UNIVERSITÀ
DEGLI STUDI
FIRENZE

**Scuola di
Scienze Matematiche,
Fisiche e Naturali**
Corso di Laurea in
Matematica

Grandi deviazioni per catene di Markov

Relatore
Francesca Romana Nardi

Candidato
Davide Baldelli

Anno Accademico 2019/2020

Indice

Introduzione

L'obiettivo della tesi è presentare un teorema che descrive il comportamento asintotico delle catene di Markov per lo studio di eventi rari.

Le catene di Markov, o processi Markoviani, sono famiglie di variabili aleatorie che descrivono l'evoluzione temporale di un processo, per le quali la probabilità di transizione da uno stato ad un altro dipende solo dallo stato immediatamente precedente, e non da tutta la cronologia degli stati. La varietà di fenomeni di cui queste sono modelli è molto ampia.

La teoria delle grandi deviazioni è una branca della teoria delle probabilità che si occupa della descrizione di eventi rari. Facciamo un breve excursus per conoscere il terreno su cui si è sviluppata. Consideriamo una famiglia di variabili aleatorie (v.a.) indipendenti e identicamente distribuite (i.i.d.); nei comuni libri di testo di calcolo delle probabilità, esistono due risultati che descrivono il comportamento asintotico della media campionaria: la legge dei grandi numeri e il teorema del limite centrale. Il primo asserisce che la media campionaria delle v.a. converge al valore atteso di tutte le variabili. Il teorema del limite centrale, invece, è un risultato più generale e individua una distribuzione asintotica alla quale la somma di v.a. converge, quantificando la probabilità che la somma differisca dalla sua media di una quantità "normale". Per dirlo con la matematica, siano (X_i) v.a. i.i.d. di valore atteso μ e deviazione standard σ e sia $S_n = X_1 + ... + X_n$, allora la legge dei grandi numeri dice che

$$\lim_{n\to\infty} \frac{1}{n} S_n = \mu,$$

mentre, per il teorema del limite centrale

$$\frac{1}{\sigma\sqrt{n}}(S_n - n\mu) \xrightarrow[n\to\infty]{l} Z,$$

dove con Z si intende la distribuzione normale standard e la convergenza è in legge. Quindi, le deviazioni che rientrano nel campo dell'ultimo teorema sono nell'ordine di $\sqrt{n}$, mentre la teoria delle grandi deviazioni si occupa di deviazioni di ordine almeno n. Ad esempio, se consideriamo l'evento

$$\{S_n \geq (\mu + a)n\} \qquad a > 0,$$

il teorema del limite centrale implica che la sua probabilità tende a 0 quando $n \to \infty$. Il nostro scopo è quello di quantificare la "velocità" con cui questo accade. Nei casi che studieremo, come spesso succede, il decadimento è esponenziale in n, più precisamente:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq (\mu + a)n) = -I(a) < 0.$$

dove I è detta funzione di *rate* e quantifica la suddetta velocità. La sua individuazione è uno dei principali obiettivi della teoria. Il teorema che raggiunge questo scopo nel caso di v.a. i.i.d. è il teorema di Cramer, che qui abbiamo omesso. Nel secondo capitolo descriveremo, però, il teorema di Sanov, che ne è una generalizzazione, attraverso l'utilizzo di un altro funzionale delle (X_i): la misura empirica (che verrà introdotta nel primo capitolo).
La teoria delle grandi deviazioni ha svariate applicazioni. Ad esempio, in probabilità, considerando un cammino aleatorio in un ambiente aleatorio unidimensionale, lo studio della funzione di *rate* permette di arrivare a una descrizione dettagliata dell'andamento del cammino. Un'altra interessante applicazione coinvolge invece la conduzione del calore in un mezzo aleatorio, cioè con sorgenti e pozzi casuali. Un uso relativo alla meccanica statistica riguarda l'analisi di sistemi composti da un gran numero di particelle interagenti (come in termodinamica o in fluidodinamica). La descrizione può essere accostata ai teoremi sopracitati, in particolare il comportamento macroscopico corrisponde alla legge dei grandi numeri, la descrizione delle deviazioni "normali" rientra nel campo del teorema del limite centrale ma le fluttuazioni più rare sono oggetto della teoria delle grandi deviazioni, il cui studio diventa interessante, ad esempio, se si investiga il comportamento del sistema in scale temporali maggiori di quelle per cui uno studio macroscopico sia efficace.

Il teorema principale della tesi è quindi uno dei più importanti risultati della teoria delle grandi deviazioni nel caso in cui le variabili aleatorie X_i non siano indipendenti, ma rientrino nella classe delle catene di Markov. In quest' ultime l'evoluzione del processo nel tempo futuro è funzione solamente del presente e non di tutta la storia passata. Vedremo che, durante lo sviluppo della tesi, saranno necessarie nozioni di probabilità, teoria della misura, calcolo differenziale e topologia.

Capitolo 1

Il Teorema di grandi deviazioni per catene di Markov

Nel primo capitolo introduciamo i fondamenti matematici implicati nella formulazione del teorema di nostro interesse, per poi presentarne l'enunciato.

1.1 Alcune definizioni preliminari

Iniziamo dando la definizione di processo stocastico e di catena di Markov. Questo genere di processi aleatori saranno i protagonisti del teorema principale dell'elaborato. Successivamente presentiamo uno strumento matematico attraverso il quale vengono spesso studiate le grandi deviazioni: la misura empirica. Infine, diamo le definizioni su cui si fonda tutta la teoria delle grandi deviazioni e la maggior parte dei risultati presentati nella tesi.

Definizione (processo stocastico)**.** Si chiama *processo stocastico* una famiglia di variabili aleatorie (X_t) definite su uno stesso spazio di probabilità, dove t varia in un sottoinsieme $T \subset \mathbb{R}^+$

I processi stocastici sono quindi insiemi ordinati di variabili aleatorie che variano secondo un parametro reale (tipicamente il tempo).

Definizione (catena di Markov)**.** Si chiama *Catena di Markov* un processo stocastico tale che:

- $T = \mathbb{N}$ oppure T è un sottointervallo di $\mathbb{N}$.
- Le variabili aleatorie X_n assumono valori nello stesso insieme $\Gamma \subset \mathbb{N}$.
- per ogni $i, j \in \Gamma$ e per ogni $n \in T$ esistono dei numeri $P_{i,j}(n)$ tali che

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, ..., X_1 = i_1) = \\ = \mathbb{P}(X_{n+1} = j \mid X_n = i) = P_{i,j}(n).$$

