APPUNTI DI PROCESSI STOCASTICI

Fabio Antonelli

Capitolo 1

RICHIAMI DI PROBABILITÀ

1.1 Indipendenza

Definizione 1.1.1 : *Uno* **spazio di probabilità** *è una terna* $(\Omega, \mathcal{F}, P)$*, dove*

1. Ω *è un insieme (spazio degli eventi)*

2. $\mathcal{F}$ *è una* $\sigma-$*algebra su* Ω*, ovvero è una famiglia di sottoinsiemi di* Ω *tale che*

   (a) $\emptyset, \Omega \in \mathcal{F}$;

   (b) *se* $A \in \mathcal{F}$*, allora anche* $A^c \in \mathcal{F}$;

   (c) *se* $\{A_i\}_{i \geq 1}$ *è una famiglia al più numerabile di insiemi di* $\mathcal{F}$*, allora anche* $\bigcup_{i \geq 1} A_i \in \mathcal{F}$.

3. $P : \mathcal{F} \longrightarrow [0, 1]$ *è una probabilità, ovvero una funzione di insieme tale che*

   (a) $P(\Omega) = 1$;

   (b) *se* $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathcal{F}$*,* $A_i \cap A_j = \emptyset$ *per ogni* $i \neq j$*, allora*

$$P\Big(\bigcup_{i \in \mathbb{N}} A_i\Big) = \sum_{i \in \mathbb{N}} P(A_i)$$

Se Ω è costituito da una quantità al più numerabile di elementi, si dice che è uno spazio discreto di probabilità. In particolare, se $\mathrm{card}(\Omega) = n < +\infty$, allora di solito si assume l'insieme delle parti di Ω, $\mathcal{P}(\Omega)$, come la $\sigma-$algebra di riferimento.

Si può dimostrare facilmente che l'intersezione di una qualsiasi famiglia di $\sigma-$algebre (anche infinita) è ancora una $\sigma-$algebra.

Se $\mathcal{A}$ è una famiglia di insiemi di Ω, viene detta la $\sigma-$algebra generata da $\mathcal{A}$, la più piccola $\sigma-$algebra contenente $\mathcal{A}$ e si indica con $\sigma(\mathcal{A})$, ovvero

$$\sigma(\mathcal{A}) = \bigcap\{\mathcal{F},\, \sigma - \text{algebra},\, \mathcal{A} \subseteq \mathcal{F}\}$$

Osservazione 1.1.1 : *usando la proprietà 3b, si ha $P(A^c) = 1 - P(A)$, da cui si ottiene*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \forall\, A, B \in \mathcal{F}$$
$$P\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) = 1 - P\Big(\bigcup_{n\in\mathbb{N}} A_n^c\Big) \quad \forall\, \{A_n\} \subseteq \mathcal{F}$$

Definizione 1.1.2 : *siano $A, B \in \mathcal{F}$ con $P(B) > 0$, si dice* **probabilità di A condizionata a B**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Definizione 1.1.3 : *Due eventi A e B si dicono* **indipendenti** *se*

$$P(A \cap B) = P(A)P(B).$$

Una famiglia di eventi $\{A_i\}_{i\in I}$ si dice **indipendente** *se comunque scelti un intero $k \leq card(\Omega)$ e $A_{i_1}, \dots A_{i_k}$ insiemi distinti della famiglia, risulta*

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

$\{A_i\}_{i\in I}$ *è una famiglia di eventi* **indipendenti a due a due** *se per ogni $i \neq j$ risulta*

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

Mentre l'indipendenza di una famiglia di eventi implica direttamente l'indipendenza a due a due della stessa, il contrario non è vero come mostra il seguente esempio

Esempio 1.1.1 : *Consideriamo uno spazio di probabilità equiprobabile costituito da 4 elementi $\Omega = \{1, 2, 3, 4\}$ ed i seguenti eventi $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, $A_3 = \{1, 4\}$. Allora si ha*

$$P(A_i) = \frac{1}{2} \quad \forall\, i = 1, 2, 3\,; \quad P(A_i \cap A_j) = \frac{1}{4} = P(A_i)P(A_j) \quad \forall\, i \neq j = 1, 2, 3$$

ma $P(A_1 \cap A_2 \cap A_3) = \dfrac{1}{4} \neq P(A_1)P(A_2)P(A_3) = \dfrac{1}{8}$.

Sia $\{E_i\}_{i=1}^n$ una partizione finita dello spazio di probabilità Ω e sia $A \in \mathcal{F}$, allora vale la formula delle probabilità totali

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i).$$

Osserviamo che la formula vale anche se $P(E_i) = 0$, qualora si definisce un valore arbitrario per $P(A|E_i)$.

Vale infine la Formula di Bayes per qualsiasi A, B tale che $P(A), P(B) > 0$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

Se A e B sono eventi indipendenti allora $P(A|B) = P(A)$. Ricordiamo infine che eventi disgiunti non sono indipendenti.

1.2 Variabili Aleatorie

Sia $\mathcal{B}(\mathbb{R})$ la $\sigma-$algebra dei Boreliani su $\mathbb{R}$ ovvero la $\sigma-$algebra generata da tutti gli aperti di $\mathbb{R}$. Una funzione $X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è detta misurabile se per ogni $S \in \mathcal{B}(\mathbb{R})$ si ha che $X^{-1}(S) \in \mathcal{F}$. Una variabile aleatoria è una funzione misurabile.

Ogni variabile aleatoria induce una probabilità sullo spazio $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, data da

$$P_X(S) = P(X^{-1}(S)), \quad \forall\, S \in \mathcal{B}(\mathbb{R}).$$

In questa maniera la terna $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ è uno spazio di probabilità, che è anche uno spazio metrico, mentre lo spazio di partenza non è necessariamente dotato di alcuna struttura topologica. La probabilità P_X su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è detta **legge** della variabile aleatoria X.

Poichè gli insiemi aperti sono generati dai semiintervalli illimitati della retta reale e gli aperti generano la $\sigma-$algebra $\mathcal{B}(\mathbb{R})$, per caratterizzare la legge di X basta valutare

$$F_X(t) = P_X((-\infty, t]) = P(X^{-1}((-\infty, t])) = P(X \leq t) \quad \forall\, t \in \mathbb{R}.$$

La funzione F_X è detta **funzione di ripartizione** di X.

Proprietà della funzione di ripartizione F_X:

1. $\lim_{t \to +\infty} F_X(t) = 1\,; \quad \lim_{t \to -\infty} F_X(t) = 0;$
2. è non decrescente;
3. è continua a destra con limiti a sinistra.

Osservazione 1.2.1 : *Le tre proprietà della funzione di ripartizione implicano che F_X può avere al più una quantità numerabile di discontinuità di salto.*

Ogni volta che la funzione di ripartizione presenta un salto in un punto t_0, la probabilità che $X = t_0$ è non nulla e pari a $F_X(t_0) - F_X(t_0-)$, in questo caso si dice che t_0 è un atomo per la legge di X; nel caso in cui invece la funzione di ripartizione risulti continua su tutta la retta reale, allora si dice che essa è non atomica.

Definizione 1.2.1 : *Se una funzione di ripartizione F_X è derivabile in $\mathbb{R}$, allora la sua derivata, f_X, è detta* **funzione di densità** *di X e chiaramente gode delle seguenti due proprietà:*

1. $f_X(t) \geq 0$, *per ogni* $t \in \mathbb{R}$;
2. $\displaystyle\int_{\mathbb{R}} f_X(t)dt = 1$.

Definizione 1.2.2 : *una famiglia finita di variabili aleatorie $\{X_i\}_{i=1}^n$ si dice* **indipendente** *se*

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \ldots P(X_n \in B_n) \tag{1.1}$$

per ogni scelta di $B_1, \ldots B_n \in \mathcal{B}(\mathbb{R})$.

Qualora la famiglia fosse illimitata, per avere indipendenza è necessario che la condizione (1.1) sia verificata per ogni $n \in \mathbb{N}$.

Definizione 1.2.3 : *Sia X una variabile aleatoria, la $\sigma-$algebra generata da essa e' data da*

$$\sigma(X) = \sigma(\{X^{-1}(S), S \in \mathcal{B}(\mathbb{R})\}).$$

La definizione 1.2.2 vuol dire che le $\sigma-$algebre generate da ognuna delle variabili aleatorie sono indipendenti, ovvero X ed Y variabili aleatorie indipendenti è equivalente a dire che se $\mathcal{A} = \sigma(X)$ e $\mathcal{G} = \sigma(Y)$, allora

$$P(A \cap G) = P(A)P(G), \quad \forall\, A \in \mathcal{A},\, G \in \mathcal{G}.$$

1.3 Lemma di Borel-Cantelli

In quest'ultima sezione introduciamo due lemmi fondamentali in teoria della misura e di largo impiego nei capitoli seguenti

Definizione 1.3.1 *Sia $\{A_n\}_{n\in\mathbb{N}}$ una successione di eventi in uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, definiamo*

$$\limsup_n A_n = \bigcap_{n\geq 1} \bigcup_{k=n}^{+\infty} A_k, \qquad \liminf_n A_n = \bigcup_{n\geq 1} \bigcap_{k=n}^{+\infty} A_k$$

Per le proprietà di $\sigma-$algebra, $\limsup_n A_n$, $\liminf_n A_n$ appartengono a $\mathcal{F}$.

