



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE, NATURALI  
CORSO DI LAUREA TRIENNALE IN MATEMATICA  
A. A. 2009/10

TESI DI LAUREA

# Il teorema ergodico

CANDIDATO  
Mauro Timini

RELATORE  
Francesco Caravenna



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Il teorema ergodico</b>	<b>3</b>
1.1 Richiami di probabilità . . . . .	3
1.2 Impacchettamenti e ricoprimenti . . . . .	8
1.3 Trasformazioni stazionarie ed ergodiche . . . . .	10
1.4 Il teorema ergodico . . . . .	16
1.4.1 Enunciato del teorema ergodico . . . . .	16
1.4.2 Il teorema ergodico : caso binario . . . . .	17
1.4.3 Il teorema ergodico: caso generale . . . . .	20
<b>2 Applicazioni ed esempi</b>	<b>31</b>
2.1 Strumenti tecnici . . . . .	31
2.2 Processi stazionari . . . . .	35
2.3 Catene di Markov . . . . .	36



# Introduzione

In questa tesi dimostriamo un risultato importante relativo al calcolo della probabilità: il teorema ergodico. La teoria ergodica ha le sue origini nella meccanica statistica: più precisamente, venne introdotta da Boltzmann (1844-1906) verso la fine dell'Ottocento, in riferimento ai sistemi meccanici complessi ai quali era attribuita la proprietà di assumere, nella loro evoluzione spontanea, tutti gli stati dinamici microscopici compatibili con il loro stato macroscopico. In base a tale teoria, la media “temporale” di ogni osservabile del sistema doveva coincidere con la media “spaziale” di tale osservabile rispetto a un’opportuna distribuzione di probabilità sullo spazio delle configurazioni del sistema. Una formulazione più tecnica di quest’idea venne fornita poco più tardi da Josiah Willard Gibbs (1839-1903) ed, in seguito, tale teoria venne sviluppata da un punto di vista matematico, considerandone gli aspetti probabilistici.

La nostra trattazione riguarda proprio quest’ultimo approccio. Il risultato fondamentale di questa teoria è costituito dal teorema ergodico, dimostrato da George David Birkhoff (1884-1944). Tale teorema afferma che, data una trasformazione  $T$  definita su uno spazio di probabilità ed a valori nel medesimo, se tale applicazione soddisfa determinate proprietà, allora la media temporale di ogni osservabile calcolata lungo la traiettoria  $\{x, Tx, T^2x, \dots\}$  (dove  $x$  è un punto dello spazio) coincide, facendo tendere all’infinito il numero di iterazioni, con l’integrale dell’osservabile rispetto alla misura di probabilità fissata sullo spazio. In particolare, questo teorema fornisce una generalizzazione della legge dei grandi numeri per un’ampia classe di processi stocastici (ad esempio, per le catene di Markov), al di là del caso classico di variabili indipendenti ed identicamente distribuite.

La dimostrazione del teorema ergodico che forniremo non è quella classica, ma si basa su alcuni argomenti di tipo combinatorio e discreto, sviluppati in [1, Sez.I.3]. La dimostrazione è strutturata in due parti: dapprima consideriamo il caso di processi discreti e poi estendiamo la trattazione a processi stocastici a valori in uno spazio generico.

La tesi è organizzata nel modo seguente:

- Nel Capitolo 1 richiamiamo alcuni concetti di base, sviluppiamo alcuni strumenti relativi ai ricoprimenti aleatori ed, infine, forniamo la dimostrazione del teorema ergodico.
- Nel Capitolo 2 presentiamo alcuni modi operativi di dimostrare l’ergodicità (in particolare definendo le proprietà di mescolanza e di mescolanza debole) ed applichiamo tali metodi a due casi importanti: i processi i.i.d. e la Catene di Markov.



# Capitolo 1

## Il teorema ergodico

### 1.1 Richiami di probabilità

In questo paragrafo richiameremo alcuni concetti di base della teoria del calcolo delle probabilità che ci saranno utili in seguito. Sia  $X$  un insieme. Definiamo  $\sigma$ -algebra su  $X$  un sottoinsieme  $\Sigma$  di  $\mathcal{P}(X)$  tale che :

- i)  $X \in \Sigma$ ,
- ii)  $A \in \Sigma \Rightarrow X \setminus A \in \Sigma, \forall A \in \Sigma$ ,
- iii) Se  $\{A_n\}_{n \in \mathbb{N}}$  è una successione di elementi di  $\Sigma$ , allora  $\bigcup_{n \in \mathbb{N}} A_n \in \Sigma$ .

Si può verificare facilmente che, comunque si scelga  $X$ , esistono sempre almeno due  $\sigma$ -algebre su  $X$  :  $\Sigma_1 = \{\emptyset, X\}$  ( $\sigma$ -algebra banale) e  $\Sigma_2 = \mathcal{P}(X)$  ( $\sigma$ -algebra discreta). Dato un insieme  $X$ , pertanto, è sempre possibile associarvi una  $\sigma$ -algebra  $\Sigma$  : si definisce *spazio misurabile* la coppia  $(X, \Sigma)$ . Sia, dunque,  $(X, \Sigma)$  uno spazio misurabile. Consideriamo ora un'applicazione

$$\mu : \Sigma \rightarrow [0, +\infty]$$

tale che:

- i)  $\mu(\emptyset) = 0$ .
- ii) Se  $\{A_n\}_{n \in \mathbb{N}}$  è una successione di elementi a due a due disgiunti di  $\Sigma$  (cioè tali che  $A_n \cap A_m = \emptyset, \forall m, n \in \mathbb{N}$  tali che  $m \neq n$ ), allora

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Tale funzione viene detta *misura su  $(X, \Sigma)$* . Uno *spazio di misura* è una terna  $(X, \Sigma, \mu)$ , dove  $X$  è un insieme,  $\Sigma$  è una  $\sigma$ -algebra su  $X$  e  $\mu$  è una misura su  $\Sigma$ . Se, inoltre, supponiamo valga l'ipotesi aggiuntiva  $\mu(X) = 1$ , allora si dice che  $\mu$  è una *probabilità su  $\Sigma$*  ed  $(X, \Sigma, \mu)$  è detto *spazio di probabilità*. Ora consideriamo due proprietà fondamentali della misura che risulteranno utili in seguito.

**Lemma 1** (Continuità dal basso). *Sia  $(X, \Sigma, \mu)$  uno spazio di misura. Data una successione  $\{A_n\}_{n \in \mathbb{N}}$  di elementi di  $\Sigma$  tale che  $A_n \subseteq A_{n+1}$ ,  $\forall n \geq 0$ , si ha che:*

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$$

*Dimostrazione.* Definiamo la successione  $\{B_m\}_{m \in \mathbb{N}}$  in questo modo:

$$B_0 := A_0 \text{ e } B_m := A_m \setminus A_{m-1}, \forall m \geq 1.$$

Si può osservare facilmente che, dato  $n \in \mathbb{N}$ ,  $A_n = \bigcup_{m=0}^n B_m$  e che, quindi,  $\bigcup_{n=0}^{\infty} A_n = \bigcup_{n=0}^{\infty} \bigcup_{m=0}^n B_m = \bigcup_{m=0}^{\infty} B_m$ . Pertanto, in virtù delle proprietà della misura  $\mu$  si ha che

$$\begin{aligned} \mu\left(\bigcup_{n=0}^{\infty} A_n\right) &= \mu\left(\bigcup_{m=0}^{\infty} B_m\right) = \sum_{m=0}^{\infty} \mu(B_m) \\ &= \lim_{n \rightarrow \infty} \sum_{m=0}^n \mu(B_m) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{m=0}^n B_m\right) = \lim_{n \rightarrow \infty} \mu(A_n). \end{aligned}$$

□

**Lemma 2** (Continuità dall'alto). *Sia  $(X, \Sigma, \mu)$  uno spazio di misura. Data una successione  $\{A_n\}_{n \in \mathbb{N}}$  di elementi di  $\Sigma$  tale che  $A_n \supseteq A_{n+1}$ ,  $\forall n \geq 0$ , si ha che:*

$$\mu\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$$

*Dimostrazione.* Innanzitutto osserviamo che, dato  $n \in \mathbb{N}$ , se  $A_n \supseteq A_{n+1}$ , allora  $A_n^c \subseteq A_{n+1}^c$ . Utilizzando la proprietà di continuità dal basso, quindi, possiamo concludere che

$$\begin{aligned} \mu(X) - \mu\left(\bigcap_{n \in \mathbb{N}} A_n\right) &= \mu\left(X \setminus \bigcap_{n \in \mathbb{N}} A_n\right) = \mu\left(\left(\bigcap_{n \in \mathbb{N}} A_n\right)^c\right) = \mu\left(\bigcup_{n \in \mathbb{N}} A_n^c\right) = \lim_{n \rightarrow \infty} \mu(A_n^c) \\ &= \lim_{n \rightarrow \infty} \mu(X \setminus A_n) = \lim_{n \rightarrow \infty} (\mu(X) - \mu(A_n)) = \mu(X) - \lim_{n \rightarrow \infty} \mu(A_n). \end{aligned}$$

Dunque

$$\mu\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n).$$

□

Un risultato fondamentale che si può derivare dalla proprietà di continuità dal basso di una misura e che ci sarà di estrema utilità in seguito è dato dal *Lemma di Borel-Cantelli*, che ora enunceremo.

**Lemma 3** (Borel-Cantelli). *Sia  $(X, \Sigma, P)$  uno spazio di probabilità. Se  $\{C_n\}_{n \in \mathbb{N}}$  è una successione di elementi di  $\Sigma$  tale che  $\sum P(C_n) < \infty$ , allora*

$$P\left(\bigcap_{n \geq 1} \bigcup_{m \geq n} C_m\right) = 0,$$

ossia per quasi tutti gli  $x \in X$  esiste un naturale  $N = N(x)$  tale che  $x \notin C_n, \forall n \geq N$ .

*Dimostrazione.* Sia  $G_n := \bigcup_{m \geq n} C_m$ . Osserviamo che  $G_n \supset G_{n+1}$ ,  $\forall n \in \mathbb{N}$ . Per continuità dal basso, si ha che

$$P\left(\bigcap_{n \geq 1} G_n\right) = \lim_{n \rightarrow \infty} P(G_n).$$

Notiamo che

$$P(G_n) = P\left(\bigcup_{m \geq n} C_m\right) \leq \sum_{m \geq n} P(C_m).$$

Osserviamo che, poichè per ipotesi la serie  $\sum P(C_n)$  converge, allora dev'essere

$$\lim_{n \rightarrow \infty} \sum_{m \geq n} P(C_m) = 0,$$

quindi  $P\left(\bigcap_{n \geq 1} G_n\right) = 0$ . □

Siano, ora,  $(X, \Sigma)$  ed  $(Y, \mathcal{F})$  due spazi misurabili : una funzione  $f : X \rightarrow Y$  si dice *misurabile* se  $f^{-1}(A) \in \Sigma$ ,  $\forall A \in \mathcal{F}$ . Dato uno spazio di misura  $(X, \Sigma, \mu)$ , una funzione misurabile  $f$  a valori reali è detta *semplice* se esistono  $a_1, a_2, \dots, a_n \in \mathbb{R}$  ed  $A_1, A_2, \dots, A_n \in \Sigma$  tali che  $f(x) = \sum_{i=1}^n a_i \chi_{A_i}(x)$ ,  $\forall x \in X$ . Se una funzione misurabile è definita su uno spazio di probabilità  $(X, \Sigma, P)$ , allora tale funzione è detta *variabile casuale*. Un *processo stocastico a tempo discreto* è una sequenza  $X_1, X_2, \dots, X_n, \dots$  di variabili casuali definite su uno spazio di probabilità  $(X, \Sigma, P)$ .

Supponiamo ora che  $\{X_n\}_{n \in \mathbb{N}}$  sia un processo stocastico definito su uno spazio di probabilità  $(X, \Sigma, P)$  ed a valori in uno spazio misurabile  $(E, \mathcal{E})$ . Si dimostra che, dato tale spazio misurabile, è ben definito lo spazio prodotto  $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ , ove  $\mathcal{E}^{\mathbb{N}}$  è la  $\sigma$ -algebra prodotto. L'applicazione

$$\begin{aligned} X : (X, \Sigma, P) &\rightarrow (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \\ \omega &\mapsto (X_1(\omega), X_2(\omega), \dots, X_n(\omega), \dots) \end{aligned}$$

è misurabile se e solo se  $X_n$  è misurabile  $\forall n \in \mathbb{N}$ .

Alcuni dei risultati che ora ci apprestiamo a richiamare si basano principalmente sul concetto di *integrale secondo Lebesgue* di una funzione misurabile. Al fine di non discostarci troppo dal nostro lavoro, tuttavia, diamo per scontata la definizione di tale nozione, essendo particolarmente laboriosa ed essendo facilmente reperibile nella maggior parte dei testi di base riguardanti la teoria della probabilità (si consideri, a titolo di esempio, [2, pag. 49]).

**Lemma 4** (Diseguaglianza di Markov). *Sia  $f$  una funzione nonnegativa ed integrabile su uno spazio di probabilità  $(X, \Sigma, \mu)$  e siano  $\delta, \epsilon > 0$ . Se  $\int f d\mu \leq \epsilon\delta$ , allora  $f(x) \leq \epsilon$ , eccetto che per un insieme di misura al più  $\delta$ .*

*Dimostrazione.* Sia  $D := \{x \in X \text{ tali che } f(x) > \epsilon\}$ . Consideriamo la funzione caratteristica

$$\chi_D(x) = \begin{cases} 1 & \text{se } f(x) > \epsilon \\ 0 & \text{se } f(x) \leq \epsilon \end{cases}$$

ed osserviamo che chiaramente  $\chi_D(x) \leq \frac{f(x)}{\epsilon}$ ,  $\forall x \in X$ . Dunque

$$\begin{aligned} \mu(D) &= \int_X \chi_D(x) d\mu(x) \\ &\leq \int_X \frac{f(x)}{\epsilon} d\mu(x) = \frac{1}{\epsilon} \int_X f(x) d\mu(x) \leq \frac{1}{\epsilon} \epsilon \delta = \delta. \end{aligned}$$

□

Ricordiamo, ora, che, dato uno spazio di misura  $(X, \Sigma, \mu)$ , si definisce  $L^1$  l'insieme di tutte le funzioni  $f$  misurabili su tale spazio e tali che  $\int |f| d\mu < +\infty$ . Si prova che tale insieme è uno spazio vettoriale normato completo.

Ora enunciamo i *teoremi di passaggio al limite sotto il segno di integrale*: tali teoremi indicano alcune condizioni necessarie per cui, data una successione convergente di funzioni misurabili, sia possibile ottenere informazioni riguardo il limite degli integrali della successione stessa.

**Teorema 1** (Teorema di convergenza monotona). *Sia  $(X, \Sigma, \mu)$  uno spazio di misura e sia  $\{f_n\}_{n \in \mathbb{N}}$  una successione di funzioni misurabili tale che:*

- i)  $f_n(x) \geq 0$ ,  $\forall x \in X$ ,  $\forall n \in \mathbb{N}$
- ii)  $f_n(x) \leq f_{n+1}(x)$ ,  $\forall x \in X$ ,  $\forall n \in \mathbb{N}$
- iii) *Esiste una funzione  $f$  misurabile nonnegativa tale che*

$$f(x) = \lim_{n \rightarrow \infty} f_n(x), \quad \forall x \in X.$$

Allora

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

*Dimostrazione.* Per la dimostrazione si veda, ad esempio, [2, §5.3 (pag. 49)] □

Ora, data una generica successione  $\{f_n\}_{n \in \mathbb{N}}$  di funzioni definite su uno spazio misurabile  $(X, \Sigma, \mu)$ , definiamo, per ogni elemento  $x \in X$ , il *limite inferiore* ed il *limite superiore* rispettivamente in questo modo:

- $\liminf_{n \in \mathbb{N}} f_n(x) = \sup_{m \in \mathbb{N}} \inf_{n \geq m} f_n(x)$

- $\limsup_{n \in \mathbb{N}} f_n(x) = \inf_{m \in \mathbb{N}} \sup_{n \geq m} f_n(x)$ .

Ora che abbiamo a disposizione questi concetti, possiamo procedere enunciando un lemma dimostrabile facilmente utilizzando il teorema precedente e la proprietà di monotonia degli integrali.