L'insieme Γ si chiama l'insieme degli stati della catena. L'ultimo punto della definizione, noto come *proprietà di Markov*, è la caratteristica fondamentale delle catene di Markov: le v.a. (X_n) non sono indipendenti e la conoscenza del valore assunto da X_n dà delle informazioni su quello che sarà il valore di X_{n+1}, ma la conoscenza supplementare dei valori di $X_{n-1}, X_{n-2}...$ non fornisce informazioni ulteriori.
Per i nostri scopi aggiungeremo qualche ipotesi semplificativa. Inanzitutto, supponiamo che l'insieme degli stati sia finito, cioè $\Gamma = \{1, 2, ..., r\}$. Inoltre, ci occuperemo di catene *omogenee nel tempo*, cioè per le quali $P_{i,j}(n)$ non dipende da n. Questo ci permette di definire la matrice di transizione $P = (P_{i,j})$, il cui ordine è pari alla cardinalità di Γ e tale che:

- $P_{ij} \geq 0$ per ogni $j \in \Gamma$;
- $\sum_{j \in \Gamma} P_{ij} = 1$, infatti

$$\sum_{j\in\Gamma} P_{ij} = \sum_{j\in\Gamma} \mathbb{P}(X_{n+1} = j | X_n = i) =$$
$$= \mathbb{P}\left(\bigcup_{j\in\Gamma} \{X_{n+1} = j\} | X_n = i\right) = \mathbb{P}(X_{n+1} \in \Gamma | X_n = i) = 1.$$

Viceversa si può dimostrare che data una matrice di transizione e una distribuzione iniziale ν esiste sempre una catena di Markov ad essa associata. Addirittura noi supporremo che $P_{i,j} > 0$ per ogni $i, j \in \Gamma$.
Un'ultima nozione importante della teoria che studia questi processi è quella di distribuzione stazionaria: una distribuzione π si dice stazionaria se $\pi = \pi P$. Per il teorema di Markov, quando la matrice di transizione associata a una catena a stati finiti ha tutti i termini positivi, allora la distribuzione stazionaria associata π è unica e

$$\lim_{m\to\infty} \mathbb{P}(X_{n+m} = j | X_n = i) = \pi_j \qquad \forall\, i \in \Gamma,$$

ovvero, da qualunque distribuzione si parta, X_n converge in legge alla distribuzione invariante. Per dimostrare il teorema principale della tesi imporremo inoltre che la distribuzione iniziale sia proprio quella stazionaria. (per un approfondimento sulle catene di Markov si rimanda a [1]).

Introduciamo adesso un utile strumento per registrare progressivamente le realizzazioni delle variabili aleatorie $X_1, X_2, ..., X_n$: la misura empirica,

$$L_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_i},$$

dove con δ_{X_i} si intende la delta di Dirac.

Osservazione. Per comprendere il significato della misura empirica si noti che, data una particolare realizzazione delle v.a. $X_1, X_2, ..., X_n$, la misura empirica valutata in uno stato $j \in \Gamma$ è il numero di volte che quello stato si è manifestato fratto le realizzazioni totali, cioè la sua frequenza. Di conseguenza L_n è una misura di probabilità su Γ.

Scriviamo

$$\mathfrak{M}(\Gamma) := \{\nu = (\nu_1, ..., \nu_r) \in [0, 1]^r : \sum_{s=1}^{r} \nu_i = 1\}$$

per indicare l'insieme di misure di probabilità su Γ a cui L_n appartiene. Definiamo, poi, su questo spazio, la distanza

$$d(\nu,\mu) := \frac{1}{n}\sum_{s=1}^{r} |\nu_s - \mu_s|,$$

con la quale $\mathfrak{M}(\Gamma)$ diventa uno spazio polacco, cioè metrico, completo e separabile.
In modo del tutto analogo possiamo definire la misura empirica di coppia che registra invece le realizzazioni successive delle v.a., cioè le frequenze delle transizioni tra gli stati per ogni istante di tempo:

$$L_n^2 := \frac{1}{n}\sum_{k=1}^{n} \delta_{(X_i,X_{i+1})},$$

con la convenzione che $X_{n+1} = X_1$, cioè che le condizioni al bordo siano periodiche. La misura empirica di coppia appartiene, invece, all'insieme

$$\tilde{\mathfrak{M}}(\Gamma\times\Gamma) := \{\nu = (\nu_{s,t}) : \sum_{s,t} \nu_{s,t} = 1 \;\wedge\; \sum_{s} \nu_{s,t} = \sum_{t} \nu_{s,t}\},$$

cioè all'insieme delle misure di probabilità su $\Gamma \times \Gamma$ tali che le marginali coincidono. La relativa distanza è

$$d(\nu,\mu) := \frac{1}{2}\sum_{s,t} |\nu_{s,t} - \mu_{s,t}|$$

con cui anche questo diventa uno spazio polacco.

Volgiamo adesso la nostra attenzione a cosa si intende con soddisfare il principio delle grandi deviazioni.

Definizione (funzione di *rate*)**.** Sia $\mathcal{X}$ uno spazio polacco. Si dice che una funzione $I : \mathcal{X} \to [0,\infty]$ è una funzione di *rate*, se:

- $I \not\equiv \infty$.
- I è semicontinua inferiormente.
- I ha insiemi di livello compatti.

Definizione (principio delle grandi deviazioni)**.** Si dice che una successione di misure di probabilità (P_n) soddisfa il principio delle grandi deviazioni con *rate* n e funzione di *rate* I, se:

- I è una funzione di *rate*.
- $\limsup_{n\to\infty} \frac{1}{n}\log P_n(C) \le -I(C) \quad \forall C \subseteq \mathcal{X}$ chiuso.
- $\liminf_{n\to\infty} \frac{1}{n}\log P_n(A) \ge -I(A) \quad \forall A \subseteq \mathcal{X}$ aperto.

dove si intende

$$I(S) = \inf_{x\in S} I(x) \quad \forall S \subseteq \mathcal{X}.$$

Si può dimostrare che se una famiglia di misure di probabilità (P_n) soddisfa il principio delle grandi deviazioni, allora la funzione di *rate* associata è unica.

Osservazione. Notiamo che secondo l'ultima definizione possiamo considerare la funzione di *rate* come la "velocità" con cui $P_n(\cdot)$ decade esponenzialmente quando n va all'infinito.

Nonostante in generale sia restrittivo, in molti esempi gli ultimi due punti della definizione si possono riassumere in

$$\lim_{n\to\infty} \frac{1}{n} \log P_n(S) = -\inf_{x\in S} I(x) \quad \forall \textit{ boreliano } S \subset \mathcal{X}.$$

Questa formula può essere interpretata osservando che l'evento calcolato da $P_n(S)$ ha lo stesso peso probabilistico dell'evento nel quale S è sostituito da un intorno "piccolo" $U \subset S$ di $\bar{x}$, dove $\bar{x}$ è l'elemento di S che realizza il minimo di I. Infatti se $\bar{x}$ minimizza I in S, allora sicuramente minimizza I in U, e quindi

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log P_n(S) = -\inf_{x\in S} I(x) = \\ = -\inf_{x\in U} I(x) = \lim_{n\to\infty} \frac{1}{n} \log P_n(U). \end{aligned}$$

Questo significa che, tra tutte le realizzazioni in S, la probabilità $P_n(S)$ ha il "costo" della più "economica" realizzazione di S. Questo fatto illustra un principio chiave della teoria delle grandi deviazioni: ogni grande deviazione si realizza nel meno improbabile di tutti i modi improbabili.