Per essi valgono le seguenti proprietà

(a) $\limsup_n A_n = (\liminf_n A_n^c)^c$;

(b) $P(\liminf_n A_n) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P(\limsup_n A_n)$

Lemma 1.3.1 : *(Primo Lemma di Borel-Cantelli)*
Sia $\{A_n\}_{n\in\mathbb{N}}$ una successione di eventi in uno spazio di probabilità $(\Omega, \mathcal{F}, P)$. Se $\sum_{n\geq 1} P(A_n) < +\infty$, allora $P(\limsup_n A_n) = 0$.

Dimostrazione: Per definizione e per subadditività si ha per ogni $n \in \mathbb{N}$

$$P\Big(\bigcap_{n\geq 1} \bigcup_{k=n}^{+\infty} A_k\Big) \leq P\Big(\bigcup_{k=n}^{+\infty} A_k\Big) \leq \sum_{k\geq n} P(A_k),$$

ma per avere la convergenza della serie necessariamente si deve avere

$$\sum_{k\geq n} P(A_k) \longrightarrow 0,$$

da cui la tesi. $\square$

Osserviamo che se $\omega \in \limsup_n A_n$, allora per definizione $\omega \in \bigcup_{k\geq n} A_k$ per ogni $n \in \mathbb{N}$, dunque $\omega \in A_k$ per qualche $k \geq n$ per ogni scelta di $n \in \mathbb{N}$. In altre parole, questo vuol dire che si può trovare una successione di indici $\{k_n\}$ tale che $\omega \in A_{k_n}$ per ogni $n \in \mathbb{N}$, ovvero l'elemento ω appartiene agli insiemi della successione per infiniti valori dell'indice, cioè cade nella successione infinitamente spesso (i.s.). Una notazione alternativa è quindi $\limsup_n A_n = \{\omega \in A_k,\, i.s.\}$. Analogamente, se abbiamo una successione di v.a. $\{X_n\}$ ed una successione di insiemi di Borel $\{B_n\}$ in $\mathbb{R}$, scriviamo

$$\{X_n \in B_n,\, i.s.\} = \limsup_n \{\omega :\ X_n \in B_n\}.$$

1.4 Esercizi capitolo 1

1. Sia Ω un insieme con card(Ω) =n, dimostrare che $\mathcal{P}(\Omega)$ è una $\sigma-$algebra.

2. Dimostrare l'osservazione 1.1.1.

3. Dimostrare le proprietà della funzione di ripartizione.

4. Dimostrare che l'intersezione di una famiglia di $\sigma-$algebre è ancora una $\sigma-$algebra.

5. Dimostrare che la $\sigma-$algebra generata da una variabile aleatoria è una $\sigma-$algebra.

6. Dimostrare che se $\{A_n\}$ è una famiglia di insiemi di $\mathcal{F}$ tale che per ogni $n \in \mathbb{N}$ $A_n \subseteq A_{n+1}$ ($A_{n+1} \supseteq A_n$) e si pone $A = \bigcup_{n\in\mathbb{N}} A_n$ ($A = \bigcap_{n\in\mathbb{N}} A_n$), allora

$$P(A) = \lim_{n\in\mathbb{N}} P(A_n).$$

7. Sia $\{A_n\}$ è una famiglia di insiemi di $(\Omega, \mathcal{F}, P)$, dimostrare che gli insiemi $\limsup_n A_n$, $\liminf_n A_n$ appartengono a $\mathcal{F}$. Dimostrare inoltre
   (a) $\limsup_n A_n = (\liminf_n A_n^c)^c$;
   (b) $P(\liminf_n A_n) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P(\limsup_n A_n)$

8. Dimostrare la formula delle probabilità totali e la formula di Bayes.

9. Per qualsiasi successione di insiemi $\{A_n\}_n \subseteq \mathcal{F}$ dimostrare che

$$P(\bigcup_{n=1}^{+\infty} A_n) \leq \sum_{n=1}^{+\infty} P(A_n).$$

Capitolo 2

CATENE DI MARKOV

2.1 Preliminari e Definizioni

Le variabili aleatorie servono per descrivere un fenomeno aleatorio statico, i processi stocastici invece descrivono fenomeni aleatori dinamici, ovvero con un'evoluzione temporale.

Definizione 2.1.1 : *Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e $T \subseteq \mathbb{R}^+$; $\{X_t\}_{t \in T}$ è detto processo stocastico se per ogni $t \in T$ fissato, $X_t(\cdot)$ è una variabile aleatoria e per ogni $\omega \in \Omega$ fissato, l'applicazione $X_.(\omega) : T \longrightarrow \mathbb{R}$ è una funzione Borel misurabile, che viene detta* **traiettoria** *del processo.*

L'insieme T può essere discreto o continuo, in quel che segue sarà sempre discreto e usualmente assimilato a $\mathbb{N}$.

Rovina del giocatore - 1 .

Due persone A e B giocano a carte, il primo con capitale iniziale a e l'altro con capitale iniziale b. Ad ogni partita il giocatore A ha probabilità p di vincere indipendentemente da tutte le altre partite. Il gioco si ferma quando uno dei due giocatori va in fallimento.

Teoricamente il gioco può andare avanti all'infinito, ma procede comunque in maniera discreta di partita in partita, dunque $T = \mathbb{N}$.

Qual è il processo stocastico sottostante?

Come modellizzare la successione di partite?

Gli elementi a nostra disposizione per rappresentare il fenomeno sono p, $1-p$ ed $m = a + b$ il capitale totale.

Affinchè il gioco termini ad un certa partita, è necessario che il capitale detenuto da uno dei due giocatori risulti zero a quella partita (e conseguentemente m per l'altro giocatore). Indichiamo con X_n = capitale di A dopo la partita n, chiaramente X_n

può prendere i valori $0, 1, \ldots, m$ (per ogni $n \in \mathbb{N}$). Se $X_n = 0$ si ottiene la rovina di A al tempo n, se $X_n = m$ si ottiene la rovina di B.

Come possiamo descrivere la dinamica di gioco?

Lo possiamo fare per passi, se $1 \leq i \leq m-1$

$$\begin{aligned}
P(X_{n+1} = i+1 | X_n = i) &= p, \\
P(X_{n+1} = i-1 | X_n = i) &= q = 1-p, \\
P(X_{n+1} = j | X_n = i) &= 0, \quad \forall\, j \neq i-1, i+1
\end{aligned}$$

altrimenti

$$\begin{aligned}
P(X_{n+1} = j | X_n = 0) &= 0, \forall\, j \neq 0, \quad P(X_{n+1} = 0 | X_n = 0) = 1 \\
P(X_{n+1} = j | X_n = m) &= 0, \forall\, j \neq m, \quad P(X_{n+1} = m | X_n = m) = 1,
\end{aligned}$$

ovvero abbiamo descritto come passare da una configurazione di gioco ad un'altra in un passo.

Supponiamo che dopo la partita n il capitale di A sia pari ad $m-1$, a questo punto sappiamo che al prossimo giro vi e' probabilità p di terminare il gioco ed intuitivamente comprendiamo che non ha nessuna rilevanza se al livello n ci si sia arrivati vincendo tutte le partite oppure ritornando al capitale iniziale per cinque volte, od in qualsiasi altra maniera. Possiamo dire che una volta che siamo in una certa configurazione di gioco, quel che verrà dopo dipende unicamente dalla configurazione corrente e non dal modo in cui ci si è arrivati.

Definizione 2.1.2 : *Una* **catena di Markov** *è un processo stocastico su* $(\Omega, \mathcal{F}, P)$ *tale che*

1. $T \subseteq \mathbb{N}$ *(a meno di isomorfismi);*
2. *le variabili aleatorie* $\{X_n\}_{n \in T}$ *assumono tutte valori nello stesso spazio* E *discreto, detto* **spazio degli stati** *(per comodità supporremo che* $E \subseteq \mathbb{N}$*);*
3. *Per ogni* $i_1, \ldots, i_{n-1}, i, j \in E$*, vale la* **proprietà di Markov***:*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1) = P(X_{n+1} = j | X_n = i) = p_{ij}(n)$$

L'ultima proprietà della definizione vuol dire che la probabilità che il sistema si trovi in un certo stato al tempo $n+1$ non dipende da tutta la storia passata, ma soltanto dal sistema all'istante precedente. Le probabilità $\{p_{ij}(n)\}_{i,j \in E}$ sono dette **probabilità di transizione ad un passo al tempo** n. Se risulta che le $p_{ij}(n)$ sono indipendenti da n, allora si dice che la catena di Markov è **omogenea**. In quel che segue tratteremo solamente catene di Markov omogenee.

Osservazione 2.1.1 : *La proprietà di Markov* **non implica** *che le variabili aleatorie che la costituiscono siano indipendenti, infatti*

$$P(X_2 = j, X_1 = i) = P(X_2 = j|X_1 = i)P(X_1 = i) = p_{ij}P(X_1 = i)$$

e dalla definizione non vi è nessun motivo per cui $p_{ij} = P(X_2 = j)$.

D'altra parte questa proprietà si può riesprimere dicendo che il passato ed il futuro sono indipendenti dato il presente, per esempio

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

dove nel penultimo passaggio si è utilizzata la proprietà di Markov.