**Lemma 5** (Lemma di Fatou). *Sia  $(X, \Sigma, \mu)$  uno spazio di misura e sia  $\{f_n\}_{n \in \mathbb{N}}$  una successione di funzioni misurabili nonnegative. Allora*

$$\int \liminf_{n \in \mathbb{N}} f_n d\mu \leq \liminf_{n \in \mathbb{N}} \int f_n d\mu.$$

*Dimostrazione.* Per la dimostrazione si veda, ad esempio, [2, §5.4 (pag. 52)] □

Un altro risultato importante e fondamentale per la nostra trattazione è dato dal seguente teorema.

**Teorema 2** (Teorema di convergenza dominata). *Sia  $(X, \Sigma, \mu)$  uno spazio di misura e sia  $\{f_n\}_{n \in \mathbb{N}}$  una successione di funzioni misurabili tali che :*

i)  $f_n \in L^1, \forall n \in \mathbb{N}$

ii) *Esiste una funzione misurabile  $f$  tale che, dato  $x \in X$ ,*

$$\lim_{n \rightarrow \infty} f_n(x) = f(x)$$

iii) *Esiste una funzione nonnegativa  $g \in L^1$  tale che  $|f_n(x)| \leq g(x), \forall x \in X, \forall n \in \mathbb{N}$*

*Sotto queste ipotesi, si ha che allora  $f \in L^1$  e che  $\lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0$ , dunque  $\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu$ .*

*Dimostrazione.* Anche in questo caso, si consiglia di vedere [2, §5.9 (pag. 54)]. □

Questo teorema ci permette di dimostrare un lemma importante che dovremo utilizzare in seguito e che ci apprestiamo ad enunciare.

**Lemma 6** (Integrabilità uniforme). *Sia  $(X, \Sigma, \mu)$  uno spazio di probabilità e sia  $f \in L^1(X, \Sigma, \mu)$ . Allora, per ogni  $\eta > 0$ , esiste  $\delta > 0$  tale che, comunque si scelga  $A \in \Sigma$ , si ha che, se  $\mu(A) < \delta$ , allora  $\int_A |f(x)| d\mu(x) \leq \eta$ .*

*Dimostrazione.* Osserviamo innanzitutto che, per il teorema di convergenza dominata, si ha che

$$\begin{aligned} \lim_{M \rightarrow \infty} \int_{\{x \in X : |f(x)| > M\}} |f(x)| d\mu(x) &= \\ &= \lim_{M \rightarrow \infty} \int_X |f(x)| \chi_{\{x \in X : |f(x)| > M\}}(x) d\mu(x) = 0. \end{aligned}$$

Dunque questo significa che

$$\forall \eta > 0, \exists M_0 \text{ tale che } \int_{\{x \in X : |f(x)| > M_0\}} |f(x)| d\mu(x) \leq \frac{\eta}{2}.$$

Sia ora  $A \in \Sigma$ , osserviamo che

$$\begin{aligned} & \int_A |f(x)| d\mu(x) \\ &= \int_{A \cap \{x \in X : |f(x)| > M_0\}} |f(x)| d\mu(x) + \int_{A \cap \{x \in X : |f(x)| \leq M_0\}} |f(x)| d\mu(x) \\ &\leq \frac{\eta}{2} + M_0 P(A \cap \{x \in X : |f(x)| \leq M_0\}) \\ &\leq \frac{\eta}{2} + M_0 P(A). \end{aligned} \quad (1.1)$$

Definiamo, quindi,  $\delta := \frac{\eta}{2M_0}$ : se  $A$  è tale che  $\mu(A) < \delta$ , dunque, dalla (1.1) otteniamo che

$$\int_A |f(x)| d\mu(x) \leq \eta.$$

□

## 1.2 Impacchettamenti e ricopimenti

In questa sottosezione lasciamo momentaneamente in disparte la probabilità per concentrarci su alcuni risultati di combinatoria necessari nella dimostrazione del teorema ergodico.

In particolare, consideriamo l'insieme  $\mathbb{N}$  dei numeri naturali e tutti e soli gli intervalli del tipo  $[n, m] = \{j \in \mathbb{N} | n \leq j \leq m\}$ .

**Definizione 1.** Si dice *ricoprimento forte* di  $\mathbb{N}$  un insieme  $\mathcal{C}$  definito da una funzione a valori interi  $n \mapsto m(n)$  (ove  $m(n) \geq n, \forall n \in \mathbb{N}$ ) e formato da tutti gli intervalli nella forma  $[n, m(n)], n \in \mathbb{N}$  (il termine forte sta ad indicare il fatto che ogni numero naturale debba essere estremo sinistro di un intervallo appartenente a  $\mathcal{C}$ ).

Osserviamo che ogni ricoprimento forte ammette un sottoricoprimento  $\mathcal{C}'$  formato da elementi *disgiunti* di  $\mathcal{C}$ : è sufficiente, ad esempio, definire tale sottoinsieme come  $\mathcal{C}' := \{[n_i, m(n_i)], \text{ dove } n_1 = 1 \text{ e } n_{i+1} = 1 + m(n_i)\}, i \geq 1$ . Il problema che ci poniamo, dunque, è il seguente: dato un ricoprimento forte  $\mathcal{C}$ , è possibile ottenere un sottoricoprimento  $\mathcal{C}'$  di  $\mathcal{C}$ , formato da intervalli disgiunti, appartenenti ad  $[1, K]$  e tali che la loro unione generi tutto l'intervallo? In generale, la risposta è negativa: tuttavia è possibile ottenere un sottoricoprimento che ci permetta di ricostruire la maggior parte di tale intervallo.

Iniziamo da un caso particolare che ci permetta di verificare immediatamente ed in maniera diretta l'affermazione appena fatta, così da fissare bene le idee: supponiamo

che gli intervalli in  $\mathcal{C}$  abbiano tutti la stessa lunghezza, sia  $L$ . In tal caso, definiamo il sottoricoprimento  $\mathcal{C}'$  in questo modo:

$$\mathcal{C}' = \left\{ [iL + 1, (i + 1)L] : 0 \leq i \leq \frac{(K - L)}{L} \right\}.$$

Si può osservare facilmente che questo insieme, per come è definito ed essendo costituito da intervalli consecutivi e disgiunti di lunghezza  $L$ , ricopre tutto l'intervallo  $[1, K]$  fatta eccezione per, al massimo, gli ultimi  $L - 1$  elementi.

Inoltre, osserviamo che, dato  $\delta > 0$ , se  $K$  ed  $L$  sono tali che  $K > \frac{L}{\delta}$ , allora la porzione di intervallo che non viene ricoperta ha cardinalità al massimo  $L - 1 < L < K\delta$  e pertanto il rapporto tra la sua lunghezza e quella di  $[1, K]$  è strettamente minore di  $\delta$ .

Ora consideriamo un caso più generale: sia  $[1, K]$  un dato intervallo e sia  $\mathcal{C}$  un ricoprimento forte i cui elementi abbiano lunghezza variabile.

Il nostro obiettivo è far vedere che, se la maggior parte degli intervalli di  $\mathcal{C}$  aventi estremo sinistro in  $[1, K]$  ha lunghezza non superiore ad  $L$ , allora è possibile estrarre un sottoricoprimento  $\mathcal{C}'$  incluso interamente in  $[1, K]$  e tale che l'unione dei suoi elementi ricostruisca quasi completamente tale intervallo, a meno di una frazione di esso, la quale non supera comunque un certo valore stimabile. Innanzitutto, è opportuno introdurre alcune definizioni, al fine di tradurre in termini rigorosi ciò che intendiamo provare e, quindi, fornire una dimostrazione valida.

**Definizione 2.** Sia  $\mathcal{C}$  un ricoprimento forte di  $\mathbb{N}$ . Un intervallo  $[1, K]$  è detto  $(L, \delta)$ -fortemente ricoperto se, dati  $L > 0$ ,  $\delta > 0$ ,

$$\frac{|\{n \in [1, K] : m(n) - n + 1 > L\}|}{K} \leq \delta.$$

**Definizione 3.** Una collezione  $\mathcal{C}'$  di intervalli inclusi in un dato intervallo  $[1, K]$  è detta  $(1 - \delta)$ -impacchettamento di  $[1, K]$  se gli intervalli di  $\mathcal{C}'$  sono disgiunti e se la loro unione ha cardinalità maggiore od uguale a  $(1 - \delta)K$ .

Ora abbiamo a disposizione tutti gli elementi per dimostrare il lemma principale di questa sottosezione.

**Lemma 7** (Lemma di impacchettamento). *Sia  $\mathcal{C}$  un ricoprimento forte di  $\mathbb{N}$  e sia  $\delta > 0$  dato. Se  $K > \frac{L}{\delta}$  e se  $[1, K]$  è  $(L, \delta)$ -fortemente coperto da  $\mathcal{C}$ , allora esiste un sottoricoprimento  $\mathcal{C}' \subset \mathcal{C}$  che è un  $(1 - 2\delta)$ -impacchettamento di  $[1, K]$ .*

*Dimostrazione.* Innanzitutto, per ipotesi sappiamo che  $K > \frac{L}{\delta}$  e che  $|\{n \in [1, K] : m(n) - n + 1 > L\}| \leq \delta K$ . Ora cerchiamo di costruire un  $(1 - 2\delta)$ -impacchettamento dell'intervallo  $[1, K]$ . L'idea è quella di procedere iterativamente da sinistra a destra, selezionando ad ogni passo il primo intervallo di lunghezza non superiore ad  $L$  e che sia disgiunto da quello selezionato in precedenza, fermandoci quando l'estremo destro dell'ultimo intervallo estratto dista da  $K$  per una quantità minore di  $L$ . Più precisamente, poniamo  $m(0) = n_0 = 0$  e definiamo per induzione:

$$n_i = \min\{j \in [1 + m(n_{i-1}), K - L] : m(j) - j + 1 \leq L\}, \quad \forall i \geq 1,$$

fermandoci quando la condizione non è più soddisfatta. Se  $I$  è l'insieme degli indici  $j$  che selezioniamo, osserviamo che chiaramente  $m(n_j) - n_j + 1 \leq L, \forall j \in I$  e che quindi  $m(n_{|I|}) \leq K$ . Il procedimento si arresta quando si verifica uno dei seguenti casi:

- $I$  è tale che  $m(n_{|I|}) > K - L$ : osserviamo che  $1 + m(n_{|I|}) > K - L$ , pertanto non possiamo selezionare ulteriori intervalli disgiunti dai precedenti e di lunghezza inferiore ad  $L$  poiché, anche qualora ve ne fossero, non siamo sicuri che questi siano inclusi completamente in  $[1, K]$ , essendo la distanza del loro estremo sinistro da  $K$  inferiore di  $L$
- Non esiste  $j \in [1 + m(n_{|I|}), K - L]$ , tale che  $m(j) - j + 1 \leq L$ : anche se esistessero degli intervalli disgiunti dai precedenti ed aventi estremo sinistro a distanza maggiore di  $L$  da  $K$ , tuttavia, siccome tali intervalli hanno tutti lunghezza strettamente maggiore di  $L$ , in generale non è detto che siano completamente inclusi in  $[1, K]$ .

Ora vogliamo provare che l'insieme  $\mathcal{C}' = \{[n_i, m(n_i)] : 1 \leq i \leq |I|\}$  sia effettivamente un  $(1 - 2\delta)$ -impacchettamento di  $[1, K]$ . Innanzitutto, osserviamo che gli elementi di  $\mathcal{C}'$  sono tra loro disgiunti per costruzione e, in virtù delle osservazioni fatte, sono tutti inclusi in  $[1, K]$ . Resta da provare, pertanto, che la loro unione  $\mathcal{U}$  abbia lunghezza almeno  $(1 - 2\delta)K$ .

Osserviamo che l'intervallo  $(K - L, K]$  ha lunghezza massima  $L - 1 < L < \delta K$ , dunque sicuramente

$$|(K - L, K] \setminus \mathcal{U}| \leq \delta K.$$

Ora consideriamo l'intervallo  $[1, K - L]$ . Per come abbiamo definito  $\mathcal{C}'$  (e per come abbiamo costruito  $\mathcal{U}$ ), si ha che:

$$\text{Se } j \in [1, K - L] \setminus \mathcal{U}, \text{ allora } m(j) - j + 1 > L.$$

Ma allora, se applichiamo l'ipotesi di  $(L, \delta)$ -ricoprimento forte, posso concludere che

$$|[1, K - L] \setminus \mathcal{U}| \leq \delta K.$$

Quindi si ha sicuramente che

$$\begin{aligned} |[1, K] \setminus \mathcal{U}| &= |([1, K - L] \setminus \mathcal{U}) \cup ((K - L, K] \setminus \mathcal{U})| = \\ &= |[1, K - L] \setminus \mathcal{U}| + |(K - L, K] \setminus \mathcal{U}| \leq \delta K + \delta K = 2\delta K \end{aligned}$$

e così facendo abbiamo dimostrato la tesi.  $\square$

### 1.3 Trasformazioni stazionarie ed ergodiche

In questa sezione consideriamo le funzioni definite su un generico spazio di probabilità  $(X, \Sigma, \mu)$  ed a valori nel medesimo. In particolare, definiamo in maniera rigorosa i concetti di trasformazione stazionaria e di trasformazione ergodica, insieme ad alcune proprietà che saranno necessarie in seguito.

**Definizione 4** (Trasformazione stazionaria). Sia  $(X, \Sigma, \mu)$  uno spazio di probabilità. Un'applicazione  $T : X \rightarrow X$  si dice *trasformazione stazionaria* se è misurabile e se

$$\mu(T^{-1}B) = \mu(B), \quad \forall B \in \Sigma.$$

**Definizione 5** (Trasformazione ergodica). Una trasformazione stazionaria  $T$  si dice *ergodica* se, dato  $B \in \Sigma$ ,

$$T^{-1}B = B \Rightarrow \mu(B) = 0 \text{ o } \mu(B) = 1.$$

**Lemma 8** (Definizioni equivalenti di ergodicità). *Sia  $(X, \Sigma, \mu)$  uno spazio di probabilità. Data una trasformazione stazionaria  $T : X \rightarrow X$ , sono equivalenti :*

- a)  $T$  è ergodica.
- b)  $T^{-1}B \subseteq B \Rightarrow \mu(B) = 0$  o  $\mu(B) = 1$ .
- c)  $T^{-1}B \supseteq B \Rightarrow \mu(B) = 0$  o  $\mu(B) = 1$ .
- d)  $\mu(T^{-1}B \Delta B) = 0 \Rightarrow \mu(B) = 0$  o  $\mu(B) = 1$ ,  
ove  $A \Delta B = (A \cup B) \setminus (A \cap B)$  è la differenza simmetrica tra  $A$  e  $B$ .
- e) Se  $f(Tx) = f(x)$  quasi certamente, allora  $f$  è costante quasi certamente.
- f) Se  $f(Tx) \geq f(x)$  quasi certamente (oppure  $f(Tx) \leq f(x)$  quasi certamente), allora  $f$  è costante quasi certamente.

*Dimostrazione.* Dimostriamo, inizialmente, l'equivalenza di a) e b). Osserviamo che b)  $\Rightarrow$  a) è ovvia : basta notare che, se  $T : X \rightarrow X$  è un'applicazione tale che, dato  $B \in \Sigma$ ,  $T^{-1}B = B$ , allora in particolare  $T^{-1}B \subseteq B$ . In virtù delle ipotesi assunte, quindi, si ha che  $\mu(B) = 0$  o  $\mu(B) = 1$  e dunque abbiamo dimostrato l'ergodicità di  $T$ .

Procediamo ora con la dimostrazione dell'implicazione inversa : sia  $T$  una trasformazione ergodica e sia  $B \in \Sigma$  tale che  $T^{-1}B \subseteq B$ . Siccome l'applicazione inversa  $T^{-1}$  preserva le relazioni insiemistiche, osserviamo che

$$T^{-1}B \subseteq B \Rightarrow T^{-(n+1)}B \subseteq T^{-n}(B), \quad \forall n \in \mathbb{N}. \quad (1.2)$$

Prima di proseguire con la dimostrazione osserviamo che  $\bigcap_{n \geq 1} T^{-n}B = \bigcap_{n \geq 0} T^{-n}B$ . Si può verificare facilmente, infatti, che  $\bigcap_{n \geq 1} T^{-n}B \supseteq \bigcap_{n \geq 0} T^{-n}B$ : tale diseguaglianza è banale poiché il secondo membro corrisponde al primo intersecato con un nuovo evento. La relazione  $\bigcap_{n \geq 1} T^{-n}B \subseteq \bigcap_{n \geq 0} T^{-n}B$ , invece, è facilmente ottenibile a partire da (1.2).