Come spesso succede in teoria delle grandi deviazioni, vedremo che nei risultati che incontreremo lo spazio polacco $\mathcal{X}$ coinvolto nella definizione precedente è proprio $\mathfrak{M}(\Gamma)$ o $\tilde{\mathfrak{M}}(\Gamma \times \Gamma)$, che la successione di misure di probabilità è proprio $P_n(\cdot) = \mathbb{P}(L_n \in \cdot)$ o $\mathbb{P}(L_n^2 \in \cdot)$ e che l'obiettivo dello studio sarà l'individuazione della funzione di *rate*.

1.2 L'enunciato del teorema principale

Siamo pronti per enunciare il teorema centrale della tesi.

Teorema (grandi deviazioni per catene di Markov). *Sia (X_i) una catena di Markov e sia $P = (P_{s,t})_{s,t\in\Gamma}$ la sua matrice di transizione. Supponiamo che la catena sia a stati finiti, omogenea nel tempo e tale che i termini della matrice di transizione $P_{i,j} > 0$. Allora la famiglia di misure di probabilità (P_n^X) definita da*

$$P_n^X(S) = \mathbb{P}(L_n^2 \in S) \quad \forall \textit{ boreliano } S \subset \tilde{\mathfrak{M}}(\Gamma \times \Gamma)$$

soddisfa il principio delle grandi deviazioni su $\tilde{\mathfrak{M}}(\Gamma \times \Gamma)$ con rate n e funzione di rate

$$I_2^P(\nu) = \sum_{s,t} \nu_{s,t} \log\left(\frac{\nu_{s,t}}{\bar{\nu}_s P_{s,t}}\right),$$

dove $\bar{\nu}_s = \sum_t \nu_{s,t}$.

Si può dimostrare che la funzione di *rate* I_2^P è finita, continua, positiva e si annulla solo in $\nu = (\pi \otimes P)$, dove π è la distribuzione stazionaria di P e $(\pi \otimes P)_{st} = \pi_s P_{st}$. Osserviamo che, infatti, per quanto asserito sulle distribuzioni stazionarie nel paragrafo precedente,

$$\pi_s P_{st} = \lim_{n\to\infty} \mathbb{P}(X_n = s | X_1 = i)\mathbb{P}(X_{n+1} = t | X_n = s) \qquad \forall i \in \Gamma,$$

quindi, indipendentemente dalla distribuzione iniziale, è naturale aspettarsi che L_n^2 converga a $(\pi \otimes P)$. Notiamo adesso che se consideriamo un intorno aperto U di $(\pi \otimes P)$, allora

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n^X(U) \geq -\inf_{x\in U} I_2^P(x) = -I_2^P((\pi \otimes P)) = 0$$

e, poiché $P_n^X(\cdot) \leq 1$ in quanto è una misura di probabilità, l'unico caso in cui l'ultima disuguaglianza abbia senso è che

$$\lim_{n\to\infty} P_n^X(U) = 1,$$

cioè, per $n \to \infty$, L_n^2 si trova in un intorno aperto di $(\pi \otimes P)$ con probabilità 1, in accordo con le aspettative. Tutto questo per osservare che la teoria delle grandi deviazioni funziona anche in prossimità dei punti dove converge la famiglia di misure di probabilità, e che, per le effettive grandi deviazioni, fornisce una funzione che descrive la velocità del decadimento esponenziale della probabilità.

Spostiamo adesso la nostra attenzione su qualcuno dei risultati fondamentali della teoria delle grandi deviazioni per v.a. i.i.d., per arrivare infine a dimostrare il teorema appena enunciato.

Capitolo 2

Due risultati dalla teoria delle grandi deviazioni

In questo capitolo analizziamo due risultati di teoria delle grandi deviazioni necessari per dimostrare il teorema di nostro interesse: il teorema di Sanov e il lemma di Varadhan. Il primo descrive le grandi deviazioni per la misura empirica e per la misura empirica di coppia applicate al caso di una famiglia di v.a. indipendenti e identicamente distribuite; il secondo illustra e giustifica una tecnica largamente utilizzata che ci permette di generare un principio delle grandi deviazioni a partire da uno già noto: il *tilting*.

2.1 Il teorema di Sanov

Nella seguente sezione illustriamo due versioni del teorema di Sanov. La prima versione individua la funzione di *rate* per la famiglia di misure di probabilità definita da $\mathbb{P}(L_n \in \cdot)$; la seconda versione, quella che servirà a noi, fa la stessa cosa per $\mathbb{P}(L_n^2 \in \cdot)$. Questo secondo teorema ha un forte somiglianza col teorema sulle grandi deviazioni per catene di Markov, con la sostanziale differenza che tra le v.a. di una catena di Markov sussiste una dipendenza, mentre nel caso del teorema di Sanov si suppone che le v.a. siano i.i.d. Vedremo che questo più semplice risultato sarà un fondamentale appoggio sul quale si fonda la dimostrazione del teorema principale della tesi. Ristabiliamo allora con quali condizioni proseguiamo in questo paragrafo.
Sia (X_i) una famiglia di v.a. tale che:

$X_i \in \Gamma = \{1, ..., r\}$
$X_1, X_2, ...$ sono i.i.d. con legge marginale $\rho = (\rho_s)_{s \in \Gamma}$
$\rho_s > 0 \quad \forall s \in \Gamma$,

e consideriamo lo spazio polacco $\mathfrak{M}(\Gamma)$ con la relativa distanza, come definiti nel paragrafo 1.1. Osserviamo preliminarmente che, per la legge dei grandi numeri,

$$\lim_{n \to \infty} d(L_n, \rho) = 0,$$

dove $\rho = (\rho_1, ..., \rho_r)$. Enunciamo allora il teorema:

Teorema (Sanov)**.** *Sia* (X_i) *una famiglia di variabili aleatorie i.i.d. che soddisfano le ipotesi appena descritte. Sia* $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. *Allora, per ogni* $a>0$,

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(L_n \in B_a^c(\rho)) = -\inf_{\nu\in B_a^c(\rho)} I_\rho(\nu),$$

dove $B_a(\rho) = \{\nu \in \mathfrak{M}(\Gamma) : d(\nu,\rho) \le a\}$, $B_a^c(\rho) = \mathfrak{M}(\Gamma) \setminus B_a(\rho)$, *e*

$$I_\rho(\nu) = \sum_{s=1}^r \nu_s \log\left(\frac{\nu_s}{\rho_s}\right)$$

con la convenzione che $\inf_\emptyset I_\rho = \infty$

Dimostrazione. Sia

$$K_n = \Big\{k = (k_1, ..., k_r) \in \mathbb{N}_0^r : \sum_{s=1}^r k_s = n\Big\},$$

e osserviamo che $L_n \in \frac{1}{n}K_n \subset \mathfrak{M}(\Gamma)$ per ogni $n \in \mathbb{N}$. Calcoliamo adesso $\mathbb{P}(L_n = \frac{1}{n}k)$. Osservo che $\mathbb{P}(L_n(s) = \frac{k_s}{n})$ è la probabilità che su n realizzazioni delle v.a. $X_1, ..., X_n$ si verifichi k_s volte lo stato ρ_s. Quindi $\mathbb{P}(L_n = \frac{1}{n}k)$ ha distribuzione multinomiale di parametri $((\rho_1, ..., \rho_r), n)$:

$$\mathbb{P}\Big(L_n = \frac{1}{n}k\Big) = \binom{n}{k_1, ..., k_r} \prod_{s=1}^r \rho_s^{k_s} = n! \prod_{s=1}^r \frac{\rho_s^{k_s}}{k_s!} \quad k \in K_n.$$