Se $\mathrm{card}(E) = m < +\infty$, allora la famiglia delle probabilità di transizione dà luogo ad una matrice $m \times m$, $\mathcal{P} = (p_{ij})_{i,j\in E}$ detta **matrice di transizione** e tale che

1. $p_{ij} \geq 0$ per ogni $i, j \in E$;
2. $\sum_{j\in E} p_{ij} = \sum_{j\in E} P(X_{n+1} = j|X_n = i) = 1$ come probabilità condizionata.

Una matrice che rispetti le precedenti due proprietà in generale si dice **matrice stocastica**.

La matrice di transizione è un elemento fondamentale di una catena di Markov, poichè permette di trasporre molti dei problemi riguardanti le dinamiche di Markov in termini di algebra lineare, con la conseguente potenza di calcolo che essa implica.

Rovina del giocatore - 2

Nel caso del precedente esempio della rovina del giocatore ci rendiamo subito conto che lo spazio degli stati è $E = \{0, 1, \ldots, m\}$ e che la matrice di transizione associata è data da

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & . & . & . & . & 0 \\ q & 0 & p & . & . & . & 0 \\ 0 & q & 0 & p & . & . & 0 \\ \vdots & & & & & & \\ 0 & . & . & . & q & 0 & p \\ 0 & 0 & . & . & . & . & 1 \end{pmatrix}$$

Per rappresentare un fenomeno aleatorio che gode della proprietà di Markov abbiamo quindi bisogno della matrice di transizione e di un ulteriore ingrediente: la configurazione iniziale in cui si trova il sistema.

Assumendo 0 come il tempo iniziale, chiamiamo

$$\pi_i^0 = P(X_0 = i), \quad i \in E$$

la **distribuzione iniziale** della catena. Per esempio, nel caso della rovina del giocatore la distribuzione iniziale è data da

$$\pi_a^0 = 1, \pi_i^0 = 0, \quad \forall\, i \neq a.$$

A questo punto è chiaro che data una catena di Markov si identificano facilmente la matrice di transizione associata e la distribuzione iniziale; questi sono strumenti importanti per ricavare informazioni sulle variabili aleatorie che costituiscono la catena.

Una prima domanda importante da porsi è come calcolare le distribuzioni marginali della successione

$$\pi_i^n = P(X_n = i) \quad \text{e} \quad P(X_{n_1} = i_1, \dots, X_{n_k} = i_k).$$

Notiamo che se conosciamo la distribuzione iniziale e la matrice di transizione, allora sfruttando la formula delle probabilità totali per la v.a. X_1 si ottiene

$$\pi_j^1 = P(X_1 = j) = \sum_{i \in E} P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_{i \in E} p_{ij} \pi_i^0 \qquad j \in E$$

ovvero in notazione matriciale

$$(\pi^1)^T = (\pi^0)^T \mathcal{P}$$

e iterando la relazione precedente si ottiene

$$(\pi^n)^T = (\pi^0)^T \mathcal{P}^n. \tag{2.1}$$

Gli elementi della matrice $\mathcal{P}^n$ rappresentano le probabilità di transizione ad n passi

$$p_{ij}^{(n)} := P(X_{k+n} = j | X_k = i),$$

infatti, grazie alla proprietà di Markov, otteniamo

$$\begin{aligned}
P(X_{k+n} = j | X_k = i) &= \frac{P(X_{k+n} = j, X_k = i)}{P(X_k = i)} \\
&= \frac{\sum_{h \in E} P(X_{k+n} = j, X_{k+n-1} = h, X_k = i)}{P(X_k = i)} \\
&= \sum_{h \in E} P(X_{k+n} = j | X_{k+n-1} = h, X_k = i) P(X_{k+n-1} = h | X_k = i) \\
&= \sum_{h \in E} p_{ih}^{(n-1)} p_{hj}
\end{aligned}$$

ed iterando ci si rende conto che le probabilità di transizione ad n passi sono proprio gli elementi della $n-$esima potenza della matrice $\mathcal{P}$.

Analogamente si possono ricostruire tutte le distribuzioni congiunte marginali mediante le potenze della matrice di transizione e la distribuzione iniziale: per $0 < n_1 < n_2 < \cdots < n_k$

$$\begin{aligned} & P(X_{n_1} = i_1, \ldots, X_{n_k} = i_k) \\ = \;& P(X_{n_1} = i_1, \ldots, X_{n_{k-1}} = i_{k-1}) P(X_{n_k} = i_k | X_{n_{k-1}} = i_{k-1}, \ldots, X_{n_1} = i_1) \\ = \;& P(X_{n_1} = i_1, \ldots, X_{n_{k-1}} = i_{k-1}) p^{(n_k - n_{k-1})}_{i_{k-1} i_k} \\ \vdots & \\ = \;& \sum_{k \in E} \pi^0_k p^{(n_1)}_{k i_1} \ldots p^{(n_k - n_{k-1})}_{i_{k-1} i_k}. \end{aligned}$$

Tutte queste sono proprietà distribuzionali del processo e osserviamo che due catene di Markov distinte (che descrivono fenomeni diversi) possono avere le stesse matrice di transizione e distribuzione iniziale e, conseguentemente, le stesse distribuzioni marginali.

D'altra parte vale anche il converso nel senso del

Teorema 2.1.1 (di Kolmogorov) *(vedere [B])*

Date una matrice di transizione $\mathcal{P}$ su uno spazio degli stati E ed una legge di probabilità su E, π^0, allora esistono

(i) uno spazio di probabilità $(\Omega, \mathcal{F}, P)$;

(ii) una successione di variabili aleatorie $\{X_n\}_{n \geq 0}$ definite su $(\Omega, \mathcal{F}, P)$ a valori in E, tali che $P(X_0 = k) = \pi^0_k$ e $\{X_n\}_{n \geq 0}$ è una catena di Markov con matrice di transizione $\mathcal{P}$.

Esempi 2.1.1 :

1. *Un esempio generale di catena di Markov è dato dalla seguente successione*

$$X_0, \quad X_n = f(X_{n-1}, \xi_n)$$

dove X_0 è una variabile aleatoria data, $\{\xi_n\}_n$ una successione di variabili aleatorie indipendenti ed a loro volta indipendenti da X_0 e f è una funzione Boreliana di due variabili. Infatti

$$\begin{aligned} & P(X_n = k | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ = \;& P(f(X_{n-1}, \xi_n) = k | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ = \;& P(f(i_{n-1}, \xi_n) = k | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ = \;& P(f(i_{n-1}, \xi_n) = k) = P(f(i_{n-1}, \xi_n) = k | X_{n-1} = i_{n-1}) \\ = \;& P(X_n = k | X_{n-1} = i_{n-1}), \end{aligned}$$

dove gli ultimi passaggi si sono ottenuti per ricorrenza.

È quindi chiaro che se $\{\xi_n\}_n$ *è una successione di variabili aleatorie indipendenti allora*

$$X_n = \xi_1 + \dots \xi_n, \quad X_n = \xi_1 \dots \xi_n$$

sono esempi di catene di Markov.

2. *Nei lanci ripetuti di una moneta si vince se si ottengono sei teste consecutive. Questo gioco è rappresentato da una catena di Markov (se si individua opportunamente lo spazio degli stati) e quindi si può scrivere la matrice di transizione associata. La dinamica è di Markov poichè l'unica cosa che conta è il numero di teste consecutive uscite fino al lancio precedente e non il numero di lanci effettuati per ottenerle (ovvero non conta la storia passata). Infatti se* $X_n =$ *il numero di teste consecutive ottenute fino al tempo* n *e*

$$\xi_n = \begin{cases} 1 & se\ T \\ -5 & se\ C \end{cases}$$

al tempo n*, allora la successione* $\{\xi_n\}$ *è indipendente e*

$$X_n = f(X_{n-1}, \xi_n), \qquad f(x,y) = \begin{cases} \max(x+y, 0), & per\ x \leq 5 \\ 6 & per\ x \geq 6, \end{cases}$$

verificando quanto esposto nell'esempio precedente. Lo spazio degli stati in questo caso è $E = \{0, 1, 2, 3, 4, 5, 6\}$ *e le probabilità di transizione sono date da*

$$\begin{aligned} P(X_n = i+1 | X_{n-1} = i) &= \frac{1}{2}, \qquad \forall 1 = 0, \dots 5 \\ P(X_n = 0 | X_{n-1} = i) &= \frac{1}{2}, \qquad \forall 1 = 0, \dots 5 \\ P(X_n = 6 | X_{n-1} = 6) &= 1 \end{aligned}$$

e conseguentemente la matrice di transizione associata è

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

3. *Una funzione di una catena di Markov non è necessariamente una catena di Markov. Consideriamo una catena di Markov* $\{X_n\}$ *con matrice di transizione*

$$\mathcal{P}_X = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{pmatrix}$$

e distribuzione iniziale $P(X_0 = 1) = 1$ *e si definisca la successione di variabili aleatorie*

$$Y_n = \begin{cases} 1 & X_n = 1 \\ 2 & X_n = 2, 3 \end{cases}$$

allora quest'ultima non rappresenta una catena di Markov (se lo fosse, sarebbe omogenea come la catena di partenza), infatti

$$\begin{aligned} P(Y_2 = 1 | Y_1 = 2) &= P(X_2 = 1 | X_1 = 2, 3) \\ &= \frac{P(X_2 = 1, X_1 = 2, 3)}{P(X_1 = 2, 3)} \\ &= \frac{p_{12}p_{21} + p_{13}p_{31}}{p_{12} + p_{13}} = \frac{11}{60} \end{aligned}$$

mentre

$$\begin{aligned} P(Y_3 = 1 | Y_2 = 2, Y_1 = 2) &= \frac{P(X_3 = 1, X_2 = 2, 3, X_1 = 2, 3)}{P(X_2 = 2, 3, X_1 = 2, 3)} \\ &= \frac{p_{12}p_{22}p_{21} + p_{12}p_{23}p_{31} + p_{13}p_{32}p_{21} + p_{13}p_{33}p_{31}}{p_{12}p_{22} + p_{12}p_{23} + p_{13}p_{32} + p_{13}p_{33}} \\ &= \frac{131}{735} \end{aligned}$$

Per le catene di Markov finite (ed anche numerabili) si può usare un metodo di rappresentazione molto comodo, associando un grafo, dove ogni stato rappresenta un nodo del grafo e si disegna un ramo uscente dal nodo i al nodo j se $p_{ij} \neq 0$. Chiaramente la distribuzione iniziale deve essere assegnata separatamente.