In base a queste osservazioni, quindi, se definiamo l'insieme  $C := \bigcap_{n \geq 0} T^{-n}B$ , si può notare che

$$\begin{aligned} T^{-1}C &= T^{-1}\left(\bigcap_{n \geq 0} T^{-n}B\right) = \bigcap_{n \geq 0} T^{-1}(T^{-n}B) = \\ &= \bigcap_{n \geq 0} T^{-(n+1)}B = \bigcap_{n \geq 1} T^{-n}B = \bigcap_{n \geq 0} T^{-n}B = C. \end{aligned}$$

Poichè  $T$  è per ipotesi ergodica, dev'essere per forza  $\mu(C) = 0$  oppure  $\mu(C) = 1$ . Osserviamo che per la continuità dall'alto si ha che

$$\mu(C) = \mu\left(\bigcap_{n \geq 0} T^{-n}B\right) = \lim_{n \rightarrow \infty} \mu(T^{-n}B) = \lim_{n \rightarrow \infty} \mu(B) = \mu(B).$$

Dunque  $\mu(B) = 0$  oppure  $\mu(B) = 1$  ed abbiamo dimostrato anche questa implicazione.

L'equivalenza di b) e c) è immediata. Basta considerare, infatti, che

$$T^{-1}B \supseteq B \iff (T^{-1}B)^c = T^{-1}(B^c) \subseteq B^c.$$

Supponendo, pertanto, che valga b), se  $T^{-1}B \supseteq B$ , allora, in virtù dell'osservazione fatta sopra,  $\mu(B^c) = 0$  oppure  $\mu(B^c) = 1$ , ossia  $\mu(B) = 1$  oppure  $\mu(B) = 0$  e quindi c) è verificata. Supponiamo, ora, che valga c) e che  $T^{-1}B \subseteq B$ : quest'ultimo fatto equivale ad affermare che  $T^{-1}B^c \supseteq B^c$ . Avendo supposto valida c), dunque, concludiamo che  $\mu(B^c) = 0$  oppure  $\mu(B^c) = 1$ , ovvero  $\mu(B) = 1$  oppure  $\mu(B) = 0$  e dunque b) è verificata.

Dimostriamo ora che b) è equivalente a d). Innanzitutto, dimostrare d) $\Rightarrow$ b) è semplice: osserviamo che, se  $T^{-1}B \subseteq B$ , allora

$$\mu(T^{-1}B \Delta B) = \mu(B \setminus T^{-1}B) = \mu(B) - \mu(T^{-1}B) = \mu(B) - \mu(B) = 0.$$

Se supponiamo valga d), dunque,  $\mu(B) = 0$  oppure  $\mu(B) = 1$ .

Ora proviamo la validità dell'implicazione inversa: supponiamo, dunque, che valga b). Definiamo  $C := \bigcup_{n \geq 0} T^{-n}B$  ed osserviamo che

$$T^{-1}C = T^{-1}\left(\bigcup_{n \geq 0} T^{-n}B\right) = \bigcup_{n \geq 1} T^{-n}B \subseteq \bigcup_{n \geq 0} T^{-n}B = C,$$

questo implica che  $\mu(C) = 0$  oppure  $\mu(C) = 1$ . Ora notiamo che, per come abbiamo definito  $C$  e per la proprietà di continuità dall'alto, si ha che

$$\mu(C) = \lim_{N \rightarrow \infty} \mu\left(\bigcup_{n=0}^N (T^{-n}B)\right). \quad (1.3)$$

Sia  $C_N := \bigcup_{n=0}^N (T^{-n}B)$ . Ora consideriamo i seguenti risultati che ci saranno utili nel proseguimento della dimostrazione:

i) Dati due insiemi  $A$  e  $B$ ,

$$\begin{aligned} \mu(A) - \mu(B) &= \mu(A \cap B) + \mu(A \setminus B) - (\mu(B \cap A) + \mu(B \setminus A)) \\ &= \mu(A \setminus B) - \mu(B \setminus A). \end{aligned}$$

Dunque  $|\mu(A) - \mu(B)| \leq \mu(A \setminus B) + \mu(B \setminus A) = \mu(A \Delta B)$ .

ii) Dati  $n$  insiemi  $A_1, A_2, \dots, A_n$  ed un insieme  $B$ , si può verificare che

$$\left( \bigcup_{n=1}^N A_n \right) \Delta B \subseteq \bigcup_{n=1}^N (A_n \Delta B).$$

iii) Dati due insiemi  $A$  e  $B$ , si verifica facilmente che

$$A \Delta B \subseteq (A \Delta C) \cup (C \Delta B).$$

Ora osserviamo che, essendo  $T$  stazionaria e poiché l'inversa di un'applicazione preserva le relazioni insiemistiche, in base alle nostre ipotesi possiamo notare che

$$\begin{aligned} 0 &= \mu(T^{-1}B \Delta B) = \mu(T^{-1}(T^{-1}B \Delta B)) = \mu(T^{-2}B \Delta T^{-1}B) \\ &= \mu(T^{-1}(T^{-2}B \Delta T^{-1}B)) = \dots = \mu(T^{-(n-1)}B \Delta T^{-n}B), \quad \forall n \in \mathbb{N} \end{aligned}$$

In base ad i) e ii), dunque, possiamo notare che

$$\begin{aligned} |\mu(C_N) - \mu(B)| &\leq \mu(C_N \Delta B) \\ &\leq \mu\left(\bigcup_{n=0}^N (T^{-n}B \Delta B)\right) \leq \sum_{n=0}^N \mu(T^{-n}B \Delta B). \quad (1.4) \end{aligned}$$

Ora osserviamo che, in base a iii), dato  $n \in \mathbb{N}$ ,

$$T^{-n}B \Delta B \subseteq (T^{-n}B \Delta T^{-(n-1)}B) \cup (T^{-(n-1)}B \Delta B).$$

Otteniamo, dunque, la seguente diseguaglianza

$$\begin{aligned} \mu(T^{-n}B \Delta B) &\leq \mu(T^{-n}B \Delta T^{-(n-1)}B) + \mu(T^{-(n-1)}B \Delta B) \\ &= \mu(T^{-(n-1)}B \Delta B). \end{aligned}$$

Osserviamo che tale maggiorazione può essere ripetuta per tutti gli  $n \in \mathbb{N}$ , così da ottenere la catena di diseguaglianze:

$$\begin{aligned} \mu(T^{-n}B \Delta B) &\leq \mu(T^{-(n-1)}B \Delta B) \leq \mu(T^{-(n-2)}B \Delta B) \\ &\leq \dots \leq \mu(T^{-1}B \Delta B). \end{aligned}$$

Siccome, per ipotesi,  $\mu(T^{-1}B \Delta B) = 0$ , allora, in base all'osservazione precedente, deve essere  $\mu(T^{-n}B \Delta B) = 0$ ,  $\forall n \in \mathbb{N}$ . Ma allora osserviamo che, per la (1.4), si ha che  $|\mu(C_N) - \mu(B)| = 0$ ,  $\forall N \in \mathbb{N}$ . Dunque  $\mu(C_N) = \mu(B)$ ,  $\forall N \in \mathbb{N}$  e, tornando alla (1.3), otteniamo che

$$\mu(C) = \lim_{N \rightarrow \infty} \mu(C_N) = \lim_{N \rightarrow \infty} \mu(B) = \mu(B).$$

Dunque possiamo concludere, in quanto, se  $\mu(C) = \mu(B)$ , segue che  $\mu(B) = 1$  oppure  $\mu(B) = 0$ . La validità di d) a partire da b) è quindi dimostrata.

Ora completiamo la dimostrazione provando che d) ed e) sono equivalenti. Innanzitutto supponiamo che valga d) e consideriamo una generica funzione semplice

$$f(x) = \sum_{i=1}^n a_i \chi_{B_i}(x), \text{ ove } a_i \neq a_j \text{ e } B_i \cap B_j = \emptyset, \forall i \neq j, i, j \in \{1, \dots, n\}$$

(ricordiamo che ogni funzione semplice può essere espressa in questa forma). Supponiamo che  $f(x) = f(Tx)$  quasi certamente. Osserviamo che

$$f(Tx) = \sum_{i=1}^n a_i \chi_{B_i}(Tx) = \sum_{i=1}^n a_i \chi_{T^{-1}(B_i)}(x). \quad (1.5)$$

Dato  $i \in \{1, \dots, n\}$ , se  $x \in (B_i \Delta T^{-1}(B_i))$ , ciò significa che  $x$  non può appartenere sia a  $B_i$  che a  $T^{-1}(B_i)$ : in altre parole, non può verificarsi che  $x$  e  $Tx$  appartengano entrambi all'insieme  $B_i$ . Questo fatto implica, in base alla (1.5), che sicuramente  $f(x) \neq f(Tx)$ . Possiamo affermare, dunque, che  $B_i \Delta T^{-1}(B_i) \subseteq \{x : f(x) \neq f(Tx)\}$  e che, avendo supposto  $f(x) = f(Tx)$  quasi certamente, si ha dunque:

$$\begin{aligned} \mu(B_i \Delta T^{-1}(B_i)) &\leq \mu(\{x : f(x) \neq f(Tx)\}) = 0 \quad \forall i \in \{1, \dots, n\} \\ &\Rightarrow \mu(B_i \Delta T^{-1}(B_i)) = 0 \quad \forall i \in \{1, \dots, n\}. \end{aligned}$$

Siccome abbiamo supposto che d) sia vera, segue che  $\mu(B_i) = 0$  oppure  $\mu(B_i) = 1$ ,  $\forall i \in \{1, \dots, n\}$ . Tuttavia, essendo gli insiemi  $B_i$  a due a due disgiunti, questo significa che esiste al più un solo indice  $\bar{j} \in \{1, \dots, n\}$  tale che  $\mu(B_{\bar{j}}) = 1$  e  $\mu(B_i) = 0$ ,  $\forall i \neq \bar{j}$ . Questo risultato implica che la funzione semplice  $f$ , per come è definita, è quasi certamente costante. Dunque il risultato vale per le funzioni semplici. Sia ora  $f$  una funzione generica per cui supponiamo che  $f(Tx) = f(x)$  quasi certamente. Definiamo, per ogni numero naturale  $n$ , la funzione

$$f_n(x) := \sum_{k=-n2^n}^{+n2^n} \frac{k}{2^n} \chi_{\{x : f(x) \in (\frac{k-1}{2^n}, \frac{k}{2^n}] \}}(x).$$

Osserviamo che si tratta di una funzione semplice e che, per ogni  $n \in \mathbb{N}$ , il valore che assume  $f_n$  in ciascun  $x \in X$  dipende dal valore assunto da  $f$  in tale punto. In particolare, dato  $n \in \mathbb{N}$  e dato  $x \in X$ , se  $f(Tx) = f(x)$  allora

$$\begin{aligned} f_n(Tx) &= \sum_{k=-n2^n}^{+n2^n} \frac{k}{2^n} \chi_{\{x : f(Tx) \in (\frac{k-1}{2^n}, \frac{k}{2^n}] \}}(x) \\ &= \sum_{k=-n2^n}^{+n2^n} \frac{k}{2^n} \chi_{\{x : f(x) \in [\frac{k-1}{2^n}, \frac{k}{2^n}] \}}(x) = f_n(x). \end{aligned}$$

Dunque, se  $f(Tx) = f(x)$  quasi certamente, allora  $f_n(Tx) = f_n(x)$  quasi certamente,  $\forall n \in \mathbb{N}$ . Pertanto, in virtù del risultato ottenuto per le funzioni semplici, possiamo

affermare che, per ogni numero naturale  $n$ ,  $f_n(x) = c_n \in \mathbb{R}$ ,  $\forall x \in A_n$ , ove  $A_n$  è un insieme tale che  $\mu(A_n) = 1$ . Notiamo, inoltre, che, fissato  $x \in X$ ,  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ . Definiamo, quindi, l'insieme  $A := \bigcap_{n \geq 0} A_n$  ed osserviamo che  $\mu(A) = 1$ . Infatti osserviamo che  $A^c = \bigcup_{n \geq 0} A_n^c$ , quindi  $\mu(A^c) \leq \sum_{n \geq 0} \mu(A_n^c) = 0$ , pertanto  $\mu(A^c) = 0$ , ovvero  $\mu(A) = 1$ . Dunque, se  $x \in A$ , allora la successione  $\{f_n(x)\}_{n \in \mathbb{N}}$  è una successione di numeri reali  $\{c_n\}_{n \in \mathbb{N}}$  convergente ad  $f(x)$ . Poichè questo fatto vale per ogni  $x \in A$ , allora per forza  $f(x)$  dev'essere costante su tutto  $A$ , altrimenti si avrebbe che la successione  $\{c_n\}_{n \in \mathbb{N}}$  converge a più valori, il che è assurdo. Pertanto  $f$  è costante su  $A$ , dunque a meno di un insieme di misura nulla.

Supponiamo ora che e) sia valida e sia  $B \in \Sigma$  tale che  $\mu(T^{-1}B\Delta B) = 0$ . Osserviamo che

$$0 = \mu(T^{-1}B\Delta B) = \int |\chi_{T^{-1}B}(x) - \chi_B(x)|d\mu(x),$$

da cui segue facilmente che

$$\begin{aligned} |\chi_{T^{-1}B}(x) - \chi_B(x)| &= 0, \text{ q.c. } \Rightarrow \chi_{T^{-1}B}(x) = \chi_B(x) \text{ q.c.} \\ &\iff \chi_B(Tx) = \chi_B(x) \text{ q.c.} . \end{aligned}$$

Dunque, supponendo e) vera, segue che  $\chi_B(x)$  è quasi certamente costante. Siccome tale funzione può assumere i soli valori 0 ed 1, allora o  $\chi_B(x) = 1$  q.c. (ed in tal caso  $\mu(B) = 1$ ), oppure  $\chi_B(x) = 0$  q.c. (ed in tal caso  $\mu(B) = 0$ ). Dunque abbiamo provato la validità di d).

Ora dimostriamo che e) ed f) sono equivalenti. L'implicazione  $f) \Rightarrow e)$  è ovvia poichè basta notare che  $f(Tx) = f(x)$  quasi certamente  $\Rightarrow f(Tx) \geq f(x)$  (oppure  $f(Tx) \leq f(x)$ ) quasi certamente.

Ora dimostriamo che e) implica f). Supponiamo, dunque, che valga e) che  $f(Tx) \geq f(x)$  quasi certamente (nel caso in cui  $f(Tx) \leq f(x)$  quasi certamente, si procede in maniera analoga). Definiamo la funzione  $g(x) := f(Tx) - f(x)$ : osserviamo che, in virtù dell'ipotesi,  $g(x) \geq 0$  quasi certamente. Notiamo che, in virtù della stazionarietà di  $T$

$$\int g(x)d\mu(x) = \int f(Tx)d\mu(x) - \int f(x)d\mu(x) = 0.$$

In base a queste due proprietà, dunque, dev'essere  $g(x) = 0$  quasi certamente, ma allora  $f(Tx) = f(x)$  quasi certamente, dunque  $f$  è costante quasi certamente.  $\square$

Sia ora  $(E, \mathcal{E})$  uno spazio misurabile e sia  $(E^\mathbb{N}, \mathcal{E}^\mathbb{N})$  lo spazio prodotto definito a partire da  $E$  (costituito dagli elementi  $(x_1, x_2, \dots, x_n, \dots)$  tali che  $x_n \in E$ ,  $\forall n \in \mathbb{N}$ ) e munito della  $\sigma$ -algebra prodotto  $\mathcal{E}^\mathbb{N}$ . Definiamo, quindi, l'operazione di *shift*

$$\begin{aligned} T : E^\mathbb{N} &\rightarrow E^\mathbb{N} \\ (x_1, x_2, \dots, x_n, \dots) &\mapsto (x_2, x_3, \dots, x_n, \dots). \end{aligned}$$

Consideriamo ora un processo stocastico  $\{X_n\}_{n \in \mathbb{N}}$  a valori in  $E$ , cioè

$$X_n : (X, \Sigma, P) \rightarrow (E, \mathcal{E}), \quad \forall n \in \mathbb{N},$$

ove  $(X, \Sigma, P)$  è uno spazio di probabilità. Definiamo, su  $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ , la *misura di Kolmogorov* per il processo  $\{X_n\}_{n \in \mathbb{N}}$  in questo modo:

$$\begin{aligned}\mu(A_1 \times A_2 \times \dots \times A_n \times \dots) := \\ P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n, \dots), \quad \forall \{A_i\}_{i \in \mathbb{N}}, A_i \in \mathcal{E}\end{aligned}$$

( $\mu$  è detta *legge della successione*  $\{X_n\}_{n \in \mathbb{N}}$ ). Tale processo, dunque, si dice *ergodico* se l'operazione di shift è ergodica rispetto alla misura di Kolmogorov associata al processo  $\{X_n\}_{n \in \mathbb{N}}$ .