Sia $\nu^n(k) = \frac{1}{n}k \in \mathfrak{M}(\Gamma)$, per $k \in K_n$. Poniamo quindi

$$Q_n(a) = \max_{k\in K_n : \nu^n(k)\in B_a^c(\rho)} \left(n! \prod_{s=1}^r \frac{\rho_s^{k_s}}{k_s!}\right).$$

Ora poiché

$$\mathbb{P}(L_n \in B_a^c(\rho)) = \sum_{\nu^n(k)\in B_a^c(\rho)} \mathbb{P}(L_n = \nu^n(k))$$

$$e$$

$$\mathbb{P}(L_n = \nu^n(k)) \le Q_n(a) \;\; per \; \nu^n(k) \in B_a^c,$$

allora chiaramente,

$$Q_n(a) \le \mathbb{P}(L_n \in B_a^c(\rho)) \le |K_n| Q_n(a).$$

Stimiamo adesso $Q_n(a)$ ricordando l'approssimazione di Stirling:

$$\lim_{n\to\infty} \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{n!} = 1,$$

cioè, in una formulazione più adatta ai nostri scopi,

$$\log(n!) = n\log(n) - n + O(\log(n)) \quad per \; n \to \infty.$$

Usando anche che $\sum_{s=1}^r k_s = n$, abbiamo che

$$\frac{1}{n}\log\left(n!\prod_{s=1}^r \frac{\rho_s^{k_s}}{k_s!}\right) = \frac{1}{n}\log n! - \frac{1}{n}\sum_{s=1}^r \log k_s! + \sum_{s=1}^r \frac{k_s}{n}\log\rho_s =$$
$$=\frac{1}{n}(n\log(n)-n) - \frac{1}{n}\sum_{s=1}^n (k_s\log(k_s)-k_s) + \sum_{s=1}^r \frac{k_s}{n}\log\rho_s + O\left(\frac{\log n}{n}\right) =$$
$$=\log n - \sum_{s=1}^r \frac{k_s}{n}\log k_s + \sum_{s=1}^r \frac{k_s}{n}\log\rho_s + O\left(\frac{\log n}{n}\right) =$$
$$=\sum_{s=1}^r \frac{k_s}{n}\log n - \sum_{s=1}^r \frac{k_s}{n}\log k_s + \sum_{s=1}^r \frac{k_s}{n}\log\rho_s + O\left(\frac{\log n}{n}\right) =$$
$$=\sum_{s=1}^r \frac{k_s}{n}\log\rho_s - \sum_{s=1}^r \frac{k_s}{n}\log\frac{k_s}{n} + O\left(\frac{\log n}{n}\right) =$$
$$=\sum_{s=1}^r \frac{k_s}{n}\left(\log\rho_s - \log\frac{k_s}{n}\right) + O\left(\frac{\log n}{n}\right) = -\sum_{s=1}^r \nu_s^n(k)\log\left(\frac{\nu_s^n(k)}{\rho_s}\right) + O\left(\frac{\log n}{n}\right)$$

dove la somma all'ultimo membro è $-I_\rho(\nu_s^n(k))$. Sia quindi $\bar\nu^n(k)$ la distribuzione che realizza il massimo $Q_n(a)$, che, per quanto appena dimostrato, è la stessa distribuzione che realizza il minimo di $I_\rho(\nu)$ in $B_a^c \cap \frac{1}{n}K_n$. Osserviamo inoltre che $|K_n| \le n^r$, quindi che $\frac{1}{n}\log|K_n| \le r\frac{\log n}{n}$, arriviamo quindi a

$$\frac{1}{n}\log Q_n(a) = -\sum_{s=1}^r \bar\nu_s^n(k)\log\left(\frac{\bar\nu_s^n(k)}{\rho_s}\right) + O\left(\frac{\log n}{n}\right) =$$
$$= -\min_{k\in K_n:\nu^n(k)\in B_a^c(\rho)} I_\rho(\nu^n(k)) + O\left(\frac{\log n}{n}\right) \le \frac{1}{n}\log\mathbb{P}(L_n \in B_a^c(\rho)) \le$$
$$\le\frac{1}{n}\log|K_n|Q_n(a) = -\min_{k\in K_n:\nu^n(k)\in B_a^c(\rho)} I_\rho(\nu^n(k)) + O\left(\frac{\log n}{n}\right) + r\frac{\log n}{n}.$$

Quindi, in modo più compatto, possiamo scrivere che

$$\frac{1}{n}\mathbb{P}(L_n \in B_a^c(\rho)) = -\min_{k\in K_n:\nu^n(k)\in B_a^c(\rho)} I_\rho(\nu^n(k)) + O\left(\frac{\log n}{n}\right)$$

Per completare la dimostrazione dobbiamo notare che:

$\bigcup_{n\in\mathbb{N}}\{\nu_n(k) : k \in K_n\}$ è denso in $\mathfrak{M}(\Gamma)$,

$\nu \mapsto I_\rho(\nu)$ è continua su $\mathfrak{M}(\Gamma)$.

Infatti, per la densità di K in $\mathfrak{M}(\Gamma)$ esiste una successione (k_n), con $k_n \in K_n$, tale che

$$\lim_{n\to\infty} d(\nu^n(k_n),\nu) = 0$$

e per la continuità di I_ρ

$$\lim_{n\to\infty} I_\rho(\nu^n(k_n)) = I_\rho(\nu).$$

Adesso, poiché $B_a^c(\rho)$ è un insieme aperto, questo implica che

$$\limsup_{n\to\infty} \min_{k\in K_n:\nu^n(k)\in B_a^c(\rho)} I_\rho(\nu^n(k)) \le I_\rho(\nu) \quad \forall \nu \in B_a^c(\rho)$$

e ottimizzando su ν,

$$\limsup_{n\to\infty} \min_{k\in K_n:\nu^n(k)\in B_a^c(\rho)} I_\rho(\nu^n(k)) \le \inf_{\nu\in B_a^c(\rho)} I_\rho(\nu)$$

e, chiaramente

$$\inf_{\nu\in B_a^c(\rho)} I_\rho(\nu) \le \liminf_{n\to\infty} \min_{k\in K_n:\nu^n(k)\in B_a^c(\rho)} I_\rho(\nu^n(k)).$$

Riordinando i termini, abbiamo dimostrato che

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\mathbb{P}(L_n \in B_a^c(\rho)) = \\
&= \lim_{n\to\infty} - \min_{k\in K_n:\nu^n(k)\in B_a^c(\rho)} I_\rho(\nu^n(k)) + O\left(\frac{\log n}{n}\right) = -\inf_{\nu\in B_a^c} I_\rho(\nu)
\end{aligned}$$

□

Osservazione. La scelta della distanza nella dimostrazione è flessibile, come anche B_a^c può essere rimpiazzato da un insieme aperto arbitrario.