Esempi 2.1.2 :

1. *Rovina del giocatore:*

2. *Per il gioco delle sei teste consecutive nel lancio di una moneta, si ottiene il seguente grafo*

2.2 Esercizi

1. Si abbia una catena di Markov con spazio degli stati E e siano $i, j \in E$. Dimostrare che se esistono due interi $n_1(i)$ e $n_2(j)$ tali che $p_{ii}^{(n_1)} > 0$ e $p_{jj}^{(n_2)} > 0$, scegliendo $n = m.c.m.(n_1, n_2)$ allora si ha necessariamente $p_{ii}^{(n)} > 0$ e $p_{jj}^{(n)} > 0$.

2. Per una catena di Markov $\{X_n\}$, con matrice di transizione $\mathcal{P}$ e distribuzione iniziale π_0, calcolare $P(X_1 = j | X_3 = i)$ usando π_0 e le probabilità di transizione.

3. Per una catena di Markov $\{X_n\}_n$, dimostrare che per ogni $n \in \mathbb{N}$, X_{n+1}, X_{n-1} sono indipendenti condizionatamente a X_n.

4. La successione di v.a. X_n a valori in $\{0, 1, 2\}$ è definita nel modo seguente:

$$X_0 = 1$$

   Per ogni $n \geq 1$ si lancia una moneta equa

$$X_{n+1} = \begin{cases} (X_n + 1)\text{mod}3 & \text{se esce testa} \\ 1 & \text{altrimenti} \end{cases}$$

   (a) Dimostrare che si tratta di una CM.
   (b) Calcolare la probabilità dell'evento $\{X_1 = 1, X_3 = 1\}$.
   (c) Calcolare le probabilità di transizione.

2.3 Classificazione degli stati

Abbiamo visto che una catena di Markov è completamente identificata da $\mathcal{P}$ e π^0, quindi una volta passato l'istante iniziale, l'unica cosa che governa l'evoluzione del sistema dinamico è la matrice di transizione.

È quindi importante capire le proprietà della matrice $\mathcal{P}$ e come queste si traducono in proprietà degli stati o, per meglio dire, dell'evoluzione del sistema. La dinamica della catena di Markov è dunque caratterizzata da come gli stati sono collegati (in senso di grafo) tra di loro.

D'ora in poi indicheremo con $P_i(\cdot) = P(\cdot | X_0 = i)$.

Definizione 2.3.1 : *per ogni $i, j \in E$ diciamo che i* **comunica** *con j (e scriviamo $i \to j$) se esiste un $n \in \mathbb{N}$ tale che $p_{ij}^{(n)} > 0$, ovvero se abbiamo probabilità positiva di passare dallo stato i allo stato j in un numero finito di passi.*

La proprietà di comunicare è transitiva, infatti se $i \to j$ e $j \to h$ allora anche $i \to h$, poichè esistono due interi $m(i,j)$, $n(j,h)$ tali che $p_{ij}^{(m)} > 0$ e $p_{jh}^{(n)} > 0$ e dunque

$$p_{ih}^{(m+n)} = \sum_{k=1}^{|E|} p_{ik}^{(m)} p_{kh}^{(n)} \geq p_{ij}^{(m)} p_{jh}^{(n)} > 0$$

essendo $\mathcal{P}^{m+n} = \mathcal{P}^m \mathcal{P}^n$.

Definizione 2.3.2 : *un sottoinsieme $C \subseteq E$ si dice una* **classe chiusa** *se gli stati in C non comunicano con gli stati in C^c, ovvero se*

$$\sum_{j \in C} p_{ij} = 1 \qquad \forall\, i \in C.$$

Se $C = \{i\}$ è chiusa allora si dice che lo stato i è **assorbente**.

L'ultima definizione vuol dire che una volta che si arriva nella classe C ad un qualche tempo, non se ne esce più.

Esempio 2.3.1 : *Consideriamo la catena di Markov rappresentata dal seguente grafo*

ovvero dalla seguente matrice stocastica

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

In questo caso, come evidenzia il grafo, la classe $\{4,5,6\}$ è chiusa (verificare la definizione). Notiamo che nella matrice $\mathcal{P}$ è evidenziata la sottomatrice stocastica

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

in corrispondenza delle righe e colonne 4,5,6.

Chiaramente se consideriamo l'intero spazio degli stati E, questo è banalmente una classe chiusa.

Definizione 2.3.3 : *Una classe chiusa $C \subseteq E$ è detta* **irriducibile** *se tutti gli stati ad essa appartenenti comunicano tra loro, ovvero se per ogni $i, j \in C$ esiste un intero $n = n(i, j)$ tale che $p_{ij}^{(n)} > 0$. Se E risulta essere l'unica classe irriducibile, allora si dice che la catena di Markov è* **irriducibile**.

I concetti di chiusura ed irriducibilità per i sottoinsiemi dello spazio degli stati sono diversi. Per esempio E è automaticamente una classe chiusa, ma non è detto che sia anche irriducibile, come dimostra l'esempio 2.3.1.

Vogliamo sottolineare che se due stati i e j comunicano tra loro, questo vuol dire che in effetti esistono due interi $n_1(i, j)$ e $n_2(j, i)$ tali che $p_{ij}^{(n_1)} > 0$ e $p_{ji}^{(n_2)} > 0$, ma scegliendo $n = m.c.m.(n_1, n_2)$ allora si ha necessariamente $p_{ij}^{(n)} > 0$ e $p_{ji}^{(n)} > 0$ (verificare). Questo ragionamento può essere esteso ad un qualsiasi numero finito di stati, ma a priori non ad una quantità numerabile di stati

Esempi 2.3.1 :

1. *Catena di Markov con un'unica classe chiusa e irriducibile non banale*

2. *Catena di Markov irriducibile*

3. *Stato assorbente*

Per capire la dinamica delle catene di Markov è importante individuare le classi chiuse ed irriducibili, con quale probabilità ci si può finire, così come il tempo medio che si impiega ad arrivare in esse. Per formalizzare il nostro discorso, data una CM $\{X_n\}_n$, introduciamo le variabili aleatorie (non negative)

$$\tau_j = \min_{n \geq 1}\{X_n = j\}, \qquad j \in E$$

ognuna delle quali definisce il tempo di primo arrivo della catena nello stato j. Come vedremo in seguito, queste v.a. sono esempi di **tempi di arresto**.

In generale $\tau_j = 1, 2, \ldots$, con la convenzione che $\tau_j = +\infty$ sull'insieme $\{\omega : X_n(\omega) \neq j \, \forall n \in \mathbb{N}\}$.

Indichiamo con

$$\rho_{ij} = P_i(\tau_j < +\infty)$$

la probabilità di passare per la prima volta per lo stato j partendo da i. Quindi ρ_{ii} sarà detta la probabilità di primo ritorno in i.

Definizione 2.3.4 : *Se $\rho_{ii} < 1$, lo stato i è detto* **transiente**. *Se $\rho_{ii} = 1$ lo stato i è detto* **persistente** *o* **ricorrente**

Notiamo che se i è uno stato assorbente, allora necessariamente $\rho_{ii} = 1$, poichè $p_{ii} = 1$, che è come dire che $P_i(\tau_i = 1) = 1$.

Come menzionato precedentemente, oltre a calcolare ρ_{ii}, ci interessa calcolare anche il tempo medio di ritorno nello stato i, definito da $E_i(\tau_i)$, ovvero ci interessa capire se esso è integrabile oppure no. Osserviamo che se i è uno stato transiente allora, per definizione $P_i(\tau_i = +\infty) > 0$, perciò si ha immediatamente che $E_i(\tau_i) = +\infty$. Invece l'integrabilità di τ_i induce un'ulteriore distinzione tra gli stati persistenti.