## 1.4 Il teorema ergodico

### 1.4.1 Enunciato del teorema ergodico

**Teorema 3** (Teorema ergodico). *Sia  $T$  una trasformazione stazionaria su uno spazio di probabilità  $(X, \Sigma, \mu)$  e sia  $f$  una funzione integrabile, allora la media  $\frac{1}{n} \sum_{i=1}^n f(T^{i-1}x)$  converge quasi certamente ed in  $L^1$ , per  $n \rightarrow \infty$ , verso una funzione  $T$ -invariante  $f^*(x)$ .*

**Corollario 1.** *Se  $T$  è una trasformazione ergodica, allora  $f^*(x) = \int f d\mu$ ,  $\forall x \in X$ .*

*Dimostrazione.* Sia  $f_n(x) := \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x)$ .

Siccome  $f_n$  converge in  $L^1$  ad  $f^*$ , allora

$$\left| \int f^n d\mu - \int f^* d\mu \right| = \left| \int (f^n - f^*) d\mu \right| \leq \int |f_n - f^*| d\mu \rightarrow 0, \text{ per } n \rightarrow \infty,$$

dunque  $\int f_n d\mu \rightarrow \int f^* d\mu$ , per  $n \rightarrow \infty$ .

Inoltre, notiamo che,  $\forall n \in \mathbb{N}$ ,

$$\int f_n(x) d\mu = \int \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) d\mu = \frac{1}{n} \sum_{i=1}^n \int f(T^{i-1}x) d\mu = \int f(x) d\mu,$$

poiché, essendo  $T$  una trasformazione stazionaria,

$$\int f(T^n(x)) d\mu = \int f(x) d\mu, \forall n \in \mathbb{N}.$$

Dunque dev'essere  $\int f d\mu = \int f^* d\mu$ .

Inoltre, se supponiamo che  $T$  sia una trasformazione ergodica, essendo  $f^*$   $T$ -invariante, allora, in virtù del Lemma 4,  $f^*(x)$  è costante su tutto il dominio, pertanto  $f^*(x) = c \in \mathbb{R}$ ,  $\forall x \in X$ , dunque :

$$\int f(x) d\mu = \int f^*(x) d\mu = \int c d\mu = \mu(X)c = c,$$

pertanto, nel caso in cui  $T$  sia anche ergodica la funzione  $f^*(x)$  (media temporale) è costante su tutto il dominio ed ha valore pari a  $\int f d\mu$  (media spaziale).  $\square$

### 1.4.2 Il teorema ergodico : caso binario

**Teorema 4** (Teorema ergodico: caso binario). *Se  $\{X_n\}$  è un processo ergodico tale che  $\{X_n\} \in \{0, 1\}$ ,  $\forall n \in \mathbb{N}$ , allora la media  $\frac{X_1 + X_2 + \dots + X_n}{n}$  converge quasi certamente al valore costante  $E(X_1) = \mu(1) = \mu(\{x : x_1 = 1\})$ .*

*Dimostrazione.* Il nostro obiettivo è provare che :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = \mu(1) \text{ quasi ovunque,} \quad (1.6)$$

ove, ricordiamo,  $\mu(1) = \mu\{x : x_1 = 1\} = E(X_1)$  e  $\mu_1$  è la legge di  $X$  su  $\{0, 1\}^{\mathbb{N}}$ . Supponiamo, per assurdo, che (1.6) sia falsa.

Più precisamente, supponiamo che l'insieme

$$B = \left\{ x : \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i > \mu(1) + \epsilon \right\}$$

sia tale che  $\mu(B) > 0$  (il caso complementare, in cui si considera l'insieme degli  $x \in X$  tali che  $\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i < \mu(1) - \epsilon$ , si tratta in modo analogo). Osserviamo che

$$\limsup_{n \rightarrow \infty} \frac{x_1 + x_2 + \dots + x_n}{n} = \limsup_{n \rightarrow \infty} \frac{x_2 + x_3 + \dots + x_{n+1}}{n}.$$

Infatti, notiamo che

$$\frac{x_2 + x_3 + \dots + x_{n+1}}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} + \frac{x_{n+1}}{n} - \frac{x_1}{n}.$$

Inoltre, poichè  $x_n \in \{0, 1\}$ ,  $\forall n \in \mathbb{N}$ , si ha che  $\lim_{n \rightarrow \infty} \frac{x_{n+1}}{n} - \frac{x_1}{n} = 0$ , dunque

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{x_2 + x_3 + \dots + x_{n+1}}{n} &= \\ \limsup_{n \rightarrow \infty} \frac{x_1 + x_2 + \dots + x_n}{n} + \lim_{n \rightarrow \infty} \frac{x_{n+1}}{n} - \frac{x_1}{n} &= \\ &= \limsup_{n \rightarrow \infty} \frac{x_1 + x_2 + \dots + x_n}{n} \end{aligned}$$

Pertanto, se la sequenza  $(x_1, x_2, \dots, x_n, \dots) \in B$ , allora la sequenza, ottenuta tramite uno shift,  $(x_2, x_3, \dots, x_n, \dots) \in B$  e questo fatto vale anche per tutte le operazioni di shift successive : abbiamo provato, quindi, che l'insieme  $B$  è  $T$ -invariante e, poichè abbiamo assunto che  $\mu$  sia ergodica rispetto a  $T$  e che  $\mu(B) > 0$ , allora deve valere  $\mu(B) = 1$ . Supponiamo ora che  $x = (x_1, x_2, \dots, x_n, \dots) \in B$  : essendo  $B$  un insieme  $T$ -invariante, allora

$$(x_n, x_{n+1}, \dots, x_{n+m}, \dots) \in B, \forall n \in \mathbb{N},$$

$$\iff \limsup_{m \rightarrow \infty} \frac{1}{m} \sum_{k=n}^{n+m-1} x_k > \mu(1) + \epsilon, \forall n \in \mathbb{N},$$

dunque per ogni intero  $n$  esiste un intero  $m(n) \geq n$  tale che

$$\frac{x_n + x_{n+1} + \dots + x_{m(n)}}{m(n) - n + 1} > \mu(1) + \epsilon.$$

Definiamo, pertanto, la collezione di intervalli

$$\mathcal{C}(x) := \{[n, m(n)] : n \in \mathbb{N}\} : \quad (1.7)$$

tale insieme è un ricoprimento forte di  $\mathbb{N}$ . Osserviamo, inoltre, che tale insieme è determinato in modo casuale, al variare dei valori assunti dalla sequenza  $(x_1, x_2, \dots, x_n, \dots)$ , i quali, appunto, sono determinati da una sequenza di variabili casuali. Osserviamo che, poichè  $\mu(B) = 1$ , allora

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i > \mu(1) + \epsilon, \text{ quasi certamente,}$$

ovvero ad eccezione di un sottoinsieme di  $\{0, 1\}^{\mathbb{N}}$  avente misura nulla. Dunque possiamo affermare che quasi certamente esiste un  $m(1) \in \mathbb{N}$  tale che

$$\frac{x_1 + x_2 + \dots + x_{m(1)}}{m(1)} > \mu(1) + \epsilon,$$

questo significa che la variabile casuale  $m(1)$  (il cui valore varia al variare dei valori assunti dal processo ergodico  $\{X_n\}_{n \in \mathbb{N}}$ ) assume quasi certamente un valore finito in  $\mathbb{N}$  e, pertanto, è limitata a meno di un sottoinsieme di  $\{0, 1\}^{\mathbb{N}}$  avente probabilità piccola a piacere. Più precisamente, dato  $\delta > 0$ , esiste un numero  $L$  tale che se  $D := \{x = (x_1, x_2, \dots, x_n, \dots) : m(1) > L\}$  allora  $\mu(D) < \delta^2$ .

Definiamo la funzione

$$g_K(x) := \frac{1}{K} \sum_{i=1}^K \chi_D(T^{i-1}(x)),$$

ove  $\chi_D(x) = \begin{cases} 1 & \text{se } x \in D \\ 0 & \text{se altrimenti.} \end{cases}$

Ora vogliamo calcolare l'integrale di questa funzione.

Innanzitutto osserviamo che, poichè  $T$  è una trasformazione stazionaria,  $\mu = \mu \circ T$ .

Dunque

$$\begin{aligned} \int g_K(x) d\mu &= \frac{1}{K} \sum_{i=1}^K \int \chi_D(T^{i-1}x) d\mu = \\ &= \frac{1}{K} \sum_{i=1}^K \int \chi_D(x) d\mu = \frac{1}{K} K \mu(D) = \mu(D). \end{aligned}$$

Poichè, quindi,  $\int g_K(x) d\mu = \mu(D) < \delta^2$ , per la Disegualanza di Markov si ha che l'insieme  $G_K := \{x : g_K(x) \leq \delta\}$  è tale che  $\mu(G_K) \geq (1 - \delta)$ .

Supponiamo  $x \in G_K$ , allora:

$$g_K(x) = \frac{1}{K} \sum_{i=1}^K \chi_D(T^{i-1}x) \leq \delta.$$

Ora osserviamo questo fatto importante: per definizione dell'insieme  $D$ , sia  $x = (x_1, x_2, \dots, x_n, \dots)$ , allora

$$\begin{aligned} T^{n-1}(x) = (x_n, x_{n+1}, \dots, \dots) \in D &\iff m(n) - n + 1 > L \\ &\iff |[n, m(n)]| > L. \end{aligned}$$

Se  $x \in G_K$ , dunque, possiamo scrivere

$$\begin{aligned} g_k(x) &= \frac{1}{K} \sum_{i=1}^K \chi_D(T^{i-1}(x)) = \\ &= \frac{|\{n \in [1, K] : m(n) - n + 1 > L\}|}{K} = \frac{|\{n \in [1, K] : |[n, m(n)]| > L\}|}{K} \leq \delta. \end{aligned}$$

Dunque abbiamo provato che, se  $x \in G_K$ , allora l'insieme  $\mathcal{C}(x)$ , definito in (1.7), è un  $(L, \delta)$ -ricoprimento forte dell'intervallo  $[1, K]$ .

In virtù del *Lemma di impacchettamento*, quindi, esiste un sottoinsieme

$$\mathcal{C}'(x) = \{[n_i, m(n_i)] : i \leq I(x)\} \subset \mathcal{C}(x),$$

il quale è un  $(1 - 2\delta)$ -impacchettamento di  $[1, K]$ . Poichè le  $x_i$  sono nonnegative, gli intervalli di  $\mathcal{C}'(x)$  sono disgiunti e sono tutti contenuti in  $[1, K]$ , allora:

$$\begin{aligned} \sum_{j=1}^K x_j &\geq \sum_{i=1}^{I(x)} \sum_{j=n_i}^{m(n_i)} \geq \sum_{i=1}^{I(x)} (m(n_i) - n_i + 1)[\mu(1) + \epsilon] = \\ &= [\mu(1) + \epsilon] \sum_{i=1}^{I(x)} |[n_i, m(n_i)]| \geq [\mu(1) + \epsilon](1 - 2\delta)K. \end{aligned}$$

Ora osserviamo che, poichè ovviamente  $\sum_{j=1}^K x_j \geq (\sum_{j=1}^K x_j) \chi_{G_k}(x)$ , si ha:

$$\begin{aligned} E\left(\sum_{j=1}^K x_j\right) &\geq E\left(\left(\sum_{j=1}^K x_j\right) \chi_{G_k}(x)\right) \geq \\ &\geq E([\mu(1) + \epsilon](1 - 2\delta)K \chi_{G_k}(x)) = [\mu(1) + \epsilon](1 - 2\delta)KE(\chi_{G_k}(x)) = \\ &= [\mu(1) + \epsilon](1 - 2\delta)K\mu(G_k) \geq [\mu(1) + \epsilon](1 - 2\delta)K(1 - \delta), \end{aligned}$$

cosicché:

$$\mu(1) = E\left(\frac{1}{K} \sum_{j=1}^K x_j\right) \geq [\mu(1) + \epsilon](1 - 2\delta)(1 - \delta), \forall \delta > 0.$$

Questo fatto è chiaramente assurdo, poiché, per  $\delta \rightarrow 0$ , dovrebbe avversi, per la permanenza del segno,

$$\mu(1) \geq \mu(1) + \epsilon > \mu(1)$$

In questo modo, quindi, abbiamo provato che la (1.6) è vera e, pertanto, abbiamo dimostato il teorema.  $\square$

### 1.4.3 Il teorema ergodico: caso generale

In questa sottosezione dimostriamo la validità del teorema ergodico nella sua forma più astratta. Prima di procedere con la dimostrazione vera e propria, tuttavia, è necessario provare alcuni lemmi.

**Lemma 9.** *Consideriamo uno spazio di probabilità  $(X, \Sigma, \mu)$ . Siano  $f \in L^1(X, \Sigma, \mu)$ ,  $T : X \rightarrow X$  una trasformazione stazionaria ed  $\alpha$  un numero reale arbitrario. Definiamo gli insiemi*

$$B = \left\{ x \in X : \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) > \alpha \right\}$$

$$D = \left\{ x \in X : \lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} = 0 \right\},$$

allora  $\mu(D) = 1$  e  $B \cap D = T^{-n}B \cap D$ ,  $\forall n \in \mathbb{N}$ .