Presentiamo adesso la versione del teorema che più ci interessa. A questo scopo consideriamo lo spazio metrico $\tilde{\mathfrak{M}}(\Gamma\times\Gamma)$ con la relativa distanza e osserviamo analogamente a prima che, per il teorema ergodico di Birkhoff,

$$\lim_{n\to\infty} d(L_n^2, \rho\times\rho) = 0.$$

Abbiamo omesso la dimostrazione in quanto concettualmente simile quella appena percorsa, ma complicata da un termine combinatorio che la rende di più difficile comprensione.

Teorema (Sanov per coppie)**.** *Sia (X_i) una famiglia di variabili aleatorie i.i.d. con le condizioni appena definite. Sia $L_n^2 = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i,X_{i+1})}$ con condizioni al bordo periodiche. allora per ogni $a > 0$*

$$\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}(L_n^2 \in B_a^c(\rho\times\rho)) = -\inf_{\nu\in B_a^c(\rho\times\rho)} I_\rho^2(\nu),$$

dove $B_a(\rho\times\rho) = \{\nu \in \tilde{\mathfrak{M}}(\Gamma\times\Gamma) : d(\nu,\rho\times\rho) \le a\}$ e
$B_a^c(\rho\times\rho) = \tilde{\mathfrak{M}}(\Gamma\times\Gamma) \setminus B_a(\rho\times\rho)$

$$I_\rho^2(\nu) = \sum_{s,t} \nu_{s,t} \log\left(\frac{\nu_{s,t}}{\bar{\nu}_s \rho_t}\right)$$

dove $\bar{\nu}_s = \sum_t \nu_{s,t}$.

Questo teorema sarà utilizzato nella dimostrazione nell'analogo teorema per le catene di Markov e la funzione di *rate* si otterrà attraverso una manipolazione di quella appena descritta.

2.2 Il lemma di Varadhan

In questo paragrafo dimostriamo un importante risultato generale di teoria delle grandi deviazioni: il lemma di Varadhan. Anche in questo caso ne illustriamo due versioni e di entrambe faremo uso nella dimostrazione del teorema principale.
Facciamo una piccola osservazione preliminare che ci sarà utile per la dimostrazione. Due successioni a termini positivi (α_n) e (β_n) si dicono logaritmicamente equivalenti e si scrive $\alpha_n \simeq \beta_n$, se

$$\lim_{n\to\infty} \frac{1}{n}(\log \alpha_n - \log \beta_n) = 0.$$

Notiamo adesso che se $\alpha_n = o(\beta_n)$, allora è chiaro che

$$\alpha_n + \beta_n \simeq \beta_n$$

e se $\alpha_n = O(\beta_n)$, allora

$$\alpha_n + \beta_n \simeq \beta_n \simeq \alpha_n.$$

Indichiamo questo fatto con la notazione compatta

$$\alpha_n + \beta_n \simeq \alpha_n \vee \beta_n. \tag{2.1}$$

L'uguaglianza si può estendere facilmente al caso di una quantità finita di successioni.
Siamo pronti ad enunciare il risultato di Varadhan, che nonostante si presenti come un teorema, nella letteratura è rimasto noto con la denominazione di "lemma".

Teorema (Lemma di Varadhan). *Sia $\mathcal{X}$ uno spazio polacco. Sia (P_n) una famiglia di misure di probabilità che soddisfa il principio delle grandi deviazioni su $\mathcal{X}$ con rate n e funzione di rate I. Sia $F : \mathcal{X} \to \mathbb{R}$ una funzione continua e limitata superiormente. Allora*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)}\, P_n(dx) = \sup_{x\in\mathcal{X}}[F(x) - I(x)].$$

Dimostrazione. Sia

$$J_n(S) = \int_S e^{nF(x)}\, P_n(dx) \quad \forall \textit{ boreliano } S \subseteq \mathcal{X},$$

e poniamo

$$a = \sup_{x\in\mathcal{X}} F(x), \quad b = \sup_{x\in\mathcal{X}}[F(x) - I(x)].$$

Osserviamo che $-\infty < b \leq a < \infty$, perché $I \geq 0$ e F è continua e limitata superiormente. Procediamo dimostrando l'uguaglianza attraverso due disuguaglianze e iniziamo da

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \leq b$$

Dividiamo lo spazio $\mathcal{X}$ a seconda dei valori che assume F, cioè sia $C = F^{-1}([b, a])$, e fissato un $N \in \mathbb{N}$ definiamo gli insiemi

$$C_j^N = F^{-1}([c_{j-1}^N, c_j^N]) \quad j = 1, ..., N,$$

dove $c_j^n = b + \frac{j}{N}(a-b)$ per $j = 0, 1, ..., N$. Ovviamente,

$$C = \bigcup_{j=1}^{N} C_j^N.$$

Poiché F è continua, tutti i C_j^N sono chiusi. Quindi applicando la definizione di principio delle grandi deviazioni, ottengo

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C_j^N) \leq -I(C_j^N) \quad \forall j = 1, ..., N.$$

Notiamo inoltre che, per come abbiamo definito i C_j^N abbiamo che $F(x) \leq c_j^N$ su C_j^N. Consideriamo adesso che

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) = \limsup_{n\to\infty} \frac{1}{n} \log \int_C e^{nF(x)}\, P_n(dx) \leq \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=1}^{N} \int_{C_j^N} e^{nF(x)}\, P_n(x) \leq \limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=1}^{N} e^{nc_j^N} \int_{C_j^N} P_n(dx) = \\
&= \limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=1}^{N} e^{nc_j^N}\, P_n(C_j^N)
\end{aligned}$$

Applichiamo quindi (2.2) alla famiglia di successioni

$$\alpha_n^j = e^{nc_j^N}\, P_n(C_j^N) \quad j = 1, ..., N,$$

e sia k tale che

$$\alpha_n^k = \bigvee_{j=1}^{N} \alpha_n^j.$$

Otteniamo quindi

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \leq \limsup_{n\to\infty} \frac{1}{n} \log e^{nc_k^N}\, P_n(C_k^N) = \\
&= c_k^N + \limsup_{n\to\infty} \frac{1}{n} P_n(C_k^N) \leq c_k^N - I(C_k^N) \leq \\
&\leq \max_{1\leq j\leq N} [c_j^N - I(C_j^N)].
\end{aligned}$$