Definizione 2.3.5 : *Se $\rho_{ii} = 1$, lo stato persistente i è detto*

(a) **persistente positivo** *se $E_i(\tau_i) < +\infty$;*

(b) **persistente nullo** *se $E_i(\tau_i) = +\infty$.*

La dizione "persistente" è intuitiva, infatti se si torna con sicurezza nello stato i una volta, vi si deve tornare infinite volte. Grazie alla proprietà di Markov una volta che si è arrivati in i, il passato può essere dimenticato e la catena può essere "riinizializzata". La dinamica che governa la catena ($\mathcal{P}$) è sempre la stessa, quindi con sicurezza si tornerà nello stato i una seconda volta e così via. Analogamente, per gli stati transienti, vi è una probabilità maggiore di zero di non tornare mai più nello stato. Questo è quanto dimostra la seguente

Proposizione 2.3.1 : *Sia $\{X_n\}_n$ una CM con spazio degli stati E. Allora per ogni $i \in E$ vale*

$$P_i(X_n = j,\ i.s.) = \begin{cases} 0 & sse\ \ \rho_{jj} < 1, \quad ovvero\ sse\ \ j\ \text{è}\ transiente \\ \rho_{ij} & sse\ \ \rho_{jj} = 1, \quad ovvero\ sse\ \ j\ \text{è}\ persistente \end{cases}$$

Dimostrazione: Notiamo che vale la seguente sequenza di uguaglianze

$$\begin{aligned} P_i(X_n = j,\ \text{i.s.}) &= P_i(\exists\,\text{infiniti}\,k:\ X_k = j) \\ &= P_i(\text{la successione}\,\{X_n\}_n\,\text{visita}\ j\ \text{infinite volte}) \\ &= \lim_{N\to+\infty} P_i(\text{la successione}\,\{X_n\}_n\,\text{visita}\ j\ \text{almeno}\ N\,\text{volte}). \end{aligned}$$

dove l'ultima uguaglianza è vera poichè se chiamiamo A_N l'evento { la successione $\{X_n\}_n$ visita j almeno N volte }, allora $A_{N+1} \subseteq A_N$ e $\lim\limits_{N\to+\infty} P_i(A_N) = P(\bigcap\limits_N A_N)$.

Analogamente se chiamiamo

$$B_k^N = \{\text{la successione}\,\{X_n\}_n\,\text{visita}\ j\ \text{almeno}\ N\,\text{volte entro}\ k\,\text{passi}\}$$

allora $B_k^N \subseteq B_{k+1}^N$ e $\lim\limits_{k\to+\infty} P_i(B_k^N) = P_i(\bigcup\limits_k B_k^N) = P_i(A_N)$.

D'altra parte

$$\begin{aligned} &P_i(B_k^N) = P_i(\exists\, k_1, \ldots, k_N \le k : X_{k_1} = j, \ldots, X_{k_N} = j) \\ &= \sum_{k_1,..,k_N=1}^{k} P_i(X_1 \neq j, .., X_{k_1-1} \neq j, X_{k_1} = j, X_{k_1+1} \neq j, .., X_{k_N-1} \neq j, X_{k_N} = j) \end{aligned}$$

Grazie alla proprietà di Markov si ha

$$\begin{aligned} &P_i(X_1 \neq j, \ldots, X_{k_1} = j, X_{k_1+1} \neq j, \ldots, X_{k_2} = j) \\ &= P_i(X_{k_2} = j, \ldots, X_{k_1+1} \neq j | X_{k_1} = j, \ldots, X_1 \neq j) P_i(X_{k_1} = j, \ldots, X_1 \neq j) \\ &= P_i(X_{k_2} = j, \ldots, X_{k_1+1} \neq j | X_{k_1} = j) P_i(X_{k_1} = j, \ldots, X_1 \neq j) = \rho_{jj}^{(k_2-k_1)} \rho_{ij}^{(k_1)}, \end{aligned}$$

dove abbiamo definito

$$\rho_{ij}^{(k)} := P_i(X_k = j, \ldots, X_1 \neq j) = P_i(\text{primo arrivo in j in esattamente k passi}).$$

Dunque iterando si ottiene

$$\begin{aligned} P_i(B_k^N) &= \sum_{k_1,\ldots,k_N=1}^{k} \rho_{ij}^{(k_1)} \rho_{jj}^{(k_2-k_1)} \ldots \rho_{jj}^{(k_N-k_{N-1})} \\ P_i(A_N) &= \lim_{k\to+\infty} \sum_{k_1=1}^{k} \cdots \sum_{k_N=1}^{k} \rho_{ij}^{(k_1)} \rho_{jj}^{(k_2-k_1)} \ldots \rho_{jj}^{(k_N-k_{N-1})} \\ &= \lim_{k\to+\infty} \sum_{k_1=1}^{k} \rho_{ij}^{(k_1)} \sum_{k_2=1}^{k} \rho_{jj}^{(k_2-k_1)} \cdots \sum_{k_N=1}^{k} \rho_{jj}^{(k_N-k_{N-1})}. \end{aligned}$$

Ogni somma nel precedente prodotto è positiva, dunque si può applicare la regola del prodotto nel limite e per costruzione abbiamo

$$\lim_{k\to+\infty}\sum_{k_1=1}^{k}\rho_{ij}^{(k_1)}=\rho_{ij},\qquad \lim_{k\to+\infty}\sum_{k_h=1}^{k}\rho_{jj}^{(k_h)}=\rho_{jj}\quad h=2,\dots N$$

e quindi che

$$P_i(A_N)=\rho_{ij}\rho_{jj}^{N-1}$$

da cui

$$P_i(X_n=j,\ \text{i.s.})=\lim_{N\to+\infty}P_i(A_N)=\begin{cases}\rho_{ij} & \text{sse}\quad \rho_{jj}=1\\ 0 & \text{sse}\quad \rho_{jj}<1\end{cases}$$

ovvero la tesi. $\square$

Verificare la definizione di stato persistente o transiente non è facile, introduciamo quindi alcuni criteri che ci aiuteranno nel compito. Il primo è una conseguenza della precedente proposizione.

Corollario 2.3.1 :
i è persistente sse $P_i(X_n=i,\ i.s.)=1$ sse $\sum_{n=1}p_{ii}^{(n)}=+\infty$
i è transiente sse $P_i(X_n=i,\ i.s.)=0$ sse $\sum_{n=1}p_{ii}^{(n)}<+\infty$

Dimostrazione: La proposizione 2.3.1 dimostra il primo *se e soltanto se* sia nel caso della persistenza sia della transienza, ponendo $i=j$.

Dimostriamo il secondo *se e soltanto se* nel caso della transienza, questo lo implicherà anche nel caso della persistenza.

Supponiamo che $\sum_{n=1}^{+\infty}p_{ii}^{(n)}<+\infty$. Per definizione $p_{ii}^{(n)}=P_i(X_n=i)$, dunque il Lemma di Borel-Cantelli implica che

$$P_i(X_n=i,\,i.s.)=P_i(\limsup_n\{X_n=i\})=0$$

grazie al primo *sse* , ciò vuol dire che i è transiente.

Viceversa, sia i transiente, notiamo che vale la seguente **equazione di rinnovo**

$$\begin{aligned}p_{ij}^{(n)} &= P_i(X_n=j)=P_i(X_n=j,1\le\tau_j\le n)=\sum_{k=0}^{n-1}P_i(X_n=j,\tau_j=n-k)\\ &= \sum_{k=0}^{n-1}P_i(X_n=j,X_{n-k}=j,X_{n-k-1}\neq j,\dots,X_1\neq j)\\ &= \sum_{k=0}^{n-1}\rho_{ij}^{(n-k)}p_{jj}^{(k)},\end{aligned}$$

da cui otteniamo per ogni j, sommando fino ad un indice m

$$\begin{aligned}
(2.2) \qquad \sum_{n=1}^{m} p_{ij}^{(n)} &= \sum_{n=1}^{m}\sum_{k=0}^{n-1} p_{jj}^{(k)}\rho_{ij}^{(n-k)} = \sum_{k=0}^{m-1} p_{jj}^{(k)} \sum_{n=k+1}^{m} \rho_{ij}^{(n-k)} \\
&\leq \sum_{k=0}^{m-1} p_{jj}^{(k)} \sum_{h=1}^{+\infty} \rho_{ij}^{(h)} \leq \sum_{k=0}^{m} p_{jj}^{(k)}\rho_{ij} = p_{jj}^{(0)}\rho_{ij} + \rho_{ij}\sum_{k=1}^{m} p_{jj}^{(k)}.
\end{aligned}$$

Poichè $p_{jj}^{(0)} = 1$ per convenzione, possiamo concludere per qualsiasi m

$$\sum_{n=1}^{m} p_{ij}^{(n)} \leq \rho_{ij} + \rho_{ij}\sum_{k=1}^{m} p_{jj}^{(k)}.$$

Prendendo $i = j$ nella precedente, poichè $\rho_{ii} < 1$, si può risolvere la disuguaglianza ottenendo per ogni m

$$\sum_{n=1}^{m} p_{ii}^{(n)} \leq \frac{\rho_{ii}}{1-\rho_{ii}}.$$

Spingendo $m \to +\infty$ nella precedente si ha la tesi. □

Teorema 2.3.1 : *Sia $\{X_n\}_n$ una CM irriducibile. Allora*

(a) tutti gli stati sono transienti e $\sum_{n=1}^{+\infty} p_{ij}^{(n)} < +\infty$ *per ogni* i, j.