*Dimostrazione.* Osserviamo innanzitutto che, siccome  $f \in L^1$ , si ha,  $\forall \epsilon > 0$  :

$$\begin{aligned} \infty > \int |f| d\mu &= \int_0^{+\infty} \mu(\{x \in X : |f(x)| > t\}) dt = \\ &= \sum_{n \in \mathbb{N}} \int_{\epsilon(n-1)}^{\epsilon n} \mu(\{x \in X : |f(x)| > t\}) dt \geq \sum_{n \in \mathbb{N}} \epsilon \mu(\{x \in X : |f(x)| > \epsilon n\}) \end{aligned}$$

(la prima uguaglianza segue da un'applicazione del Teorema di Fubini, mentre l'ultima diseguaglianza segue dal fatto che il valore  $\mu(\{x \in X : |f(x)| > t\})$  decresce all'aumentare di  $t \in [\epsilon(n-1), \epsilon n]$ ). Dunque si ha che

$$\sum_{n \in \mathbb{N}} \mu(\{x \in X : |f(x)| > \epsilon n\}) < +\infty.$$

Osserviamo che, essendo  $T$  una trasformazione stazionaria, allora dev'essere

$$\begin{aligned}\mu(\{x \in X : |f(x)| > \epsilon n\}) &= \\ &= \mu(T^{-n}(\{x \in X : |f(x)| > \epsilon n\})) = \\ &\quad \mu(\{x \in X : |f(T^n x)| > \epsilon n\}).\end{aligned}$$

Dunque

$$\sum_{n \in \mathbb{N}} \mu(\{x \in X : |f(T^n x)| > \epsilon n\}) = \sum_{n \in \mathbb{N}} \mu(\{x \in X : |f(x)| > \epsilon n\}) < +\infty.$$

Per il lemma di Borel-Cantelli, quindi, possiamo affermare che

$$\begin{aligned}\mu\left(\bigcap_{n \geq 1} \bigcup_{m \geq n} \{x \in X : |f(T^m x)| > \epsilon m\}\right) &= \\ &= \mu(\{x \in X : \forall n \in \mathbb{N}, \exists m \geq n \text{ tale che } |f(T^m x)| > \epsilon m\}) = 0\end{aligned}$$

Pertanto

$$\begin{aligned}1 &= \mu(\{x \in X : \exists N = N(x) \text{ tale che } |f(T^n x)| \leq \epsilon n, \forall n \geq N\}) = \\ &= \mu(\{x \in X : \exists N = N(x) \text{ tale che } \frac{|f(T^n x)|}{n} \leq \epsilon, \forall n \geq N\}).\end{aligned}$$

Ricordiamo che questo risultato vale per  $\epsilon > 0$  fissato. Definiamo, quindi, l'insieme

$$G_\epsilon := \{x \in X : \exists N = N(x) \text{ tale che } \frac{|f(T^n x)|}{n} \leq \epsilon, \forall n \geq N\}.$$

Finora abbiamo provato che  $\mu(G_\epsilon) = 1, \forall \epsilon > 0$ . Si può notare facilmente che, dati  $\epsilon_1, \epsilon_2 > 0$ , se  $\epsilon_1 < \epsilon_2$ , allora  $G_{\epsilon_1} \subseteq G_{\epsilon_2}$ . Osserviamo, inoltre, che

$$\begin{aligned}\bigcap_{\epsilon > 0} G_\epsilon &= \{x \in X : \forall \epsilon > 0, \exists N_\epsilon = N_\epsilon(x) \text{ tale che } \frac{|f(T^n x)|}{n} \leq \epsilon, \forall n \geq N_\epsilon\} \\ &= \{x \in X : \lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} = 0\} = D.\end{aligned}$$

Osserviamo, dunque, che per continuità dall'alto si ha che

$$\mu(D) = \mu\left(\bigcap_{\epsilon > 0} G_\epsilon\right) = \lim_{\epsilon \rightarrow 0} \mu(G_\epsilon) = 1.$$

Sia ora  $x \in B \cap D$ , ovvero, ricordiamo, sia  $x \in X$  tale che :

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1} x) > \alpha \text{ e } \lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} = 0.$$

Innanzitutto proviamo che  $T(x) \in B$ . Osserviamo che

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}(T(x))) &= \frac{1}{n} \sum_{i=2}^{n+1} f(T^{i-1}x) = \\ &= \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) + \frac{1}{n} \sum_{i=2}^{n+1} f(T^{i-1}x) - \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) = \\ &= \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) + \left( \frac{f(T^n x)}{n} - \frac{f(x)}{n} \right). \end{aligned} \quad (1.8)$$

Notiamo che, per come abbiamo definito l'insieme  $D$  e poiché, per ogni  $x$  fissato, si ha chiaramente  $\lim_{n \rightarrow \infty} \frac{f(x)}{n} = 0$ , allora se  $x \in B \cap D$  si ha che

$$\lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} - \frac{f(x)}{n} = \lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} - \lim_{n \rightarrow \infty} \frac{f(x)}{n} = 0 + 0 = 0.$$

Dunque, in virtù della (1.8) e del fatto che  $x \in B$ , abbiamo

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}(T(x))) &= \\ &= \limsup_{n \rightarrow \infty} \left( \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) + \left( \frac{f(T^n x)}{n} - \frac{f(x)}{n} \right) \right) = \\ &= \limsup_{n \rightarrow \infty} \sum_{i=1}^n f(T^{i-1}x) + \lim_{n \rightarrow \infty} \left( \frac{f(T^n x)}{n} - \frac{f(x)}{n} \right) = \\ &= \limsup_{n \rightarrow \infty} \sum_{i=1}^n f(T^{i-1}x) > \alpha. \end{aligned} \quad (1.9)$$

Dunque  $T(x) \in B$ . Osserviamo che, inoltre,  $T(x) \in D$ , poiché, siccome per ipotesi  $x \in B \cap D$ , si ha

$$\lim_{n \rightarrow \infty} \frac{f(T^n(T(x)))}{n} = \lim_{n \rightarrow \infty} \frac{f(T^{n+1}x)}{n} = \lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} = 0 \quad (1.10)$$

Dunque  $x \in B \cap D \Rightarrow T(x) \in B \cap D$ . Osserviamo che, inoltre, per la (1.9) e la (1.10) vale anche l'implicazione inversa, dunque  $x \in B \cap D \iff T(x) \in B \cap D$ . Dunque  $B \cap D = T^{-1}(B \cap D) = (T^{-1}B) \cap (T^{-1}D) = T^{-1}B \cap D$ , poiché, per la (1.10),  $D = T^{-1}D$ . Osserviamo che, iterando questa proprietà, otteniamo che

$$\begin{aligned} B \cap D &= T^{-1}(B \cap D) = T^{-1}(T^{-1}(B \cap D)) = T^{-2}(B \cap D) \\ &= T^{-2}B \cap T^{-2}D = T^{-2}B \cap D = \dots = T^{-n}B \cap D = \dots \end{aligned}$$

dunque  $B \cap D = T^{-n}B \cap D, \forall n \in \mathbb{N}$ . □

**Lemma 10.** *Sia  $T$  una trasformazione stazionaria su uno spazio di probabilità  $(X, \Sigma, \mu)$ , sia  $f \in L^1(X, \Sigma, \mu)$  e sia  $\alpha$  un numero reale arbitrario. Se definiamo l'insieme*

$$B = \left\{ x : \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) > \alpha \right\},$$

*allora*

$$\int_B f(x) d\mu(x) \geq \alpha \mu(B). \quad (1.11)$$

*Dimostrazione.* Se  $\mu(B) = 0$ , il lemma è banalmente verificato. Consideriamo, dunque, il caso in cui  $\mu(B) > 0$ . Definiamo l'insieme

$$D = \left\{ x \in X : \lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} = 0 \right\}.$$

Sappiamo, dal lemma precedente, che  $\mu(D) = 1$ , dunque  $\mu(B \cap D) = \mu(B)$ . Inoltre, definiamo la *misura condizionata*

$$\mu_B(A) := \frac{\mu(A \cap B)}{\mu(B)}, \quad \forall A \in \Sigma$$

(ricordiamo che abbiamo supposto  $\mu(B) > 0$ ). Osserviamo che, in base a tale definizione ed alla definizione di integrale, si può concludere facilmente che

$$\int_B f(x) d\mu_B(x) = \frac{1}{\mu(B)} \int_B f(x) d\mu(x).$$

Siccome nella (1.11) l'integrale è calcolato sull'insieme  $B$ , in virtù delle affermazioni precedenti osserviamo che

$$\int_B f(x) d\mu(x) = \int_{B \cap D} f(x) d\mu(x) = \mu(B) \int_{B \cap D} f(x) d\mu_B(x).$$

Al fine di dimostrare il lemma, dunque, sarà sufficiente far vedere che

$$\int_{B \cap D} f(x) d\mu_B(x) \geq \alpha.$$

Ricordiamo, in base al lemma precedente, che  $B \cap D = T^{-n}B \cap D$ ,  $\forall n \in \mathbb{N}$ . Ciò equivale esattamente ad affermare che, dato  $x \in B \cap D$ , per ogni numero naturale  $n$  esiste  $m(n) \in \mathbb{N}$  tale che  $m(n) \geq n$  e che

$$\frac{\sum_{i=n}^{m(n)} f(T^{i-1}x)}{m(n) - n + 1} > \alpha.$$

Dunque, per  $x \in B \cap D$ , è sempre definito il ricoprimento forte  $\mathcal{C}(x) = \{[n, m(n)] : n \in \mathbb{N}\}$ . Osserviamo che, dato  $x \in B$ , esiste sempre  $m(1) \in \mathbb{N}$  tale che

$$\frac{\sum_{i=1}^{m(1)} f(T^{i-1}x)}{m(1)} > \alpha.$$

Questo significa che la variabile aleatoria  $m(1)$  è sempre finita, per  $x \in B$ , dunque è limitata, a meno di un insieme avente misura  $\mu_B$  piccola a piacere. Pertanto possiamo affermare che, dato  $\delta > 0$ , esiste  $L > 0$  tale che l'insieme  $E := \{x \in B : m(1) > L\}$  è tale che  $\mu_B(E) < \delta^2$ . Definiamo, per ogni  $K > 0$ ,  $K \in \mathbb{N}$ , l'insieme

$$G_K = \left\{ x \in B \cap D : \frac{1}{K} \sum_{i=1}^K \chi_{E \cap D}(T^{i-1}x) \leq \delta \right\}.$$

Osserviamo che, se definiamo  $g_K(x) := \frac{1}{K} \sum_{i=1}^K \chi_{E \cap D}(T^{i-1}x)$ , allora, per la proprietà di linearità dell'integrale e poiché  $T$  è stazionaria e  $B \cap D$  è  $T$ -invariante, abbiamo che

$$\begin{aligned} \int_{B \cap D} g_K(x) d\mu_B(x) &= \frac{1}{K} \sum_{i=1}^K \int_{B \cap D} \chi_{E \cap D}(T^{i-1}x) d\mu_B(x) = \\ &= \frac{1}{K} \sum_{i=1}^K \int_{B \cap D} \chi_{E \cap D}(x) d\mu_B(x) = \frac{1}{K} K \int_{B \cap D} \chi_{E \cap D}(x) d\mu_B(x) = \\ &= \frac{1}{K} K \mu_B(E \cap D) = \mu_B(E). \end{aligned}$$

Per la diseguaglianza di Markov, dunque, si ha che  $\mu_B(G_K) \geq 1 - \delta$ .

Supponiamo di scegliere  $K > 0$  tale che  $K \geq \frac{L}{\delta}$ : sotto queste ipotesi possiamo applicare il lemma di impacchettamento al ricoprimento forte  $\mathcal{C}$ . Più precisamente, esiste un sottoricoprimento  $\mathcal{C}'(x) = \{[n_i, m(n_i)] : i \leq I(x)\} \subset \mathcal{C}(x)$ , il quale è un  $(1 - 2\delta)$ -impacchettamento dell'intervallo  $[1, K]$ , e pertanto ricordiamo che si ottiene la seguente diseguaglianza :

$$\sum_{i=1}^{I(x)} |[n_i, m(n_i)]| \geq (1 - 2\delta)K.$$

Osserviamo che, per  $x \in G_K$

$$\begin{aligned} \sum_{j=1}^K f(T^{j-1}(x)) &\geq \sum_{j \in [1, K] \setminus \bigcup[n_i, m(n_i)]} f(T^{j-1}(x)) + \sum_{i=1}^{I(x)} \sum_{j=n(i)}^{m(n_i)} f(T^{j-1}(x)) \geq \\ &\geq \sum_{j \in [1, K] \setminus \bigcup[n_i, m(n_i)]} f(T^{j-1}(x)) + \sum_{i=1}^{I(x)} (m(n_i) - n_i + 1)\alpha = \\ &= \sum_{j \in [1, K] \setminus \bigcup[n_i, m(n_i)]} f(T^{j-1}(x)) + \alpha \sum_{i=1}^{I(x)} |[n_i, m(n_i)]| \geq \\ &\geq \sum_{j \in [1, K] \setminus \bigcup[n_i, m(n_i)]} f(T^{j-1}(x)) + \alpha K(1 - 2\delta). \end{aligned}$$

Ora, nel caso in cui  $f$  sia limitata, supponiamo  $|f(x)| < M$ ,  $M > 0$ , in base al risultato ottenuto dal lemma di impacchettamento posso affermare che

$$\begin{aligned} \sum_{j \in [1, K] \setminus \bigcup[n_i, m(n_i)]} f(T^{j-1}(x)) &\geq \sum_{j \in [1, K] \setminus \bigcup[n_i, m(n_i)]} (-M) \geq \\ &\geq -2MK\delta. \end{aligned}$$

Pertanto, in tale situazione si ha che, per  $x \in G_K$ ,

$$\sum_{j=1}^K f(T^{j-1}(x)) \geq -2MK\delta + \alpha K(1 - 2\delta)$$

, mentre, per  $x \notin G_K$ , si ha che  
e quindi

$$\int_{B \cap D} \sum_{j=1}^K f(T^{j-1}(x)) d\mu_B(x) = \int_{B \cap D} f(x) d\mu_B(x) \geq 2MK\delta + \alpha K(1 - 2\delta).$$

Dunque, facendo tendere  $\delta$  a 0, otteniamo il risultato desiderato.

Se, invece,  $f$  non è limitata, allora osserviamo che

$$\begin{aligned} \int_{B \cap D} f(x) d\mu_B(x) &= \int_{B \cap D} \frac{1}{K} \sum_{j=1}^K f(T^{j-1}(x)) d\mu_B(x) = \\ &= \int_{G_K} \frac{1}{K} \sum_{i=1}^{I(x)} \sum_{j=n_i}^{m(n_i)} f(T^{j-1}(x)) d\mu_B(x) + \int_{(B \cap D) \setminus G_K} \frac{1}{K} \sum_{j=1}^K f(T^{j-1}(x)) d\mu_B(x) + \\ &\quad + \int_{G_K} \frac{1}{K} \sum_{j \in [1, K-L] \setminus \bigcup[n_i, m(n_i)]} f(T^{j-1}(x)) d\mu_B(x) + \\ &\quad + \int_{G_K} \frac{1}{K} \sum_{j \in [K-L, K] \setminus \bigcup[n_i, m(n_i)]} f(T^{j-1}(x)) d\mu_B(x). \end{aligned}$$

Innanzitutto notiamo che, per il lemma di impacchettamento e poiché  $x \in B \cap D$ , si ha

$$\begin{aligned} \int_{G_K} \frac{1}{K} \sum_{i=1}^{I(x)} \sum_{j=n_i}^{m(n_i)} f(T^{j-1}(x)) d\mu_B(x) &\geq \int_{G_K} \frac{1}{K} \sum_{i=1}^{I(x)} (m(n_i) - n_i + 1) \alpha d\mu_B(x) \geq \frac{1}{K} \int_{G_K} (1 - 2\delta) K \alpha d\mu_B(x) \\ &= \frac{1}{K} (1 - 2\delta) K \alpha \mu_B(G_K) = (1 - 2\delta) \alpha \mu_B(G_K). \end{aligned}$$

Sia ora

$$I_1 := \int_{(B \cap D) \setminus G_K} \frac{1}{K} \sum_{j=1}^K f(T^{j-1}(x)) d\mu_B(x).$$

Osserviamo che

$$\begin{aligned} |I_1| &\leq \frac{1}{K} \int_{(B \cap D) \setminus G_K} \sum_{j=1}^K |f(T^{j-1}(x))| d\mu_B(x) = \\ &= \frac{1}{K} \sum_{j=1}^K \int_{(B \cap D) \setminus G_K} |f(T^{j-1}(x))| d\mu_B(x) = \frac{1}{K} \sum_{j=1}^K \int_{T^{1-j}((B \cap D) \setminus G_K)} |f(x)| d\mu_B(x). \end{aligned}$$

Osserviamo che, poiché  $f \in L^1$ , allora  $|I_1| < +\infty$ . Inoltre notiamo che, poiché  $\mu(G_K) \geq 1 - \delta$ , allora

$$\mu_B((B \cap D) \setminus G_K) = \mu_B(B \cap D) - \mu_B(G_K) \leq 1 - (1 - \delta) = \delta.$$

Essendo per ipotesi  $T$  stazionaria, dunque, si ha che

$$\mu(T^{1-j}((B \cap D) \setminus G_K)) = \mu((B \cap D) \setminus G_K) \leq \delta.$$

Dunque gli insiemi  $T^{1-j}((B \cap D) \setminus G_K)$ , al variare di  $j \in \{1, \dots, K\}$ , hanno la stessa misura, la quale è limitata superiormente da  $\delta$ . Ricordiamo che la scelta di  $\delta > 0$ , da cui dipende la definizione dell'insieme  $G_K$ , è arbitraria. Dunque, per il Lemma 6, scegliamo  $\delta > 0$  tale che, se  $\mu((B \cap D) \setminus G_k) < \delta$ , allora è possibile stimare  $|I_1|$  con una quantità arbitrariamente piccola. Sia ora

$$I_2 := \int_{G_K} \frac{1}{K} \sum_{j \in [1, K-L] \setminus \bigcup_{[n_i, m(n_i)]}} f(T^{j-1}(x)) d\mu_B(x).$$

Osserviamo che, per il lemma di impacchettamento, se  $j \in [1, K-L] \setminus \bigcup_{[n_i, m(n_i)]}$ , allora  $m(j) - j + 1 > L$ . Sotto queste ipotesi, dunque, se  $x \in G_K \subset B \cap D$ , allora  $T^{j-1}(x) \in E \cap D$ , per come abbiamo definito l'insieme  $E$ . Dunque, ricordando che  $T$  è stazionaria, si ha che

$$\begin{aligned} |I_2| &\leq \frac{1}{K} \int_{G_K} \sum_{j \in [1, K-L] \setminus \bigcup_{[n_i, m(n_i)]}} |f(T^{j-1}(x))| d\mu_B(x) = \\ &= \frac{1}{K} \int_{B \cap D} \sum_{j=1}^K \chi_{E \cap D}(T^{j-1}(x)) |f(T^{j-1}(x))| d\mu_B(x) = \\ &= \frac{1}{K} \sum_{j=1}^K \int_{B \cap D} \chi_{E \cap D}(T^{j-1}(x)) |f(T^{j-1}(x))| d\mu_B(x) = \\ &= \frac{1}{K} K \int_{B \cap D} \chi_{E \cap D}(x) |f(x)| d\mu_B(x) = \int_{E \cap D} |f(x)| d\mu_B(x). \end{aligned}$$

Osserviamo che, in questo caso, all'aumentare di  $L$ ,  $\mu_B(E) = \mu_B(E \cap D)$  è una quantità sempre minore, dunque per  $L$  grande,  $|I_2|$  è arbitrariamente piccolo.