Adesso, per proseguire nelle maggiorazioni osserviamo che $c_j^N \leq \inf_{x\in C_j^N} F(x) + \frac{1}{N}(a-b)$ e otteniamo

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \leq \max_{1\leq j\leq N} \left\{ \inf_{x\in C_j^N} F(x) - \inf_{x\in C_j^N} I(x) \right\} + \frac{1}{N}(a-b) \leq \\
&\leq \max_{1\leq j\leq N} \sup_{x\in C_j^N} [F(x) - I(x)] + \frac{1}{N}(a-b) = \sup_{x\in C} [F(x) - I(x)] + \frac{1}{N}(a-b) \leq \\
&\leq b + \frac{1}{N}(a-b).
\end{aligned}$$

Mandando $N \to \infty$ otteniamo che $\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \leq b$. Per estendere il risultato a $J_n(\mathcal{X})$, stimiamo $J_n(\mathcal{X} \setminus C)$. Osserviamo che $F(x) \leq b$ in $\mathcal{X} \setminus C$, quindi

$$J_n(\mathcal{X} \setminus C) = \int_{\mathcal{X}\setminus C} e^{nF(x)}\, P_n(dx) \leq e^{nb} \int_{\mathcal{X}\setminus C} P_n(dx) \leq e^{nb} \int_{\mathcal{X}} P_n(dx) = e^{nb}.$$

Di conseguenza anche per $\mathcal{X} \setminus C$ vale che

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X} \setminus C) \leq b.$$

Dunque

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) = \limsup_{n\to\infty} \frac{1}{n} \log \left[J_n(\mathcal{X} \setminus C) + J_n(C)\right],$$

a cui possiamo riapplicare (2.2) in quanto l'argomento del logaritmo del membro di destra è logaritmicamente equivalente a uno dei due addendi e qualunque dei due esso sia, vale che

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \leq b.$$

Dimostriamo adesso che

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \geq b.$$

Sia $x \in \mathcal{X}$ e sia $\epsilon > 0$ arbitrario, allora l'insieme

$$\mathcal{O}_{x,\epsilon} = \{y \in \mathcal{X} : F(y) > F(x) - \epsilon\}$$

è un intorno aperto di x per la continuità di F. Dalla definizione di principio delle grandi deviazioni otteniamo che per ogni ϵ e per ogni x,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(\mathcal{O}_{x,\epsilon}) \geq -I(\mathcal{O}_{x,\epsilon})$$

e poiché $I(\mathcal{O}_{x,\epsilon}) \leq I(x)$ otteniamo

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{O}_{x,\epsilon}) = \liminf_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{O}_{x,\epsilon}} e^{nF(y)}\, P_n(dy) \\
\geq &F(x) - \epsilon + \liminf_{n\to\infty} \log P_n(\mathcal{O}_{x,\epsilon}) \geq F(x) - \epsilon - I(\mathcal{O}_{x,\epsilon}) \geq \\
\geq &F(x) - \epsilon - I(x).
\end{aligned}$$

Notiamo adesso che $J_n(\mathcal{X}) \geq J_n(\mathcal{O}_{x,\epsilon})$ e, mandando $\epsilon \to 0$, troviamo

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \geq F(x) - I(x) \quad \forall x \in \mathcal{X},$$

cioè,

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \geq b$$

$\square$

Riportiamo adesso una versione alternativa del Lemma di Varadhan, che presenta come generare una nuova famiglia di misure di probabilità a partire da una nota e individuarne la funzione di *rate* in funzione di quella conosciuta. Questa tecnica è nota come *tilting* e ne faremo uso nell'ultimo capitolo per dimostrare il teorema principale della tesi.

Teorema (principio delle grandi deviazioni *tilted*)**.** *Sia (P_n) una famiglia di misure di probabilità che soddisfa il principio delle grandi deviazioni su uno spazio polacco $\mathcal{X}$ con rate n e funzione di rate I. Sia $F : \mathcal{X} \to \mathbb{R}$ una funzione continua e limitata superiormente. Definiamo*

$$J_n(S) = \int_S e^{nF(x)}\, P_n(dx), \qquad \forall \text{ boreliano } S \subset \mathcal{X}.$$

Allora la successione (P_n^F) di misure di probabilità definite da

$$P_n^F(S) = \frac{J_n(S)}{J_n(\mathcal{X})}, \qquad \forall \text{ boreliano } S \subset \mathcal{X}$$

soddisfa il principio delle grandi deviazioni su $\mathcal{X}$ con rate n e funzione di rate

$$I^F(x) = \sup_{y\in\mathcal{X}}[F(y) - I(y)] - [F(x) - I(x)].$$

Dimostrazione. Sappiamo dal lemma di Varadhan che

$$\lim_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) = \sup_{y\in\mathcal{X}}[F(y) - I(y)].$$

Osserviamo adesso che, definendo $b(S) = \sup_{x\in S}[F(x)-I(x)]$ per ogni boreliano $S \subset \mathcal{X}$ e ripercorrendo la prima parte della dimostrazione del lemma di Varadhan per un generico chiuso $C \subset \mathcal{X}$, dimostriamo facilmente che

$$\limsup_{n\to\infty} \frac{1}{n} J_n(C) \le b(C) \qquad \forall\, C \subset \mathcal{X} \text{ chiuso},$$

e analogamente, seguendo la seconda parte della dimostrazione del lemma di Varadhan,

$$\liminf_{n\to\infty} \frac{1}{n} J_n(A) \ge b(A) \qquad \forall\, A \subset \mathcal{X} \text{ aperto}.$$

Otteniamo quindi che

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n^F(C) = \limsup_{n\to\infty} \frac{1}{n} \log J_n(C) - \frac{1}{n} \log J_n(\mathcal{X}) \le$$
$$\le -\sup_{y\in\mathcal{X}}[F(y) - I(y)] + \sup_{x\in C}[F(x) - I(x)] \qquad \forall\, C \subset \mathcal{X} \text{ chiuso}$$

e analogamente,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n^F(A) \ge -\sup_{y\in\mathcal{X}}[F(y) - I(y)] + \sup_{x\in A}[F(x) - I(x)] \quad \forall\, A \subset \mathcal{X} \text{ aperto}.$$

Notando che

$$I^F(S) = \inf_{x\in S} I^F(x) = \inf_{x\in S}\{\sup_{y\in\mathcal{X}}[F(y) - I(y)] - [F(x) - I(x)]\} =$$
$$= \sup_{y\in\mathcal{X}}[F(y) - I(y)] - \sup_{x\in S}[F(x) - I(x)] \qquad \forall S \subset \mathcal{X} \text{ boreliano},$$

la dimostrazione è conclusa. □

Capitolo 3

La dimostrazione del teorema di grandi deviazioni per catene di Markov

Prima di dimostrare il teorema, introduciamo un utile formula che ci servirà trasformare domande che riguardano la probabilità di v.a. di una catena di Markov nelle loro analoghe per v.a. indipendenti, alle quali possiamo applicare i teoremi dimostrati nel capitolo precedente.