(b) tutti gli stati sono persistenti e $\sum_{n=1}^{+\infty} p_{ij}^{(n)} = +\infty$ *per ogni* i, j.

Dimostrazione: Se $\{X_n\}_n$ è una CM irriducibile, allora per ogni $i, j \in E$ esistono due indici k, h tali che $p_{ij}^{(k)} > 0$, $p_{ji}^{(h)} > 0$, da cui per ogni $n \in \mathbb{N}$ vale

$$\begin{aligned}
p_{ii}^{(n+k+h)} &= \sum_{l=1}^{|E|} p_{li}^{(n+h)} p_{il}^{(k)} \geq p_{ij}^{(k)} p_{ji}^{(n+h)} \\
&= p_{ij}^{(k)} \sum_{m=1}^{|E|} p_{jm}^{(n)} p_{mi}^{(h)} \geq p_{ij}^{(k)} p_{ji}^{(h)} p_{jj}^{(n)}.
\end{aligned}$$

Prendendo le somme su n, le due serie hanno lo stesso carattere.

Per la seconda parte, notiamo che dalla (2.2) si ottiene la convergenza di $\sum_{l=1}^{n} p_{ij}^{(l)}$ nel caso transiente, mentre nel caso persistente notiamo che se k è tale che $p_{ij}^{(k)} > 0$, allora per ogni $n > k$ si ha $p_{ij}^{(n)} \geq p_{ij}^{(k)} p_{ii}^{(n-k)}$. □

D'ora in poi indicheremo con T il sottoinsieme degli stati transienti dello spazio degli stati E.

Osservazioni 2.3.1 :

1. *Se E ha cardinalità finita, allora deve esistere almeno uno stato persistente, infatti se così non fosse avremmo per ogni n*

$$\sum_{j \in T} p_{ij}^{(n)} = \sum_{j \in E} p_{ij}^{(n)} = 1, \tag{2.3}$$

d'altra parte per il criterio di transienza $\sum_{n \geq 1} p_{ij}^{(n)} < +\infty$ per ogni $j \in T = E$, e per avere la convergenza della serie necessariamente si ha

$$\lim_{n \to +\infty} p_{ij}^{(n)} = 0 \qquad \forall\, j.$$

Poichè $E = T$ è uno spazio finito, si possono scambiare limite e somma nella seguente espressione

$$0 = \sum_{j \in T} \lim_{n \to +\infty} p_{ij}^{(n)} = \lim_{n \to +\infty} \sum_{j \in T} p_{ij}^{(n)} \tag{2.4}$$

e le uguaglianze (2.3) e (2.4) sono in chiara contraddizione.

2. *La relazione "i comunica con j e j comunica con i", che denotiamo con $i \leftrightarrow j$, è una relazione di equivalenza sullo spazio degli stati persistenti, che indichiamo con J, quindi essa ripartisce J in classi di equivalenza disgiunte $C_1, C_2, \dots$ che sono chiuse e irriducibili per costruzione.*

   Dunque lo spazio degli stati può essere decomposto come

$$E = T \cup C_1 \cup C_2 \cup \dots$$

Ogni classe chiusa ed irriducibile si comporta come una catena di Markov irriducibile (sottomatrice stocastica), quindi individuare le classi chiuse ed irriducibili è fondamentale per poter semplificare l'analisi della dinamica della CM.

Il criterio introdotto dal teorema 2.3.1 per studiare il carattere degli stati non è immediato, ne introduciamo uno più agevole per verificare la transienza

Proposizione 2.3.2 : *Scelto $j \in E$, se esiste uno stato i tale che j comunica con i e i non comunica con j, allora j è transiente.*

Dimostrazione: Sia j tale che esiste un i con cui comunica, ma che non comunica con esso e sia $m = \inf\{n\,;\, p_{ji}^{(n)} > 0\}$ il minimo numero di passo necessario per andare da j ad i. Poichè $p_{ji}^{(m)} > 0$, devono esistere $i_1, \ldots, i_{m-1} \neq i$ (per la minimalità di m) tali che

$$p_{ji_1} p_{i_1 i_2} \ldots p_{i_{m-1} i} > 0$$

od in altre parole l'insieme $A = \{X_1 = i_1, \ldots, X_{m-1} = i_{m-1}, X_m = i\}$ ha $P_j(A) > 0$. Per disgiunzione

$$\begin{aligned} A &= (A \cap \{\tau_j < +\infty\}) \cup (A \cap \{\tau_j = +\infty\}), \\ P_j(A) &= P_j(A \cap \{\tau_j < +\infty\}) + P_j(A \cap \{\tau_j = +\infty\}). \end{aligned}$$

Notiamo che su A, $\tau_j \geq m + 1$, dunque

$$P_j(A \cap \{\tau_j < +\infty\}) = \sum_{k=1}^{+\infty} P_j(A \cap \{\tau_j = k\}) \sum_{k=m+1}^{+\infty} P_j(A \cap \{\tau_j = k\}).$$

D'altronde, per $k \geq m + 1$, $\{\tau_j = k\} \subseteq \{X_k = j\}$ e

$$\begin{aligned} P_j(A \cap \{\tau_j = k\}) &= P_j(\{\tau_j = k\}|A) P_j(A) \\ &= P_j(\tau_j = k | X_1 = i_1, \ldots, X_{m-1} = i_{m-1}, X_m = i) P_j(A) \\ &\leq P_j(X_k = j | X_1 = i_1, \ldots, X_{m-1} = i_{m-1}, X_m = i) P_j(A) \\ &= P_j(X_k = j | X_m = i) P_j(A) = p_{ij}^{(k-m)} P_j(A) = 0 \end{aligned}$$

dove nelle ultime due uguaglianze abbiamo usato la proprietà di Markov ed il fatto che $p_{ij}^{(h)} = 0$ per ogni h, poichè i non comunica con j. Possiamo quindi concludere che $P_j(A) = P_j(A \cap \{\tau_j = +\infty\}$, ovvero che j è transiente. $\square$

Esempi 2.3.2 :

1. *Rovina del Giocatore: Sia $|E| = m$ e supponiamo che ad ogni partita si possa vincere con probabilità p e perdere con probabilità $q = 1 - p$. Quindi il grafo associato è dato da*

   Chiaramente $\{0\}$ ed $\{m\}$ sono gli unici stati assorbenti, mentre tutti gli altri sono transienti, poichè per ogni $i \neq 0, m$, i comunica con 0 ed m, ma non è vero il viceversa. La catena ovviamente non è irriducibile.

2. *Sia* $\{X_n\}$ *una CM a valori in* $\{0,1\}$ *con*

$$\mathcal{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

La CM è irriducibile con spazio degli stati finito, dunque tutti gli stati sono persistenti (positivi).

Verifichiamo che la definizione. Si ha $P_i(\tau_i < +\infty) = 1$*, poichè*

$$P_1(\tau_1 = 1) = 1-q, \qquad P_1(\tau_1 = n) = q(1-p)^{n-2}p, \quad n \geq 2$$

dunque

$$P_1(\tau_1 < +\infty) = 1 - q + \sum_{n=2}^{+\infty} q(1-p)^{n-2}p = 1$$

3. *Si ha la seguente CM a 5 stati rappresentata dal grafo*

Dal grafo è evidente che gli stati 1 e 2 comunicano con 3, ma non viceversa, mentre 3 4 5 comunicano tra loro, perciò $\{3,4,5\}$ *costituisce l'unica classe chiusa ed irriducibile e* $\{1,2\}$ *è l'insieme dei transienti. Se scriviamo la matrice di transizione*

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

allora ci rendiamo conto che questa è una matrice a blocchi $\mathcal{P} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$*, con* C *sottomatrice stocastica. Quindi se si calcolano le potenze della matrice di transizione, queste risulteranno del tipo*

$$\mathcal{P}^n = \begin{pmatrix} A^n & f_n(A,B,C) \\ 0 & C^n \end{pmatrix}$$

per qualche polinomio f_n. In altre parole $p_{ij}^n = 0$ per ogni n se $j \in \{3,4,5\}$ e $i \in \{1,2\}$. Dimostriamo esplicitamente che 2 è transiente

$$\begin{aligned} P_2(\tau_2 = 1) &= \frac{1}{3}, \quad P_2(\tau_2 = n) = \frac{1}{2}(\frac{1}{2})^{n-2}\frac{1}{3}, \quad n \geq 2 \\ P_2(\tau_2 < +\infty) &= \frac{1}{3} + \sum_{n \geq 2} \frac{1}{2}(\frac{1}{2})^{n-2}\frac{1}{3} = \frac{2}{3} < 1 \end{aligned}$$

4. *Nel caso in cui lo spazio degli stati abbia cardinalità numerabile, si può avere una CM con tutti gli stati transienti. Consideriamo il grafo*

rappresentato anche dalla matrice di dimensione infinita

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & \dots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Allora la catena ha infinite classi chiuse, ma nessuna irriducibile. Gli stati comunicano a due a due e quindi a due a due devono avere lo stesso carattere. La dimostrazione è analoga per ogni stato dispari e perciò dimostriamo che 1 è transiente.

$$P_1(\tau_1 = 1) = \frac{1}{2}, \quad P_1(\tau_1 = 2) = \frac{1}{4}, \quad P_1(\tau_1 = n) = 0 \quad \forall\, n \geq 3$$

perciò $P_1(\tau_1 < +\infty) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$ e $P_1(\tau_1 = +\infty) = \frac{1}{4} > 0$.

