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

Concludiamo con un risultato di fondamentale importanza che implica che il AIC non può essere calcolato tramite un algoritmo su un insieme infinito di stringhe.

Teorema 12.9 (**Teorema di Non-Computabilità**). *La funzione $AIC(n)$ non è ricorsiva parziale. Inoltre, nessuna funzione ricorsiva parziale, definita su un insieme infinito, può coincidere con AIC su tutto il suo dominio.*

Si può provare a calcolare il AIC per qualche stringa particolare. Ma per determinare numericamente se una generica stringa sia o non sia c-random bisogna cercare di costruire algoritmi che approssimino il AIC.

13. La complessità nei sistemi dinamici e gli algoritmi di compressione - 24/05/12

Usando la nozione di AIC possiamo introdurre una nozione di complessità per le orbite di un sistema dinamico. Iniziamo con il caso della dinamica simbolica.

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

Definizione 13.1. Data una stringa $\omega \in \Omega$, chiamiamo *complessità* di ω il limite superiore

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

La complessità di una stringa infinita si può quindi interpretare come il contenuto di informazione medio di ogni simbolo della stringa, indipendentemente dal contesto. Nel teorema seguente, vedremo che coincide con l'idea di contenuto di informazione introdotto dall'entropia di Shannon e metrica.

Sia (Ω, τ) un sistema dinamico simbolico, dove τ è lo shift. Per una misura di probabilità $\mathbb{P}$ su Ω, che sia τ-invariante, indichiamo con $h_{\mathbb{P}}(\tau)$ la sua entropia metrica.

Teorema 13.2 (Brudno)**.** *Se $\mathbb{P}$ è una misura di probabilità su Ω, τ-invariante ed ergodica, allora $K(\omega) = h_{\mathbb{P}}(\tau)$ per $\mathbb{P}$-q.o. $\omega \in \Omega$.*

Affrontiamo adesso il problema per un sistema dinamico $(\mathcal{X}, T, \mathbb{P})$ generale. Innanzitutto vediamo come definire la complessità delle orbite. Usiamo la rappresentazione simbolica di un sistema dinamico

Definizione 13.3. Dato un sistema dinamico $(\mathcal{X}, T)$, sia Z una partizione finita e misurabile di $\mathcal{X}$, e sia φ_Z la rappresentazione simbolica associata. Poniamo $AIC(x, n, Z) := AIC(\varphi_Z(x)^n)$. La *complessità* $K(x, T, Z)$ dell'orbita un punto $x \in X$, relativa alla partizione Z, è definita tramite

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

Teorema 13.4. *Sia $(\mathcal{X}, T)$ sistema dinamico e $\mathbb{P}$ misura di probabilità T-invariante ed ergodica. Data una partizione Z finita e misurabile, per $\mathbb{P}$-quasi ogni $x \in \mathcal{X}$ vale $K(x, T, Z) = h_{\mathbb{P}}(T, Z)$.*

Per proseguire l'analogia con la definizione dell'entropia di un sistema dinamico, bisognerebbe considerare l'estremo superiore al variare di tutte le partizioni finite e misurabili. Mostriamo però che per la complessità $K(x, T, Z)$ non si può considerare l'estremo superiore al variare di Z.

Proposizione 13.5. *Sia $x \in \mathcal{X}$ un punto non periodico per T, allora per ogni $N \in \mathbb{N}$ esiste una partizione Z finita e misurabile, tale che $K(x, T, Z) = \log N$.*

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

$$\int_{\Omega} K(\omega) \, d\mathbb{P} = h_{\mathbb{P}}(\tau)$$

Considerando una partizione Z per un sistema dinamico $(\mathcal{X}, T, \mathbb{P})$, con $\mathbb{P}$ misura di probabilità T-invariante, si ottiene, ragionando come prima,

$$\int_{X} K(x, T, Z) \, d\mathbb{P} = h_{\mathbb{P}}(T, Z)$$

Nel caso dei sistemi dinamici ergodici la complessità di un'orbita ci fornisce quindi un metodo algoritmico semplice per calcolare l'entropia metrica di un sistema dinamico, ad esempio usando una partizione generante. Ci sono alcune osservazioni da fare.

La prima è che il calcolo dell'entropia metrica per un sistema ergodico $(\mathcal{X}, T, \mathbb{P})$ rispetto a una partizione Z è possibile farlo anche attraverso altri metodi di "natura ergodica", ossia solo basati sulla teoria ergodica dei sistemi dinamici. Ricordiamo il metodo principale, contenuto nel Teorema di Shannon-McMillan-Breiman, che consiste nell'uguaglianza

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

che vale per $\mathbb{P}$-q.o. $x \in \mathcal{X}$, dove $Z_n(x)$ indica l'insieme della partizione iterata Z_n che contiene il punto x.