Sia ora

$$I_3 := \int_{G_K} \frac{1}{K} \sum_{j \in [K-L, K] \setminus \cup_{[n_i, m(n_i)]}} f(T^{j-1}(x)) d\mu_B(x).$$

Tale valore può essere facilmente maggiorato se osserviamo che, poichè  $G_K \subset B \cap D$

$$\begin{aligned} |I_3| &\leq \frac{1}{K} \sum_{j \in (K-L, K]} \int_{G_K} |f(T^{j-1}(x))| d\mu_B(x) \leq \\ &\leq \frac{1}{K} \sum_{j \in (K-L, K]} \int_{B \cap D} |f(T^{j-1}(x))| d\mu_B(x) = \\ &= \frac{1}{K} \sum_{j \in (K-L, K]} \int_{B \cap D} |f(x)| d\mu_B(x) \leq \\ &\leq \frac{1}{K} L \int_{B \cap D} |f(x)| d\mu_B(x). \end{aligned}$$

Dunque, per  $L$  fissato, se  $K$  è sufficientemente grande,  $I_3$  è piccolo. Ora ricordiamo che vale la relazione

$$\begin{aligned} \int_{B \cap D} f(x) d\mu_B(x) &\geq (1 - 2\delta)\alpha \mu_B(G_K) + I_1 + I_2 + I_3 \\ &\geq (1 - 2\delta)\alpha(1 - \delta) + I_1 + I_2 + I_3 \end{aligned}$$

Ora fissiamo  $\epsilon > 0$ . In base a tale valore ed alle stime ottenute, pertanto, siano :

- $\delta_0$  tale che,  $\forall \delta \leq \delta_0$ ,  $(1 - 2\delta)\alpha(1 - \delta) > \alpha - \frac{\epsilon}{4}$ ;
- $\delta_1$  tale che  $\delta_1 < \delta_0$  e che  $|I_1| < \frac{\epsilon}{4}$ ;
- $L_0$  sufficientemente grande e tale che  $|I_2| < \frac{\epsilon}{4}$ ;
- $K_0$  sufficientemente grande e tale che  $\frac{L_0}{K_0} < \delta_1$  e  $|I_3| < \frac{\epsilon}{4}$ .

Per tali valori di  $\delta_1, L_0, K_0$ , dunque, otteniamo che

$$\int_{B \cap D} f(x) d\mu_B(x) \geq \alpha - \frac{\epsilon}{4} - \frac{\epsilon}{4} - \frac{\epsilon}{4} - \frac{\epsilon}{4} = \alpha - \epsilon.$$

Ma questo significa che

$$\int_{B \cap D} f(x) d\mu_B(x) \geq \alpha - \epsilon, \quad \forall \epsilon > 0.$$

Per  $\epsilon \rightarrow 0$ , dunque, e per la permanenza del segno otteniamo che

$$\int_{B \cap D} f(x) d\mu_B(x) \geq \alpha,$$

quindi possiamo concludere. □

Ora disponiamo di tutti gli strumenti necessari per dimostrare il teorema ergodico, di cui riportiamo l'enunciato.

**Teorema 5** (Teorema ergodico). *Sia  $T$  una trasformazione stazionaria su uno spazio di probabilità  $(X, \Sigma, \mu)$  e sia  $f$  una funzione integrabile, allora la media  $\frac{1}{n} \sum_{i=1}^n f(T^{i-1}x)$  converge quasi certamente ed in  $L^1$ , per  $n \rightarrow \infty$ , verso una funzione  $T$ -invariante  $f^*(x)$ .*

*Dimostrazione.* Ricordiamo che affermare che la successione  $\frac{1}{n} \sum_{i=1}^n f(T^{i-1}x)$  sia convergente per  $n \rightarrow \infty$  equivale ad affermare che

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x).$$

Osserviamo, innanzitutto, che il risultato del lemma precedente rimane vero per qualsiasi sottoinsieme  $C \cap D \subseteq B \cap D$ , purchè tale insieme sia  $T$ -invariante: si può notare, infatti, che tutti i passaggi utilizzati nel suddetto lemma rimangono validi qualora si restrinja l'insieme  $B \cap D$  all'insieme  $C \cap D$ . Definiamo pertanto l'insieme

$$C := \left\{ x : \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) < \alpha < \beta < \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) \right\}.$$

Si noti che  $C \cap D$  è  $T$ -invariante. In base al lemma precedente, dunque, osserviamo innanzitutto che

$$\int_{C \cap D} f(x) d\mu(x) \geq \beta \mu(C)$$

Notiamo inoltre che

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) < \alpha &\iff -\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) > -\alpha \\ &\iff \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (-f(T^{i-1}x)) > -\alpha. \end{aligned}$$

Sempre in virtù del lemma precedente, quindi, concludiamo che

$$-\int_C f(x) d\mu(x) \geq -\alpha \mu(C) \iff \int_C f(x) d\mu(x) \leq \alpha \mu(C).$$

Dunque

$$\beta \mu(C) \leq \int_C f(x) d\mu(x) \leq \alpha \mu(C).$$

Siccome  $\alpha < \beta$ , pertanto, la diseguaglianza ottenuta può essere vera se e soltanto se  $\mu(C)=0$ . Osserviamo che l'insieme  $C$ , per come è definito, dipende dai parametri  $\alpha$  e  $\beta$ , dunque sia  $C = C_{\alpha,\beta}$ . Notiamo, sempre in virtù della definizione di  $C$ , che vale la seguente inclusione:

$$\left\{ x : \exists \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^{i-1}x) \right\}^c \subseteq \bigcup_{\alpha, \beta \in \mathbb{Q}, \alpha < \beta} C_{\alpha, \beta}.$$

Dunque possiamo concludere che la successione  $\frac{1}{n} \sum_{i=1}^n f(T^{i-1}x)$  converge quasi certamente per  $n \rightarrow \infty$ .

Ora dobbiamo provare che la successione converge in  $L^1$ . Definiamo, innanzitutto, l'operatore  $f \mapsto U_T(f)$  in questo modo:

$$U_T(f)(x) := f(Tx), \quad \forall x \in X, \quad f \in L^1.$$

Tale operatore è lineare, infatti, date  $f$  e  $g$  in  $L^1$ , per ogni  $x \in X$  si ha che

$$U_T(f+g)(x) = (f+g)(Tx) = f(Tx) + g(Tx) = U_T(f)(x) + U_T(g)(x).$$

Inoltre osserviamo che, essendo  $T$  stazionaria,

$$\|U_T f\|_{L^1} = \int |f(Tx)| d\mu(x) = \int |f(x)| d\mu(x) = \|f\|_{L^1}.$$

Dunque, se definiamo,  $\forall x \in X$ , la media temporale

$$A_n f(x) := \frac{1}{n} \sum_{i=1}^n U_T^{i-1} f(x)$$

osserviamo che la linearità di  $A_n$  segue facilmente dalla linearità dell'operatore  $U_T$ . Inoltre, notiamo che

$$\begin{aligned} \|A_n f\|_{L^1} &= \frac{1}{n} \int \left| \sum_{i=1}^n U_T^{i-1} f(x) \right| d\mu(x) \leq \frac{1}{n} \sum_{i=1}^n \int |U_T^{i-1} f(x)| d\mu(x) \\ &= \frac{1}{n} \sum_{i=1}^n \int |f(x)| d\mu(x) = \|f\|_{L^1}. \end{aligned} \quad (1.12)$$

Supponiamo  $f$  sia limitata, cioè esista  $M > 0$  tale che  $|f(x)| \leq M$ . Osserviamo che, allora,

$$|A_n f(x)| \leq \frac{1}{n} \sum_{i=1}^n |U_T^{i-1} f(x)| = \frac{1}{n} \sum_{i=1}^n |f(T^{i-1}x)| \leq \frac{1}{n} nM = M.$$

Siccome abbiamo provato che  $A_n f$  converge quasi certamente, dunque, poichè  $|A_n f|$  è limitata, per il teorema di convergenza dominata si ha che  $A_n f$  converge in  $L^1$ . Dunque il teorema è provato se  $f$  è limitata.

Consideriamo ora il caso generale: sia  $f \in L^1$ , in generale non limitata. Dato  $\epsilon > 0$ , dunque, sia  $g_\epsilon$  un'approssimazione limitata di  $f$ , tale che  $\|f - g_\epsilon\| \leq \frac{\epsilon}{3}$  (si può scegliere, ad esempio,  $g_\epsilon(x) = f(x)\chi_{\{|f(x)| \leq M\}}$ , prendendo  $M = M_\epsilon$  sufficientemente grande). Ora osserviamo che

$$\|f - A_n f\|_{L^1} \leq \|f - g_\epsilon\|_{L^1} + \|g_\epsilon - A_n g_\epsilon\|_{L^1} + \|A_n g_\epsilon - A_n f\|_{L^1}.$$

Ora notiamo che, in virtù di quanto osservato nella (1.12),

$$\|A_n g_\epsilon - A_n f\|_{L^1} = \|A_n(g_\epsilon - f)\|_{L^1} \leq \|g_\epsilon - f\|_{L^1}.$$

Osserviamo inoltre che, poichè, essendo  $g$  limitata,  $A_n g \rightarrow g$  in  $L^1$ , allora, per  $\epsilon > 0$ , esiste  $n_\epsilon \in \mathbb{N}$  tale che  $\|g - A_n g\|_{L^1} \leq \frac{\epsilon}{3}$ ,  $\forall n \geq n_\epsilon$ . Dunque possiamo concludere che, fissato  $\epsilon > 0$ , per ogni  $n \geq n_\epsilon$  si ha che

$$\|f - A_n f\| \leq 2\|g - f\|_{L^1} + \|A_n g - g\|_{L^1} \leq 2\frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

ovvero  $A_n f \rightarrow f$  in  $L^1$ . □

## Capitolo 2

# Applicazioni ed esempi

In questo capitolo utilizzeremo la teoria sviluppata finora al fine di dimostrare l'ergodicità di alcuni particolari processi stocastici.

### 2.1 Strumenti tecnici

In questa sezione richiamiamo alcuni concetti di calcolo delle probabilità necessari al nostro scopo. Sia, innanzitutto,  $(X, \Sigma, \mu)$  uno spazio di misura e consideriamo un'applicazione  $T : X \rightarrow X$ . Tale funzione è detta *mescolante* se

$$\lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D) = \mu(C)\mu(D), \quad \forall C, D \in \Sigma. \quad (2.1)$$

Un processo stocastico, dunque, si dice *mescolante* se l'operazione di shift è mescolante rispetto alla misura di Kolmogorov associata a tale processo.

**Lemma 11.** *Sia  $(X, \Sigma, \mu)$  uno spazio di misura e sia  $T : X \rightarrow X$  un'applicazione. Se  $T$  è mescolante, allora  $T$  è ergodica.*

*Dimostrazione.* Sia  $C \in \Sigma$  tale che  $T^{-1}C = C$ : osserviamo che, poiché  $C = T^{-1}C = \dots = T^{-n}C = \dots$ , allora  $T^{-n}C \cap D = C \cap D, \forall D \in \Sigma, \forall n \in \mathbb{N}$ . Essendo  $T$  mescolante, quindi, in virtù della (2.9) otteniamo che  $\mu(C \cap D) = \mu(C)\mu(D), \forall D \in \Sigma$ . Se poniamo  $C = D$ , quindi, otteniamo che  $\mu(C) = \mu(C)^2$  e questo è vero se e soltanto se  $\mu(C) = 0$  oppure  $\mu(C) = 1$ .  $\square$

Introduciamo ora il concetto di *mescolanza debole*. Un'applicazione  $T : X \rightarrow X$  è detta *debolmente mescolante* se

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mu(T^{-j}C \cap D) = \mu(C)\mu(D), \quad \forall C, D \in \Sigma.$$

Vediamo che, in questo caso, la condizione di mescolanza debole non solo implica la condizione di ergodicità, ma è anche del tutto equivalente a quest'ultima.

**Lemma 12.** *Sia  $(X, \Sigma, \mu)$  uno spazio di misura. Allora  $T : X \rightarrow X$  una trasformazione debolmente mescolante  $\iff T$  è ergodica.*

*Dimostrazione.* Consideriamo separatamente le due implicazioni.

[ $\implies$ ] Sia  $C \in \Sigma$  tale che  $T^{-1}C = C$ , osserviamo che allora  $C = T^{-1}C = \dots = T^{-n}C, \dots$ , dunque la condizione di mescolanza debole implica che

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mu(T^{-j}C \cap D) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mu(C \cap D) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} N \mu(C \cap D) = \mu(C \cap D) = \mu(C)\mu(D) \quad \forall D \in \Sigma. \end{aligned}$$

Se, quindi, poniamo  $D = C$ , otteniamo che  $\mu(C) = \mu(C)^2$ , ovvero  $\mu(C) = 0$  oppure  $\mu(C) = 1$ .

[ $\impliedby$ ] Supponiamo che  $T$  sia una trasformazione ergodica. In virtù del teorema ergodico, quindi, si ha che per ogni funzione integrabile  $f$  deve vale

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n f(T^j x) = \int f d\mu,$$

quasi certamente ed in  $L^1$ . Ora siano  $C, D \in \Sigma$  e poniamo  $f(x) = \chi_C(x)$ . In virtù dell'osservazione precedente si ha che

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \chi_C(T^j x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \chi_{T^{-j}C}(x) = \int \chi_C(x) d\mu = \mu(C), \quad (2.2)$$

quasi certamente ed in  $L^1$ . Osserviamo ora che

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N \mu(T^{-j}C \cap D) &= \frac{1}{N} \sum_{j=1}^N \int \chi_{T^{-j}C \cap D}(x) d\mu(x) = \int \frac{1}{N} \sum_{j=1}^N \chi_{T^{-j}C \cap D}(x) d\mu(x) \\ &= \int \frac{1}{N} \sum_{j=1}^N \chi_{T^{-j}C}(x) \chi_D(x) d\mu(x). \end{aligned}$$

Ora, in base alla (2.2) ed osservando che, ovviamente,  $|\chi_{T^{-j}C}(x)| \leq 1, \forall j$ , e che, quindi,  $\frac{1}{N} \sum_{j=1}^N \chi_{T^{-j}C}(x) \chi_D(x)$  è limitata, in base al teorema di convergenza dominata si ha che

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mu(T^{-j}C \cap D) \chi_D(x) &= \lim_{N \rightarrow \infty} \int_D \frac{1}{N} \sum_{j=1}^N \chi_{T^{-j}C}(x) d\mu(x) \\ &= \int_D \mu(C) d\mu(x) \chi_D(x) = \mu(C)\mu(D). \end{aligned}$$

Dunque abbiamo provato che, se  $T$  è una trasformazione ergodica, allora è debolmente mescolante.  $\square$

A questo punto, disponiamo di ulteriori strumenti che possono essere utili nel determinare se una data applicazione sia ergodica (e, nel caso della mescolanza debole, possiamo anche verificare che non lo sia). Ai fini pratici, tuttavia, la verifica delle proprietà di mescolanza e di mescolanza debole può essere complicata, dato che deve valere per una coppia generica di elementi appartenenti alla  $\sigma$ -algebra che stiamo considerando. A tal proposito, un aiuto notevole per la verifica di tali proprietà è fornito da un risultato, la cui dimostrazione richiede la conoscenza di alcune definizioni di base e dell'enunciato del lemma di Dynkin.