D'altronde ogni stato pari comunica con un dispari e quindi anch'essi sono transienti.

Nel seguito vedremo che è possibile cotruire una CM irriducibile con tutti gli stati transienti, se lo spazio degli stati è numerabile. Classificare gli stati di una CM a stati numerabili è più complesso. Presentiamo un ulteriore criterio che può essere utilizzato soprattutto per queste catene.

Teorema 2.3.2 : *Sia $\{X_n\}_n$ una CM irriducibile con spazio degli stati E e matrice di transizione $\mathcal{P} = (p_{ij})_{i,j\in E}$. Allora essa è persistente (transiente) se e soltanto se per ogni stato j_0 fissato arbitrariamente il sistema*

$$\begin{cases} x_i = \sum_{j\neq j_0} p_{ij}x_j, & i \neq j_0 \\ 0 \leq x_i \leq 1, & \forall\, i \in E \end{cases} \tag{2.5}$$

ammette solo la soluzione nulla (una soluzione non nulla).

Dimostrazione: Denotiamo con $\mathbf{x} = (x_1, x_2, \dots)$ il generico vettore (successione) nello spazio lineare $\mathbb{R}^{|E|}$, analogamente $\mathcal{P}$ è una matrice in $\mathbb{R}^{|E|\times|E|}$.

Consideriamo un generico sottoinsieme U dello spazio degli stati e una matrice sottostocastica definita su di esso, ovvero $\mathcal{Q} = (q_{ij})_{i,j\in U}$ tali che

1. $q_{ij} \geq 0$ per ogni $i, j \in U$;
2. $\displaystyle\sum_{j\in U} q_{ij} \leq 1$ per ogni $i \in U$.

Vogliamo dimostrare che la soluzione del sistema lineare indotto dalla matrice $\mathcal{Q}$ rappresenta la probabilità di rimanere nell'insieme U a partire da un qualsiasi stato in U.

Se $\mathcal{Q}^n$ è la $n-$ma potenza della matrice $\mathcal{Q}$, denotiamo con $q_{ij}^{(n)}$ il suo generico elemento di posto i, j e sia

$$\sigma_i^{(n)} = \sum_{j\in U} q_{ij}^{(n)}, \qquad i \in U. \tag{2.6}$$

La successione $\{\sigma_i^{(n)}\}_n$ è chiaramente nonnegativa ed inoltre monotona decrescente. Infatti per le proprietà di una matrice sottostocastica abbiamo $\sigma_i^{(1)} \leq 1$ per ogni i e

$$\sigma_i^{(n+1)} = \sum_{j\in U} q_{ij}^{(n+1)} = \sum_{j\in U}\sum_{k\in U} q_{ik}^{(n)} q_{kj} = \sum_{k\in U}\sum_{j\in U} q_{kj} q_{ik}^{(n)} = \sum_{k\in U} \sigma_k^{(1)} q_{ik}^{(n)} \leq \sum_{k\in U} q_{ik}^{(n)} = \sigma_i^{(n)},$$

dove le sommatorie si sono potute scambiare grazie alla non negatività di ogni singolo termine. Come successione monotona crescente di reali, deve esistere un limite finito per $\{\sigma_i^{(n)}\}_n$ che indichiamo con σ_i. Passando al limite nelle seguenti uguaglianze otteniamo

$$\begin{array}{lll} \sigma_i^{(n+1)} & = \sum_{j\in U}\sum_{k\in U} q_{ik}q_{kj}^{(n)} = \sum_{k\in U} q_{ik}\sum_{j\in U} q_{kj}^{(n)} & = \displaystyle\sum_{k\in U} q_{ik}\sigma_k^{(n)} \\ \downarrow & & \qquad\downarrow \\ \sigma_i & & = \displaystyle\sum_{k\in U} q_{ik}\sigma_k. \end{array}$$

Dunque $(\sigma_1, \sigma_2, \dots)$ è una soluzione del sistema associato a $\mathcal{Q}$ e tutte le componenti sono limitate, $0 \le \sigma_i \le 1$. Inoltre è anche la soluzione massimale, infatti supponiamo che $(x_1, x_2, \dots)$, $0 \le x_i \le 1$, sia un'ulteriore soluzione del sistema associato a $\mathcal{Q}$, allora si ha

$$x_i = \sum_{k \in U} q_{ik} x_k \le \sum_{k \in U} q_{ik} = \sigma_i^{(1)}$$

e per induzione si dimostra (per esercizio) che $x_i \le \sigma_i^{(n)}$ per ogni n, da cui $x_i \le \sigma_i$. D'altra parte, da (2.6), procedendo per iterazione si vede che

$$\sigma_i^{(n)} = P_i(X_1 \in U, \dots, X_n \in U), \qquad i \in U$$

e dunque spingendo n ad infinito si ottiene $\sigma_i = P_i(X_n \in U, \forall\, n)$.

La dimostrazione è completa scegliendo $U = E \setminus \{j_0\}$. In questo caso $\sigma_i = P_i(\text{non toccare mai } j_0) = 1 - \rho_{ij_0}$ per $i \ne j_0$.

Perciò si ha l'unica (per massimalità) soluzione triviale $\sigma_i = 0, \forall\, i \ne j_0$ se e soltanto se $\rho_{ij_0} = 1$ ovvero i (e quindi tutti la catena) è persistente.

Dimostriamo quest'ultima affermazione. La proposizione 2.3.1 implica che $\rho_{ij_0} > 0$ se e soltanto se $\rho_{ii} = 1$. Dunque $\rho_{ij_0} = 1$ implica che lo stato i (e quindi per irriducibilità tutta la catena) debba essere persistente.

Viceversa se $\{X_n\}_n$ è una CM irriducibile persistente, $\rho_{ii} = 1$ per ogni $i \in E$ e dimostriamo che anche $\rho_{ij} = 1$ per ogni $i, j \in E$. Per irriducibilità deve esistere un indice m tale che $p_{ji}^{(m)} > 0$ e poichè $\rho_{jj} = P_j(X_n = j, i.s.) = 1$ si ha

$$\begin{aligned}
p_{ji}^{(m)} &= P_j(X_m = i) = P_j(X_m = i, \{X_n = j, i.s.\}) \\
&= P_j(\{X_m = i\} \cap \bigcap_{N \ge 1} \bigcup_{n \ge N} \{X_n = j\}) \le P_j(\{X_m = i\} \cap \bigcup_{n > m} \{X_n = j\}) \\
&= P_j(\{X_m = i\} \cap \{\exists\, \text{almeno un } n > m : X_n = j\} \\
&= P_j(\{X_m = i\} \cap \bigcup_{n=m+1}^{\infty} \{n = \inf\{k : X_k = j\}\}) \\
&= \sum_{n > m} P_j(X_m = i, X_{m+1} \ne j, \dots, X_{n-1} \ne j, X_n = j) \\
&= \sum_{n > m} p_{ji}^{(m)} \rho_{ij}^{(n-m)} = p_{ji}^{(m)} \rho_{ij}.
\end{aligned}$$

da cui $\rho_{ij} = 1$ necessariamente.

Quanto detto (essendo un sse) ovviamente implica anche $\sigma_i > 0$, per qualche $i \ne j_0$ se e soltanto se $\rho_{ij_0} < 1$, ovvero sse la catena è transiente. $\square$

2.4 Esercizi

1. Dare la definizione di stato persistente nullo, persistente positivo e di stato transiente per una catena di Markov.

2. Dare la definizione di classe chiusa per una catena di Markov, con spazio degli stati E. Dare la definizione di classe chiusa ed irriducibile pe una CM e dare un esempio di una classe chiusa che non sia irriducibile.

3. Dimostrare che $i \leftrightarrow j$ è una relazione di equivalenza sullo spazio degli stati persistenti e che lo ripartisce in classi chiuse e irriducibili disgiunte.

4. Sia $\{X_n\}$ una CM a valori in $\{0,1\}$ con
$$\mathcal{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$
Dimostrare che $E_1(\tau_1) < +\infty$.

5. Per la CM con matrice di transizione
$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & \dots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \dots \\ \dots & \dots & \dots & \dots & & & \end{pmatrix}.$$
dimostrare che gli stati pari sono transienti.

6. Sia $\{X_n\}$ una CM con matrice di transizione
$$\mathcal{P} = \begin{pmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 & 0 \\ \frac{2}{12} & 0 & \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{3}{12} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{12} & \frac{3}{12} & \frac{1}{12} & \frac{1}{12} & 0 & \frac{6}{12} \\ 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{pmatrix}.$$
Si disegni il grafo associato e si classifichino gli stati.

7. Vi sono due urne U_1, U_2 e 4 palline, di cui due bianche e due nere. Inizialmente le palline vengono disposte a caso nelle due urne, due per urna. Quindi si lancia una moneta equa e se esce Testa allora due palline vengono scmbiate tra U_1 ed U_2, se invece esce Croce la situazione si lascia invariata.

   Sia X_n il numero di palline nere in U_1 dopo l'$n-$mo lancio.