3.1 La formula di Radon-Nikodym

Torniamo a considerare una catena di Markov con le condizioni stabilite nel paragrafo 1.1. Per comodità le riassumiamo di seguito.
Sia (X_i) una famiglia di v.a. tale che

$X_i \in \Gamma = \{1, ..., r\} \subset \mathbb{N}$,
$X_1, X_2, ...$ è una catena di Markov
con matrice di transizione $P = (P_{st})_{s,t\in\Gamma}$,
$P_{st} > 0 \; \forall \, s, t \in \Gamma$.

La distribuzione stazionaria della catena di Markov $\pi = (\pi_s)$ è unica e $\pi_s > 0$ per ogni $s \in \Gamma$. Prendiamo come distribuzione iniziale proprio quella stazionaria

$$\mathbb{P}(X_1 = s) = \pi_s \qquad \forall \, s \in \Gamma,$$

in modo che la catena di Markov sia invariante. Denotiamo con $\mathbb{P}^X$ la legge di (X_i). Introduciamo dunque una famiglia di v.a. ausiliarie i.i.d. (Y_i) con valori in Γ e con distribuzione π. Denotiamo la legge di (Y_i) con $\mathbb{P}^Y$. Sviluppiamo quindi,

$$\begin{aligned}\mathbb{P}^X[x_1, ..., x_n] =& \mathbb{P}(X_1 = x_1, ..., X_n = x_n) = \\ =& \pi_{x_1} P_{x_1,x_2} \times ... \times P_{x_{n-1},x_n} \frac{P_{x_n,x_1}}{P_{x_n,x_1}} = \\ =& \frac{\pi_{x_1}}{P_{x_n,x_1}} e^{\sum_{i=1}^{n} \log P_{x_i,x_{i+1}}}.\end{aligned}$$

Osserviamo adesso che, se $L_n^2[x_1, ..., x_n]$ è la misura empirica di coppia associata a $x_1, ..., x_n$, allora $nL_n^2[x_1, ..., x_n](s,t)$ è la quantità di volte che nella realizzazione $x_1, ..., x_n$ è successo l'evento $\{X_i = s, X_{i+1} = t\}$, quindi

$$\sum_{i=1}^{n} \log P_{x_i, x_{i+1}} = n \sum_{s,t} L_n^2[x_1, ..., x_n](s,t) \log P_{st},$$

e otteniamo che

$$\frac{\pi_{x_1}}{P_{x_n, x_1}} e^{\sum_{i=1}^{n} \log P_{x_i, x_{i+1}}} = \frac{\pi_{x_1}}{P_{x_n, x_1}} e^{n \sum_{s,t} L_n^2[x_1, ..., x_n](s,t) \log P_{st}}.$$

Attraverso simili passaggi possiamo concludere che

$$\mathbb{P}^Y[x_1, ..., x_n] = \prod_{i=1}^{n} \pi_{x_i} = e^{n \sum_{s,t} L_n^2[x_1, ..., x_n](s,t) \log \pi_t}.$$

Combinando quanto appena visto, arriviamo a

$$\begin{aligned} &\frac{d\mathbb{P}^X}{d\mathbb{P}^Y}[\,\cdot\,] = \frac{\pi_{x_1}}{P_{x_n, x_1}} e^{n \sum_{s,t} L_n^2[\,\cdot\,](s,t) \log P_{st} - L_n^2[\,\cdot\,](s,t) \log \pi_t} = \\ &= O(1) e^{nF(L_n^2[\,\cdot\,])}, \end{aligned} \tag{3.1}$$

dove

$$F(\nu) = \sum_{s,t} \nu_{st} \log\left(\frac{P_{st}}{\pi_t}\right), \qquad \nu \in \tilde{\mathfrak{M}}(\Gamma \times \Gamma).$$

Notiamo che F è limitata e continua su $\tilde{\mathfrak{M}}(\Gamma \times \Gamma)$. La formula di Radon-Nikodym (3.1), ci permette quindi di derivare il differenziale di P_n^Y da quello di P_n^X.

3.2 La dimostrazione del teorema principale e un esempio

Siamo finalmente pronti per dimostrare il teorema delle grandi deviazioni per le catene di Markov, enunciato nel paragrafo 1.2.

Dimostrazione. Applicando la formula di Radon-Nikodym, per ogni boreliano $S \subset \tilde{\mathfrak{M}}(\Gamma \times \Gamma)$,

$$\begin{aligned} &\frac{1}{n} \log P_n^X(S) = \frac{1}{n} \log \int_S \mathbb{P}_n^X(L_n^2 \in d\nu) = \\ &= O\left(\frac{1}{n}\right) + \frac{1}{n} \log \int_S e^{nF(\nu)}\, \mathbb{P}_n^Y(L_n^2 \in d\nu). \end{aligned} \tag{3.2}$$

Adesso, sappiamo dal teorema di Sanov per coppie che (P_n^Y) definita da $P_n^Y(\cdot) = \mathbb{P}^Y(L_n^2 \in \cdot)$ soddisfa il principio delle grandi deviazioni su $\tilde{\mathfrak{M}}(\Gamma \times \Gamma)$ con *rate* n e funzione di *rate*

$$I_\pi^2(\nu) = \sum_{s,t} \nu_{st} \log\left(\frac{\nu_{st}}{\bar{\nu}_s \pi_t}\right).$$

Notiamo adesso che l'integrale nell'ultimo membro di (3.2) è esattamente della forma del teorema del principio delle grandi deviazioni *tilted*, perché F è limitata e continua. Definendo quindi,

$$J_n(S) = \int_S e^{nF(\nu)}\, \mathbb{P}_n^Y(L_n^2 \in d\nu) \quad e$$
$$P_n^F(S) = \frac{J_n(S)}{J_n(\tilde{\mathfrak{M}}(\Gamma \times \Gamma))} \qquad\qquad \forall S \subset \tilde{\mathfrak{M}}(\Gamma \times \Gamma),$$

possiamo concludere che P_n^F soddisfa il principio delle grandi deviazioni su $\tilde{\mathfrak{M}}(\Gamma \times \Gamma)$ con *rate* n e funzione di *rate*

$$I^F(\nu) = \sup_{\mu \in \tilde{\mathfrak{M}}(\Gamma\times\Gamma)} [F(\mu) - I_\pi^2(\mu)] - [F(\nu) - I_\pi^2(\nu)].$$

Inoltre, per il Lemma di Varadhan, sappiamo che

$$\lim_{n\to\infty} \frac{1}{n} \log J_n(\tilde{\mathfrak{M}}(\Gamma \times \Gamma)) = \sup_{\mu \in \tilde{\mathfrak{M}}(\Gamma\times\Gamma)} [F(\mu) - I_\pi^2(\mu)].$$