**Definizione 6** (Base di una  $\sigma$ -algebra). Sia  $X$  un insieme e sia  $\Sigma$  una  $\sigma$ -algebra. Un sottoinsieme  $\mathcal{I} \subseteq \Sigma$  è detto *base di  $\Sigma$*  se

- $\Sigma$  è generata da  $\mathcal{I}$
- $A \cap B \in \mathcal{I}, \forall A, B \in \mathcal{I}$

**Definizione 7.** Sia  $X$  un insieme. Un sottoinsieme  $\mathcal{D} \subseteq X$  è detto *classe monotona su  $X$*  se soddisfa le seguenti condizioni:

- $X \in \mathcal{D}$ .
- Se  $A, B \in \mathcal{D}$  ed  $A \subseteq B$ , allora  $B \setminus A \in \mathcal{D}$ .
- Se  $\{A_n\}_{n \in \mathbb{N}}$  è una successione di elementi di  $\mathcal{D}$  tale che  $A_n \subseteq A_{n+1}, \forall n \geq 1$ , allora  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}$ .

Ora che disponiamo di queste definizioni, possiamo procedere enunciando il seguente lemma.

**Lemma 13** (Lemma di Dynkin). *Sia  $X$  un insieme, sia  $\Sigma$  una  $\sigma$ -algebra su  $X$  e sia  $\mathcal{I}$  una base di  $\Sigma$ . Ogni classe monotona  $\mathcal{D}$  su  $X$  contenente  $\mathcal{I}$  contiene anche  $\Sigma$ .*

*Dimostrazione.* Si veda, ad esempio, [2, §A1.3 (pag. 193)]. □

Ora, grazie a questi strumenti, possiamo enunciare e dimostrare un teorema che risulterà molto utile al fine di dimostrare l'ergodicità di alcuni tipi di processi stocastici.

**Teorema 6.** *Sia  $(X, \Sigma, \mu)$  uno spazio di misura e sia  $T : X \rightarrow X$  un'applicazione stazionaria. Se la condizione di mescolanza (rispettivamente di mescolanza debole) è soddisfatta per ogni  $C, D$  appartenenti ad una base di  $\Sigma$ , allora tale condizione (rispettivamente la condizione di mescolanza debole) è valida per ogni  $C, D \in \Sigma$ .*

*Dimostrazione.* Dimostriamo il teorema relativamente alla proprietà di mescolanza: la proprietà di mescolanza debole si tratta in modo analogo.

Sia, dunque,  $\mathcal{I}$  una base di  $\Sigma$ : dobbiamo provare che se, per ogni  $C, D \in \mathcal{I}$ ,

$$\lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D) = \mu(C)\mu(D), \quad (2.3)$$

allora tale proprietà è valida per ogni  $C, D \in \Sigma$ .

Innanzitutto dimostriamo che la (2.3) vale per ogni coppia  $C, D$ , con  $C \in \mathcal{I}$  e  $D \in \Sigma$ .

Sia, pertanto,  $C \in \mathcal{I}$  fissato e definiamo l'insieme

$$\Gamma_C := \{D \in \Sigma : \lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D) = \mu(C)\mu(D)\}.$$

Dimostriamo che  $\Gamma_C$  è una classe monotona. Chiaramente  $X \in \Gamma_C$ , poichè  $T$  è stazionaria, dunque

$$\lim_{n \rightarrow \infty} \mu(T^{-n}C \cap X) = \lim_{n \rightarrow \infty} \mu(T^{-n}C) = \lim_{n \rightarrow \infty} \mu(C) = \mu(C) = \mu(C)\mu(X).$$

Innanzitutto, siano  $D_1, D_2 \in \Gamma_C$  tali che  $D_1 \subseteq D_2$ , osserviamo che, dalla definizione di  $\Gamma_C$ , segue che

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu(T^{-n}C \cap (D_2 \setminus D_1)) &= \lim_{n \rightarrow \infty} \mu((T^{-n}C \cap D_2) \setminus (T^{-n}C \cap D_1)) \\ &= \lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D_2) - \lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D_1) = \mu(C)\mu(D_2) - \mu(C)\mu(D_1) \\ &= \mu(C)(\mu(D_2) - \mu(D_1)) = \mu(C)\mu(D_2 \setminus D_1) \end{aligned}$$

dunque  $D_2 \setminus D_1 \in \Gamma_C$ .

Sia, ora,  $\{D_m\}_{m \in \mathbb{N}}$  una successione di elementi di  $\Gamma_C$  tale che  $D_m \subseteq D_{m+1}$ ,  $\forall n \geq 1$  e sia  $D := \bigcup_{m \in \mathbb{N}} D_m$ : proviamo che  $D \in \Gamma_C$ . Ricordiamo che, per la continuità dal basso, si ha che  $\lim_{m \rightarrow \infty} \mu(D_m) = \mu(D)$ . Sia, dunque,  $\epsilon > 0$  e sia  $m_\epsilon \in \mathbb{N}$  tale che  $|\mu(D) - \mu(D_{m_\epsilon})| < \epsilon$ . Osserviamo che, per ogni  $n \in \mathbb{N}$ ,

$$\begin{aligned} |\mu(T^{-n}C \cap D) - \mu(T^{-n}C \cap D_{m_\epsilon})| &= |\mu(T^{-n}C \cap (D \setminus D_{m_\epsilon}))| \\ &\leq \mu(D \setminus D_{m_\epsilon}) = \mu(D) - \mu(D_{m_\epsilon}) < \epsilon \end{aligned}$$

dunque

$$\mu(T^{-n}C \cap D_{m_\epsilon}) - \epsilon \leq \mu(T^{-n}C \cap D) \leq \mu(T^{-n}C \cap D_{m_\epsilon}) + \epsilon.$$

Osserviamo che, poichè  $D_{m_\epsilon} \in \Gamma_C$ , per ipotesi

$$\lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D_{m_\epsilon}) = \mu(C)\mu(D_{m_\epsilon}),$$

dunque

$$\limsup_{n \rightarrow \infty} \mu(T^{-n}C \cap D) \leq \mu(C)\mu(D_{m_\epsilon}) + \epsilon \leq \mu(C)\mu(D) + 2\epsilon$$

$$\liminf_{n \rightarrow \infty} \mu(T^{-n}C \cap D) \geq \mu(C)\mu(D_{m_\epsilon}) - \epsilon \geq \mu(C)\mu(D) - 2\epsilon.$$

Questo fatto vale per ogni  $\epsilon > 0$  (e per il corrispondente  $n_\epsilon$ ), dunque, facendo tendere  $\epsilon$  a 0, otteniamo che

$$\mu(C)\mu(D) \leq \liminf_{n \rightarrow \infty} \mu(T^{-n}C \cap D) \leq \limsup_{n \rightarrow \infty} \mu(T^{-n}C \cap D) \leq \mu(C)\mu(D).$$

Quindi esiste  $\lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D) = \mu(D)\mu(C)$  e dunque  $D \in \Gamma_C$ . Dunque abbiamo provato che  $\Gamma_C$  è una classe monotòna: in virtù del lemma di Dynkin, dunque,  $\Gamma_C \supseteq \Sigma$ , ovvero la relazione (2.3) è soddisfatta per ogni  $C \in \mathcal{I}$  e per ogni  $D \in \Sigma$ . Viceversa, se, fissato  $D \in \Sigma$ , definiamo l'insieme

$$\tilde{\Gamma}_D := \{C \in \Sigma : \lim_{n \rightarrow \infty} \mu(T^{-n}C \cap D) = \mu(C)\mu(D)\}.$$

Si noti che  $\tilde{\Gamma}_D \supseteq \mathcal{I}$ . Si dimostra in maniera simile che  $\tilde{\Gamma}_D$  è una classe monotòna: basta osservare che, per ogni  $n \in \mathbb{N}$ , l'applicazione  $T^{-n}$  preserva le relazioni insiemistiche ed è chiusa rispetto all'unione ed all'intersezione, dunque si procede analogamente a quanto fatto prima. Quindi  $\tilde{\Gamma}_D \supseteq \Sigma$  e la relazione (2.3) è valida per ogni coppia  $C, D$  tale che  $C \in \Sigma$  e  $D \in \Sigma$ .  $\square$

Ora che disponiamo di tutti gli strumenti necessari, possiamo procedere con alcuni esempi concreti di processi ergodici.

## 2.2 Processi stazionari

Sia  $(X, \Sigma, P)$  uno spazio di probabilità e sia  $\{X_n\}_{n \in \mathbb{N}}$  un processo stocastico a valori in uno spazio di misurabile  $(E, \mathcal{E})$ , ossia tale che

$$X_n : (X, \Sigma, P) \rightarrow (E, \mathcal{E}), \quad \forall n \in \mathbb{N}.$$

Supponiamo che tale sequenza sia costituita da variabili casuali indipendenti ed identicamente distribuite, ovvero:

$$\begin{aligned} P(X_n \in A_n | X_{n-1} \in A_{n-1}, \dots, X_1 \in A_1) &= P(X_n \in A_n) \cdot \dots \cdot P(X_1 \in A_1), \\ \forall n \geq 1, \forall A_1, \dots, A_n \in \mathcal{E} \end{aligned}$$

Notiamo che un processo stocastico i.i.d. è stazionario, infatti, dati  $A_m, A_{m+1}, \dots, A_n \in \mathcal{E}$ , ove  $m, n \in \mathbb{N}$ ,  $m < n$ , si ha che

$$\begin{aligned} P(X_m \in A_m, X_{m+1} \in A_{m+1}, \dots, X_n \in A_n) \\ &= P(X_m \in A_m) \cdot P(X_{m+1} \in A_{m+1}) \cdot \dots \cdot P(X_1 \in A_1) \\ &= P(X_{m+1} \in A_m) \cdot P(X_{m+2} \in A_{m+1}) \cdot \dots \cdot P(X_2 \in A_1) \\ &= P(X_{m+1} \in A_m, X_{m+2} \in A_{m+1}, \dots, X_2 \in A_1). \end{aligned}$$

Consideriamo ora l'operazione di shift definita sullo spazio prodotto  $E^n$ , munito della  $\sigma$ -algebra prodotto  $\mathcal{E}^n$

$$\begin{array}{ccc} T : & E^\mathbb{N} & \rightarrow E^\mathbb{N} \\ & (x_1, x_2, \dots, x_n, \dots) & \mapsto (x_2, x_3, \dots, x_n, \dots). \end{array}$$

Il nostro obiettivo è dimostrare che la trasformazione di shift è ergodica rispetto alla misura di Kolmogorov associata al processo  $\{X_n\}_{n \in \mathbb{N}}$ . Innanzitutto, proviamo a vedere

se tale operazione gode della proprietà di mescolanza. Si dimostra che una base della  $\sigma$ -algebra prodotto  $\mathcal{E}^n$  è data dall'insieme costituito dagli elementi del tipo

$$A_1 \times \dots \times A_n := \{(x_1, \dots, x_n, \dots) : x_m \in A_m, \dots, x_n \in A_n\},$$

ove  $m, n \in \mathbb{N}, m \leq n, A_m, \dots, A_n \in \mathcal{E}$

(a tal proposito, si legga [1, pag.2]. Siano, quindi

$$\begin{aligned} C &:= A_1 \times \dots \times A_n \\ D &:= B_1 \times \dots \times B_n. \end{aligned}$$

Osserviamo che, dato  $N \in \mathbb{N}$  tale che  $N > n$ ,

$$T^{-N}C = \{(x_1, \dots, x_n, \dots) : x_{N+1} \in A_1, \dots, x_{N+n} \in A_n\}.$$

Dunque

$$T^{-N}C \cap D = \{(x_1, \dots, x_n, \dots) : x_1 \in B_1, \dots, x_n \in B_n, x_{N+1} \in A_1, \dots, x_{N+n} \in A_n\},$$

pertanto, in virtù dell'indipendenza delle  $X_n$  e dalla proprietà di stazionarietà del processo stocastico

$$\begin{aligned} \mu(T^{-N}C \cap D) &= P(X_1 \in B_1, \dots, X_n \in B_n, X_{N+1} \in A_1, \dots, X_{N+n} \in A_n) \\ &= P(X_1 \in B_1, \dots, X_n \in B_n) \cdot P(X_{N+1} \in A_1, \dots, X_{N+n} \in A_n) \\ &= \mu(D)\mu(T^{-N}C) = \mu(D)\mu(C) \end{aligned}$$

Dunque la proprietà di mescolanza è soddisfatta e, pertanto, le successioni di variabili casuali indipendenti ed identicamente distribuite sono processi ergodici.

### 2.3 Catene di Markov

Sia  $(X, \Sigma, P)$  uno spazio di probabilità e sia  $A$  un insieme di cardinalità finita.

Consideriamo un processo stocastico  $\{X_n\}_{n \in \mathbb{N}}$ , definito su  $(X, \Sigma, P)$  ed a valori in  $E$  e supponiamo che per ogni naturale  $n \geq 2$  e per ogni sequenza  $a_n, a_{n-1}, a_{n-2}, \dots, a_1$  di elementi di  $E$  (per cui entrambi i membri siano ben definiti) si abbia che

$$\begin{aligned} &P(X_n = a_n | X_{n-1} = a_{n-1}) \\ &= P(X_n = a_n | X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots, X_1 = a_1) : \end{aligned} \tag{2.4}$$

tale processo si dice *catena di Markov*.

Supponendo, inoltre, che il processo considerato sia tale che, comunque presi  $a, b \in E$ ,  $P(X_n = a | X_{n-1} = b)$  non dipenda dalla scelta di  $n \in \mathbb{N}$ , diciamo che la catena di Markov considerata è *omogenea*.

Sotto queste ipotesi, ricordiamo innanzitutto che  $E$  è un insieme finito, dunque possiamo supporre che, se  $|E| = n$ ,  $E = \{a_1, a_2, \dots, a_n\}$ ,  $n \in \mathbb{N}$ .

Definiamo, quindi, la matrice :

$$A = \begin{bmatrix} M_{a_1 a_1} & M_{a_1 a_2} & \dots & M_{a_1 a_n} \\ M_{a_2 a_1} & M_{a_2 a_2} & \dots & M_{a_2 a_n} \\ \vdots & \vdots & \vdots & \vdots \\ M_{a_n a_1} & \dots & \dots & M_{a_n a_n} \end{bmatrix} \quad (2.5)$$

tale che :

$$M_{a_i a_j} := P(X_n = a_j | X_{n-1} = a_i),$$

per ogni  $i, j \in \{1, 2, \dots, m-1, m\}$ .

Ricordiamo che, nelle ipotesi fatte, il valore di ciascun elemento della matrice è lo stesso al variare di  $n \in \mathbb{N}$ .

Osserviamo che la matrice, detta *matrice di transizione*, per come l'abbiamo definita, soddisfa le seguenti proprietà :

- $M_{a_i a_j} \in [0, 1], \forall a_i, a_j \in E;$
- $\sum_{a_j \in E} M_{a_i a_j} = \sum_{a_j \in E} P(X_n = a_j | X_{n-1} = a_i) = 1, \forall a_i \in E.$

Definiamo, ora, per ogni  $a \in E$ , la funzione

$$\mu_1(a) := P(X_1 = a) = P(\{\omega \in X : X_1(\omega) = a\}).$$

Il vettore così ottenuto

$$\mu_1 := (\mu_1(a_1), \mu_1(a_2), \dots, \mu_1(a_n))$$

è detto *distribuzione iniziale* del processo stocastico  $\{X_n\}_{n \in \mathbb{N}}$ .

La distribuzione iniziale si dice *stazionaria* se, per ogni  $a_i \in E$ ,

$$\mu_1(a_i) = P(X_1 = a_i) = P(X_j = a_i), \forall j. \quad (2.6)$$

Osserviamo che tale proprietà è soddisfatta se e solo se si ha che  $\mu_1 M = \mu_1$ .