La seconda osservazione è che l'utilizzo dell'AIC per il calcolo della complessità di un'orbita, ci permette di distinguere tra di loro anche sistemi dinamici con entropia metrica nulla. Infatti in questo caso si trova

$$AIC(x, n, Z) = o(n)$$

ma possiamo studiare le velocità di crescita con n della funzione $AIC(x, n, Z)$, e ottenere diversi comportamenti.

Infine ricordiamo che, in base al Teorema di Non-Computabilità per l'AIC, non esiste un algoritmo in grado di misurare la complessità delle orbite di ogni sistema dinamico. Per il calcolo della complessità dobbiamo quindi usare degli algoritmi che approssimino l'AIC. Esempi sono gli algoritmi di compressione.

13.1. **Algoritmi di compressione.** Dato un alfabeto finito S, abbiamo studiato il concetto di contenuto di informazione per stringhe di S^*. In maniera informale si può definire come una funzione $I : S^* \to \mathbb{N}$ che rappresenta la lunghezza del più piccolo messaggio da cui è possibile ricostruire la stringa iniziale. Dato un algoritmo $\mathcal{A}l$ definito su S^* che fornisce una codifica $\mathcal{A}l(s) \in \{0, 1\}^*$ di una stringa $s \in S^*$, possiamo definire la funzione contenuto di informazione associata a $\mathcal{A}l$ tramite $I_{\mathcal{A}l}(s) := |\mathcal{A}l(s)|$. Quindi possiamo definire la complessità di una stringa infinita $\omega \in S^{\mathbb{N}}$ associata a $\mathcal{A}l$ tramite

$$K_{\mathcal{A}l}(\omega) := \limsup_{n\to\infty} \frac{I_{\mathcal{A}l}(\omega^n)}{n}$$

Si tratta a questo punto di dimostrare che $K_{\mathcal{A}l}$ sia una buona nozione di complessità, nel senso che verifichi il Teorema 13.4. Ci si potrebbe poi chiedere se la funzione $I_{\mathcal{A}l}(\omega^n)$ abbia lo stesso comportamento di $AIC(\omega^n)$ nel caso di stringhe a complessità nulla.

Il primo problema è stato molto studiato ed esistono diversi algoritmi di compressione che lo risolvono. Il secondo è invece più delicato. A titolo di esempio mostriamo il funzionamento dell'algoritmo $LZ77$ introdotto da Lempel e Ziv, che funziona bene per sistemi che hanno AIC crescente con legge polinomiale n^α, con $\alpha < 1$.

L'essenza del funzionamento di $LZ77$ è l'utilizzo di puntatori che sostituiscono sotto-stringhe già apparse nella stringa da codificare. Questi puntatori si riferiscono alla posizione in cui la sotto-stringa iniziava nella parte già codificata. Una volta che la più lunga sotto-stringa già nota viene identificata, si forma una nuova parola aggiungendo un nuovo simbolo alla sotto-stringa.

Consideriamo un alfabeto $S = \{a_1, \ldots, a_r\}$ e una stringa $s \in S^*$ da codificare. Supponiamo che l'algoritmo abbia già codificato i primi p simboli $(s_1 \ldots s_p)$, e che abbia registrato h parole $\{e_1, \ldots, e_h\}$. Inizia quindi la codifica dal $(p+1)$-esimo simbolo e la prossima nuova parola sarà la $(h+1)$-esima. A questo punto l'algoritmo cerca, dal $(p+1)$-esimo simbolo, la più lunga parola che si può ottenere aggiungendo un solo simbolo $\tilde{a} \in S$ a una sotto-stringa ρ contenuta in $(s_1 \ldots s_p)$. La nuova parola avrà quindi un prefisso ρ e un simbolo finale $\tilde{a}$, ossia $e_{h+1} = \rho\ \tilde{a}$. Una volta che e_{h+1} è stata trovata, l'algoritmo la codifica tramite la tripla $(S_{h+1}, l_{h+1}, \tilde{a})$, dove S_{h+1} è la posizione in $(s_1 \ldots s_p)$ in cui comincia la sotto-stringa ρ, l_{h+1} è la lunghezza di e_{h+1}, e $\tilde{a}$ è l'ultimo carattere di e_{h+1}.

Ad esempio, vediamo come LZ77 codifica la stringa

$$s = (aababbbbaababbabbbaaababbbbb)$$

L'alfabeto $S = \{a, b\}$ e l'output è la codifica binaria delle triplette nella seconda colonna. La prima colonna indica l'ordine della nuova parola formata e la terza colonna è la parola stessa, che non è contenuta nell'output.

1	$(1, 1, a)$	$[a]$
2	$(1, 2, b)$	$[ab]$
3	$(2, 3, b)$	$[abb]$
4	$(5, 3, a)$	$[bba]$
5	$(2, 5, a)$	$[ababba]$
6	$(5, 4, a)$	$[bbba]$
7	$(1, 9, b)$	$[aababbbbb]$

Dipartimento di Matematica Applicata - Università di Pisa
E-mail address: bonanno@mail.dm.unipi.it