Ricomponendo quanto appena dimostrato possiamo infine affermare che

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log P_n^X(C) = \limsup_{n\to\infty} O\left(\frac{1}{n}\right) + \frac{1}{n} \log \int_C e^{nF(\nu)}\, \mathbb{P}_n^Y(L_n^2 \in d\nu) = \\
&= \limsup_{n\to\infty} \frac{1}{n} \log J_n(C) = \limsup_{n\to\infty} \frac{1}{n} \log \left[P_n^F(C) \cdot J_n(\tilde{\mathfrak{M}}(\Gamma \times \Gamma))\right] \leq \\
&\leq -\inf_{\nu \in C} I^F(\nu) + \sup_{\mu \in \tilde{\mathfrak{M}}(\Gamma\times\Gamma)} [F(\mu) - I_\pi^2(\mu)] = \\
&= - \sup_{\mu \in \tilde{\mathfrak{M}}(\Gamma\times\Gamma)} [F(\mu) - I_\pi^2(\mu)] + \sup_{\nu \in C} [F(\nu) - I_\pi^2(\nu)] + \sup_{\mu \in \tilde{\mathfrak{M}}(\Gamma\times\Gamma)} [F(\mu) - I_\pi^2(\mu)] = \\
&= -\inf_{\nu \in C} [I_\pi^2(\nu) - F(\nu)],
\end{aligned}$$

per ogni chiuso $C \subset \tilde{\mathfrak{M}}(\Gamma \times \Gamma)$. Allo stesso modo abbiamo che per ogni aperto $A \subset \tilde{\mathfrak{M}}(\Gamma \times \Gamma)$,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n^X(A) \geq -\inf_{\nu \in A} [I_\pi^2(\nu) - F(\nu)].$$

Abbiamo quindi dimostrato che P_n^X soddisfa il principio delle grandi deviazioni con *rate* n e funzione di *rate*

$$I_P^2(\nu) = I_\pi^2(\nu) - F(\nu) = \sum_{s,t} \nu_{st} \log\left(\frac{\nu_{st}}{\bar{\nu}_s P_{st}}\right).$$

□

Riportiamo, infine, un esempio di applicazione del teorema. Sia $\Gamma = \{1, 2\}$ e sia (X_i) a valori in Γ tale che $X_0 = 1$, cioè con distribuzione iniziale $\nu_0 = (0, 1)$ sia P la matrice di transizione definita come segue

$$P = \begin{bmatrix} p & q \\ q & p \end{bmatrix},$$

cioè tale per cui $\mathbb{P}(X_{i+1} \neq X_i) = q$. Consideriamo adesso una famiglia di variabili aleatorie (Y_i) con $Y_0 = 1$, definite da

$$Y_{i+1} = \begin{cases} 2Y_i & \textit{se } X_{i+1} \neq X_i \\ Y_i & \textit{se } X_{i+1} = X_i, \end{cases}$$

e studiamo il comportamento di $\mathbb{E}[Y_n]$ per n grande. Ora,

$$Y_n = 2^{|\{0 \leq i \leq n-1 : X_{i+1} \neq X_i\}|}.$$

Definiamo allora la v.a. $T_n = |\{0 \leq i \leq n-1 : X_{i+1} \neq X_i\}|$, che ha distribuzione binomiale $\mathbb{P}(T_n = k) = \binom{n}{k} q^k p^{n-k}$ e, quindi, media $\mathbb{E}[T_n] = nq$. Si potrebbe erroneamente pensare che

$$\mathbb{E}[Y_n] = \mathbb{E}\left[2^{|\{0 \leq i \leq n-1 : X_{i+1} \neq X_i\}|}\right] \approx 2^{\mathbb{E}[|\{0 \leq i \leq n-1 : X_{i+1} \neq X_i\}|]} = 2^{nq},$$

Ma questa approssimazione non tiene conto degli eventi rari, che con un esplosione esponenziale diventano rilevanti. Notiamo che $T_n = n[L_n^2(1,2) + L_n^2(2,1)]$, dove $L_n^2 = L_n^2[X_0, ..., X_n]$. Inoltre, fissato n, $L_n^2 \in \frac{1}{n} K_n \subset \tilde{\mathfrak{M}}(\Gamma \times \Gamma)$ con

$$K_n = \{k = (k_{st}) \in \mathbb{N}_0^{2 \times 2} : \sum_{s,t} k_{st} = n, \sum_t k_{st} = \sum_s k_{st}\}.$$

Analizziamo allora

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[Y_n] = \lim_{n \to \infty} \frac{1}{n} \log \left(\sum_{k=0}^{n} 2^k \mathbb{P}(T_n = k) \right).$$

Adesso,

$$\mathbb{P}(T_n = k) = \mathbb{P}\Big(n(L_n^2(1,2) + L_n^2(2,1)) = k\Big) = \sum_{\nu \in S_k} \mathbb{P}(L_n^2 = \nu)$$

con $S_k = \{\nu \in \frac{1}{n} K_n : n(\nu_{12} + \nu_{21}) = k\}$, dunque

$$\sum_{k=0}^{n} 2^k \mathbb{P}(T_n = k) = \sum_{k=0}^{n} 2^k \sum_{\nu \in S_k} \mathbb{P}(L_n^2 = \nu) = \sum_{\nu \in \frac{1}{n} K_n} 2^{n(\nu_{12} + \nu_{21})} \mathbb{P}(L_n^2 = \nu).$$

Ricordando (2.2), l'ultima sommatoria è logaritmicamente equivalente a $\max_{\nu \in \frac{1}{n} K_n} [2^{n(\nu_{12} + \nu_{21})} \mathbb{P}(L_n^2 = \nu)]$. Usiamo che $\bigcup_{n \in \mathbb{N}} \frac{1}{n} K_n$ è denso in $\tilde{\mathfrak{M}}(\Gamma \times \Gamma)$ e applichiamo il nostro teorema, per ottenere l'andamento asintotico corretto di $\mathbb{E}[Y_n]$:

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[Y_n] &= \lim_{n \to \infty} \frac{1}{n} \log \left(\max_{\nu \in \frac{1}{n} K_n} [2^{n(\nu_{12} + \nu_{21})} \mathbb{P}(L_n^2 = \nu)] \right) = \\ &= \lim_{n \to \infty} \sup_{\nu \in \tilde{\mathfrak{M}}(\Gamma \times \Gamma)} \left[\frac{1}{n} \log(2^{n(\nu_{12} + \nu_{21})}) + \frac{1}{n} \log(\mathbb{P}(L_n^2 = \nu)) \right] = \\ &= \sup_{\nu \in \tilde{\mathfrak{M}}(\Gamma \times \Gamma)} \left[(\nu_{12} + \nu_{21}) \log 2 - I_2^P(\nu) \right]. \end{aligned}$$

Bibliografia

[1] BALDI, P., *Calcolo delle probabilità*, McGraw - Hill, 2011.

[2] DEN HOLLANDER, F., *Large deviations*, Mathematical Institute Nijmegen University, Olanda, 1991.