Infatti, consideriamo a titolo d'esempio la prima entrata del vettore ottenuto moltiplicando  $\mu_1$  per  $M$ , il suo valore è dato da:

$$\begin{aligned}
(\mu_1 M)_{a_1} &= \mu_1(a_1)M_{a_1 a_1} + \mu_1(a_2)M_{a_2 a_1} + \mu_1(a_3)M_{a_3 a_1} + \dots + \mu_1(a_n)M_{a_n a_1} \\
&= P(X_1 = a_1)P(X_2 = a_1 | X_1 = a_1) + P(X_1 = a_2)P(X_2 = a_1 | X_1 = a_2) \\
&\quad + P(X_1 = a_3)P(X_2 = a_1 | X_1 = a_3) + \dots + P(X_1 = a_n)P(X_2 = a_1 | X_1 = a_n) \\
&= P(X_1 = a_1, X_2 = a_1) + P(X_1 = a_2, X_2 = a_1) \\
&\quad + P(X_1 = a_3, X_2 = a_1) + \dots + P(X_1 = a_n, X_2 = a_1) = P(X_2 = a_1) \quad (2.7)
\end{aligned}$$

Dunque se  $\mu_1 M = \mu_1$  si ha  $P(X_2 = a_1) = P(X_1 = a_1)$ . Dunque, è sufficiente tenere conto di questo risultato e della proprietà di omogeneità della catena di Markov per sostituire le occorrenze della variabile  $X_1$  con la variabile  $X_2$  e quelle della variabile  $X_2$  con la variabile  $X_3$ , ottenendo  $P(X_2 = a_1) = P(X_3 = a_1)$  ed iterare il procedimento fino ad ottenere  $P(X_1 = a_1) = P(X_j = a_1), \forall j \in \{1, \dots, n\}$ .

Lo stesso discorso può essere fatto per tutti gli altri valori del vettore risultante, cosicchè si può verificare direttamente che la (2.6) è equivalente a  $\mu_1 M = \mu_1$ .

Ora che abbiamo definito le proprietà di una catena di Markov, individuiamo, nella proposizione seguente, le condizioni necessarie e sufficienti affinchè tale processo sia stazionario.

**Proposizione 1.** Una catena di Markov  $\{X_n\}_{n \in \mathbb{N}}$ , le cui variabili sono definite su uno spazio di probabilità  $(X, \Sigma, P)$  ed assumono valori in un insieme finito  $E$ , è un processo stazionario se e solo se è omogenea e la sua distribuzione iniziale  $\mu_1$  è stazionaria.

*Dimostrazione.* Dimostriamo separatamente le due implicazioni:

$\Rightarrow$  Supponiamo che valga la condizione di stazionarietà, ossia, ricordiamo,

$$P(X_i = a_i, m \leq i \leq n) = P(X_{i+1} = a_i, m \leq i \leq n), \quad \forall m, n \in \mathbb{N}, \quad \forall a_m, \dots, a_n \in E.$$

Osserviamo che la validità dell'affermazione precedente implica che

$$\mu_1(a) = P(X_1 = a) = P(X_2 = a) = \dots = P(X_n = a), \quad \forall a \in E$$

e ciò, per quanto visto in precedenza, corrisponde esattamente ad affermare che la distribuzione iniziale è stazionaria.

Per quanto riguarda, invece, l'omogeneità della catena di Markov considerata, basta osservare che, dati  $n \in \mathbb{N}$  ed  $a, b \in E$ , la stazionarietà comporta che :

$$\begin{aligned}
P(X_n = a | X_{n-1} = b) &= \frac{P(X_n = a, X_{n-1} = b)}{P(X_{n-1} = b)} \\
&= \frac{P(X_{n+1} = a, X_n = b)}{P(X_n = b)} \\
&= P(X_{n+1} = a | X_n = b),
\end{aligned}$$

dunque, in generale,  $P(X_n = a | X_{n-1} = b)$  è indipendente da  $n$ .

$\Leftarrow$  Supponiamo che valga la proprietà di omogeneità e che la distribuzione iniziale sia stazionaria. Sotto queste ipotesi, dunque, è possibile osservare che, dato  $n \in \mathbb{N}$  ed  $a_1, \dots, a_n \in E$  :

$$\begin{aligned} P(X_i = a_i, m \leq i \leq n) &= \mu_1(a_1)P(X_2 = a_2 | X_1 = a_1) \cdot P(X_3 = a_3 | X_2 = a_2) \\ &\quad \cdot \dots \cdot P(X_n = a_n | X_{n-1} = a_{n-1}) \\ &= P(X_2 = a_1) \cdot P(X_3 = a_2 | X_2 = a_1) \cdot P(X_4 = a_3 | X_3 = a_2) \\ &\quad \cdot \dots \cdot P(X_{n+1} = a_n | X_n = a_{n-1}) \\ &= P(X_{i+1} = a_i, m \leq i \leq n) \end{aligned}$$

□

Sia ora  $\{X_n\}_{n \in \mathbb{N}}$  una catena di Markov omogenea ed avente distribuzione iniziale stazionaria e sia  $A \in M_{n \times n}$  la corrispondente matrice di transizione. Tale matrice è detta *irriducibile* se, per ogni coppia di indici  $(i, j)$ , ove  $i, j \in \{1, \dots, n\}$ , esiste una sequenza  $i_0, i_1, \dots, i_l$  di elementi appartenenti ad  $\{1, \dots, n\}$  tale che  $i_0 = i$ ,  $i_l = j$  ed

$$A_{i_m i_{m+1}} > 0, \text{ per } m = 0, 1, \dots, l-1.$$

Osserviamo che, in virtù delle ipotesi fatte e della definizione di Catena di Markov, supporre che valga tale condizione equivale ad affermare che, se ad un dato istante la catena di Markov assume il valore  $a_i$ , allora la probabilità che, in un istante futuro, la catena possa assumere il valore  $a_j$  è strettamente positiva. Infatti assumiamo che, dato  $k \in \mathbb{N}$ ,  $P(X_k = a_i) > 0$ , allora

$$\begin{aligned} P(X_{k+1} = a_{i_1}, X_k = a_i) &= P(X_{k+1} = a_{i_1} | X_k = a_i)P(X_k = a_i) \\ &= A_{i_0 i_1}P(X_k = a_i) > 0. \end{aligned}$$

Dunque

$$\begin{aligned} P(X_{k+2} = a_{i_2}, X_{k+1} = a_{i_1}, X_k = a_i) &= P(X_{k+2} = a_{i_2} | X_{k+1} = a_{i_1}, X_k = a_i)P(X_{k+1} = a_{i_1}, X_k = a_i) \\ &= P(X_{k+2} = a_{i_2} | X_{k+1} = a_{i_1})P(X_{k+1} = a_{i_1}, X_k = a_i) \\ &\quad A_{i_1 i_2}P(X_{k+1} = a_{i_1}, X_k = a_i) > 0 \end{aligned}$$

Iterando il procedimento e sfruttando la definizione di Catena di Markov, quindi, si ottiene che

$$P(X_{k+l} = a_j, X_{k+l-1} = a_{i_{l-1}}, \dots, X_k = a_i) > 0,$$

dunque  $P(X_{k+l} = a_j) > 0$ . Ora vogliamo dimostrare che la condizione di ergodicità di una catena di Markov omogenea ed avente distribuzione iniziale stazionaria equivale alla condizione di irriducibilità della corrispondente matrice di transizione.

**Teorema 7.** *Sia  $\{X_n\}_{n \in \mathbb{N}}$  una catena di Markov definita su uno spazio di probabilità  $(X, \Sigma, P)$  ed a valori in uno spazio finito  $E$ , omogenea ed avente distribuzione iniziale stazionaria e positiva in ogni punto di  $E$ . Allora tale catena è ergodica se e solo se la corrispondente matrice di transizione  $A$  è irriducibile.*

*Dimostrazione.* Dimostriamo entrambe le implicazioni separatamente.

⇐ Supponiamo innanzitutto che  $A$  sia irriducibile. Osserviamo che dalla condizione di irriducibilità seguono due fatti fondamentali di algebra lineare, che ci limitiamo ad elencare:

- Esiste, a meno di multipli, uno ed un solo vettore  $\pi$  tale che  $\pi A = \pi$ . Notiamo che, poichè abbiamo supposto che, se  $\mu_1$  è il vettore distribuzione iniziale,  $\mu_1 A = \mu_1$ , allora per l'unicità di  $\pi$  dev'essere per forza  $\pi = \mu_1$ .
- Vale la seguente uguaglianza:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} M^{k+1} = P, \quad (2.8)$$

ove  $P$  è la matrice avente tutte le colonne uguali al vettore  $\pi$ .

Proviamo che la catena di Markov soddisfa la condizione di debole mescolanza, ovvero, dati due insiemi misurabili

$$\begin{aligned} C &= [c_1^m] = \{(x_1, x_2, \dots, x_k, \dots) : x_1 = c_1, \dots, x_m = c_m\} \\ D &= [d_1^n] = \{(x_1, x_2, \dots, x_k, \dots) : x_1 = d_1, \dots, x_n = d_n\}, \end{aligned}$$

(ove  $m, n \in \mathbb{N}$  e  $c_1, \dots, c_m, d_1, \dots, d_n \in E$ ), dimostriamo che

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mu(T^{-j}C \cap D) = \mu(C)\mu(D). \quad (2.9)$$

Osserviamo che questo fatto equivale a dimostrare che

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=n}^{N+n-1} \mu(T^{-j}C \cap D) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu(T^{-k-n}C \cap D) = \mu(C)\mu(D)$$

Osserviamo che, dato  $k > 0$ ,

$$T^{-k-n}C = [b_{n+k+1}^{n+k+m}], \text{ ove } b_{n+k+i} = c_i, \quad 1 \leq i \leq m.$$

Ora notiamo che

$$\mu([d_1^n] \cap [b_{n+k+1}^{n+k+m}]) = P(X_1 = d_1, \dots, X_n = d_n, X_{n+k+1} = c_1, \dots, X_{n+k+m} = c_m).$$

Utilizzando, dunque, la definizione di catena di Markov e facendo dei semplici calcoli si può osservare che

$$\begin{aligned}
& \mu([d_1^n] \cap [b_{n+k+1}^{n+k+m}]) = P(X_1 = d_1, \dots, X_n = d_n, X_{n+k+1} = c_1, \dots, X_{n+k+m} = c_m) \\
&= \sum_{(d_{n+1}, \dots, d_{n+k})} P(X_1 = d_1, \dots, X_{n+1} = d_{n+1}, \dots, X_{n+k} = d_{n+k}, \dots, X_{n+k+m} = c_m) \\
&= \sum_{(d_{n+1}, \dots, d_{n+k})} (P(X_1 = d_1, \dots, X_n = d_n) \cdot P(X_{n+k+1} = c_1, \dots, X_{n+1} = d_{n+1} | X_n = d_n) \\
&\quad \cdot P(X_{n+k+m} = c_m, \dots, X_{n+k+2} = c_2 | X_{n+k+1} = c_1)) \\
&= P(X_1 = d_1, \dots, X_n = d_n) \cdot \sum_{(d_{n+1}, \dots, d_{n+k})} P(X_{n+k+1} = c_1, \dots, X_{n+1} = d_{n+1} | X_n = d_n) \\
&\quad \cdot P(X_{n+k+m} = c_m, \dots, X_{n+k+2} = c_2 | X_{n+k+1} = c_1)
\end{aligned}$$

ove  $d_{n+1}, \dots, d_{n+k} \in E$ . Ora notiamo che, sempre utilizzando la definizione di catena di Markov, possiamo facilmente osservare che

$$\begin{aligned}
U &:= P(X_1 = d_1, \dots, X_n = d_n) = \mu(D) \\
&= P(X_1 = d_1) \cdot P(X_2 = d_2 | X_1 = d_1) \cdot \dots \cdot P(X_n = d_n | X_{n-1} = d_{n-1}) \\
&= \mu_1(d_1) A_{d_1 d_2} \cdot \dots \cdot A_{d_{n-1} d_n} = \mu_1(d_1) \prod_{i=1}^{n-1} A_{d_i d_{i+1}}.
\end{aligned}$$

$$\begin{aligned}
V &:= \sum_{(d_{n+1}, \dots, d_{n+k})} P(X_{n+k+1} = c_1, X_{n+k} = d_{n+k}, \dots, X_{n+1} = d_{n+1} | X_n = d_n) \\
&= \sum_{(d_{n+1}, \dots, d_{n+k})} P(X_{n+1} = d_{n+1} | X_n = d_n) \cdot \dots \cdot P(X_{n+k+1} = c_1 | X_{n+k} = d_{n+k}) \\
&= \sum_{(d_{n+1}, \dots, d_{n+k})} \left( \prod_{i=n}^{n+k-1} A_{d_i d_{i+1}} \right) \cdot A_{d_{n+k} c_1}.
\end{aligned}$$

Osserviamo che, tramite alcune operazioni di algebra lineare, si può verificare che tale valore coincide con il termine in posizione  $(d_n, c_1)$  della matrice  $A^k$ , sia  $[A^k]_{d_n c_1}$ .

$$\begin{aligned}
W &:= P(X_{n+k+m} = c_m, \dots, X_{n+k+2} = c_2 | X_{n+k+1} = c_1) \\
&= P(X_{n+k+2} = c_2 | X_{n+k+1} = c_1) \cdot \dots \cdot P(X_{n+k+m} = c_m | X_{n+k+m-1} = c_{m-1}) \\
&= \prod_{i=1}^m A_{c_i c_{i+1}}
\end{aligned}$$

Dunque

$$\mu([d_1^n] \cap [b_{n+k+1}^{n+k+m}]) = \mu(D) [A^{k+1}]_{d_n c_1} \prod_{i=1}^m A_{c_i c_{i+1}}$$

Osserviamo che, in base a alla (2.8), si ha che

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N [A^{k+1}]_{d_n c_1} = \pi(c_1) = \mu(c_1),$$

dunque, in virtù della (2.8)

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu(T^{-k-n} D \cap C) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu([d_1^n] \cap [b_{n+k+1}^{n+k+m}]) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \mu(D) \sum_{k=0}^{N-1} [A^{k+1}]_{d_n c_1} \prod_{i=1}^m A_{c_i c_{i+1}} = \mu(D) \mu(c_1) \prod_{i=1}^m A_{c_i c_{i+1}} = \mu(D) \mu(C). \end{aligned}$$

Dunque, essendo la condizione espressa dalla (2.9) soddisfatta,  $T$  è una trasformazione ergodica.

$\Rightarrow$  Supponiamo che la catena di Markov sia ergodica. Supponiamo che  $\mu(a_j) > 0$ ,  $\forall j$  e sia  $P_i = \{x : x_1 = a_i\}$ . Osserviamo che, se  $B = T^{-1}P_j \cup T^{-2}P_j \cup \dots$ , allora  $P(B) > 0$  e  $T^{-1}B = T^{-2}P_j \cup T^{-3}P_j \cup \dots \subseteq B$ . In virtù dell'ipotesi di ergodicità, quindi,  $\mu(B) = 1$ , dunque  $\mu(B \cap P_j) > 0$ ,  $\forall j$ . Osserviamo che

$$\mu(B \cap P_i) = \mu\left(\bigcup_{n \in \mathbb{N}} (T^{-n}P_j \cap P_i)\right) > 0,$$

dunque  $\exists \bar{n} \in \mathbb{N}$  tale che  $\mu(T^{-\bar{n}}P_j \cap P_i) > 0$ . Osserviamo che

$$\begin{aligned} \mu(T^{-\bar{n}}P_j \cap P_i) &= \mu(\{(x_1, \dots, x_k, \dots) : x_{\bar{n}} = a_j, x_1 = a_i\}) \\ &= P(X_{\bar{n}} = a_j, X_1 = a_i) = P(X_{\bar{n}} = a_j | X_1 = a_i)P(X_1 = a_i) > 0 \end{aligned}$$

Dunque  $P(X_{\bar{n}} = a_j | X_1 = a_i) > 0$ , pertanto dev'esserci una sequenza  $i_0, i_1, \dots, i_{\bar{n}}$  tale che  $i_0 = i$ ,  $i_{\bar{n}} = j$  ed  $Ai_m i_{m+1} = P(X_{m+1} = a_{i_{m+1}} | X_m = a_{i_m}) > 0$ ,  $\forall m \in 0, \dots, \bar{n} - 1$ .

□

# Bibliografia

- [1] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*, American Mathematical Society (1996).
- [2] D. Williams, *Probability with Martingales*, Cambridge University Press (1991).