   (a) dimostrare che $\{X_n\}_n$ è una CM;

   (b) disegnarne il grafo e scriverne la matrice di transizione;

   (c) classificare gli stati.

8. Sia $\{X_n\}_n$ una catena di Markov con spazio degli stati $\mathbb{N}$ con le seguenti probabilità di transizione non nulle

$$p_{i0} = \frac{1}{3}, \quad p_{i,i+1} = \frac{2}{3}, \quad \forall\ i = 0, 1, ...$$

   Classificare il carattere degli stati.

9. Una ditta produce 4 tipi di componenti di un'apparecchiatura. Ogni anno il committente, sulla base della riuscita dei componenti utilizzati l'anno prima, decide quanti tipi di componenti riordinare. Se il committente e' stato completamente insoddisfatto dalla riuscita (non ordina nessun tipo di componente), chiude i rapporti con la ditta e cambia fornitore. Le probabilita' di transizione da uno stato ad un altro per $i = 1, 2, 3, 4$ sono

$$\begin{aligned}
p_{10} &= \frac{1}{2}, p_{11}\frac{1}{4}, p_{12} = \frac{1}{8}, p_{13} = \frac{1}{8}, p_{14} = 0 \\
p_{20} &= \frac{1}{4}, p_{21}\frac{1}{8}, p_{22} = \frac{1}{4}, p_{23} = \frac{1}{4}, p_{14} = \frac{1}{8} \\
p_{30} &= \frac{1}{8}, p_{31}\frac{1}{8}, p_{32} = \frac{1}{4}, p_{33} = \frac{1}{4}, p_{14} = \frac{1}{4} \\
p_{40} &= \frac{1}{8}, p_{41}\frac{1}{8}, p_{42} = \frac{1}{4}, p_{43} = \frac{3}{8}, p_{44} = \frac{1}{8}
\end{aligned}$$

   (a) Stabilire le probabilità di transizione a partire da 0, quindi scrivere la matrice delle probabilità di transizione e classificare gli stati.

   (b) Supponendo che il primo anno alla ditta siano stati commissionati tutti e quattro i tipi di componenti, calcolare la probabilità che il terzo anno il committente decida di cambiare fornitore.

10. Dimostrare che in una catena di Markov persistente $\rho_{ij} = 1$ per ogni $i, j \in E$.

2.5 Problemi di assorbimento

In questa sezione assumeremo che $|E| < +\infty$. Se prendiamo una classe chiusa $C \subseteq E$, allora per definizione questo implica che se $X_k \in C$ anche $X_n \in C$ per ogni $n \geq k$ con $P = 1$. In altre parole una volta che la catena giunge in una classe chiusa non può più uscirne.

Perciò viene naturale chiedersi se vi sono dei metodi per calcolare le probabilità di finire in una qualsiasi classe chiusa a partire da un qualche stato. Data una classe chiusa C, denotiamo con

$$\lambda_i = P_i(X_n \in C, \text{ per qualche } n > 0).$$

Trivialmente se $i \in C$ allora $\lambda_i = 1$ e se $i \in C'$ con C' altra classe chiusa disgiunta da C, $\lambda_i = 0$, poiché

$$\lambda_i = \sum_{j \in C} \sum_{n \geq 1} p_{ij}^{(n)} = 0, \quad \text{essendo } \; p_{ij}^{(n)} = 0, \; \forall\, n, fori \in C'.$$

Dunque la nostra domanda diventa interessante solo a partire dagli stati transienti.

Per esempio nel problema della rovina del giocatore, la rovina è rappresentata dallo stato assorbente $\{0\}$. È quindi naturale voler calcolare $\lambda_i = P_i($ rovina del giocatore) a partire da qualsiasi posizione di gioco i ed è ovvio che, come abbiamo detto precedentemente, $\lambda_0 = 1$ e $\lambda_m = 0$.

In generale, fissata una classe chiusa C, vale il seguente risultato

Teorema 2.5.1 : *Siano $\{X_n\}_n$ una CM tale che lo spazio degli stati abbia $|E| < +\infty$, $C \subseteq E$ una classe chiusa e si denoti con $T = \{$stati transienti che non si trovano in $C\}$ (se C è irriducibile T è costituita da tutti gli stati transienti).*

Allora il sistema lineare

$$x_i = \sum_{j \in C} p_{ij} + \sum_{r \in T} p_{ir} x_r, \qquad i \in T \tag{2.7}$$

ha come unica soluzione $x_i = \lambda_i$, dove λ_i è la probabilità di assorbimento in C partendo da i.

Dimostrazione: Procediamo per passi, dimostrando prima che le λ_i sono soluzione di (2.7) e quindi che la soluzione del sistema è unica.

Passo 1: Sia $\tau_C = \inf\{n > 0 \,:\, X_n \in C\}$, ovvero il primo istante che la CM si trova in C, quindi $\lambda_i = P_i(\tau_C < +\infty)$.

Poniamo $g_i^{(n)} = P_i(X_1 \in T, \ldots, X_{n-1} \in T, X_n \in C)$, allora $g_i^{(n)} = P_i(\tau_C = n)$. Conseguentemente

$$\lambda_i = P_i(\tau_C < +\infty) = \sum_{n=1}^{+\infty} P_i(\tau_C = n) = \sum_{n=1}^{+\infty} g_i^{(n)}.$$

D'altra parte,

$$g_i^{(n)} = P_i(X_1 \in T, \dots, X_{n-1} \in T, X_n \in C) = \sum_{\substack{h_1, \dots, h_{n-1} \in T \\ j \in C}} p_{ih_1} p_{h_1 h_2} \dots p_{h_{n-1} j}.$$

Dalla proprietà di Markov e dal prodotto righe per colonne si ha anche

$$\begin{aligned} g_i^{(n+1)} &= \sum_{\substack{r, h_1, \dots, h_{n-1} \in T \\ j \in C}} p_{ir} p_{rh_1} p_{h_1 h_2} \dots p_{h_{n-1} j} \\ &= \sum_{r \in T} p_{ir} \sum_{\substack{h_1, \dots, h_{n-1} \in T \\ j \in C}} p_{rh_1} p_{h_1 h_2} \dots p_{h_{n-1} j} = \sum_{r \in T} p_{ir} g_r^{(n)}, \end{aligned}$$

ma $g_i^{(1)} = \sum_{j \in C} p_{ij}$ e quindi

$$\begin{aligned} \lambda_i &= \sum_{n=1}^{+\infty} g_i^{(n)} = g_i^{(1)} + \sum_{n=2}^{+\infty} g_i^{(n)} = g_i^{(1)} + \sum_{n=2}^{+\infty} \sum_{r \in T} p_{ir} g_r^{(n-1)} = g_i^{(1)} + \sum_{r \in T} \sum_{n=2}^{+\infty} p_{ir} g_r^{(n-1)} \\ &= g_i^{(1)} + \sum_{r \in T} p_{ir} \sum_{n=2}^{+\infty} g_r^{(n-1)} = \sum_{j \in C} p_{ij} + \sum_{r \in T} p_{ir} \sum_{n=1}^{+\infty} g_r^{(n)} = \sum_{j \in C} p_{ij} + \sum_{r \in T} p_{ir} \lambda_r, \end{aligned}$$

dove abbiamo potuto scambiare le somme poichè una di esse è estesa su un insieme finito.

Per provare l'unicità, supponiamo di avere un'altra soluzione di (2.7): $\{x_i\}_{i \in T}$ con $0 \le x_i \le 1, \ \forall i$. Sostituendo per x_r si ha

$$\begin{aligned} x_i &= \sum_{j \in C} p_{ij} + \sum_{r \in T} p_{ir} \Big(\sum_{j \in C} p_{rj} + \sum_{h \in T} p_{rh} x_h \Big) \\ &= \sum_{j \in C} p_{ij} + \sum_{r \in T} \sum_{j \in C} p_{ir} p_{rj} + \sum_{r \in T} \sum_{h \in T} p_{ir} p_{rh} x_h \\ &= P_i(\tau_C = 1) + P_i(\tau_C = 2) + \sum_{r \in T} \sum_{h \in T} p_{ir} p_{rh} x_h = P_i(\tau_C \le 2) + \sum_{r \in T} \sum_{h \in T} p_{ir} p_{rh} x_h, \end{aligned}$$

dove di nuovo abbiamo potuto distribuire le somme, poiché gli addendi sono in numero finito. Notiamo che se r è uno stato persistente e h uno transiente allora necessariamente $p_{rh}^{(n)} = 0, \forall n$, infatti se così non fosse comunicherebbero tra loro e quindi dovrebbero avere lo stesso carattere, dunque per h transiente si ha

$$p_{ih}^{(2)} = \sum_{r \in E} p_{ir} p_{rh} = \sum_{r \in T} p_{ir} p_{rh}$$

perció scambiando le somme nel terzo termine della precedente uguaglianza si ottiene

$$\begin{aligned} x_i &= P_i(\tau_C \leq 2) + \sum_{h\in T}\sum_{r\in T} p_{ir}p_{rh}x_h = P_i(\tau_C \leq 2) + \sum_{h\in T}\sum_{r\in E} p_{ir}p_{rh}x_h \\ &= P_i(\tau_C \leq 2) + \sum_{h\in T} p_{ih}^{(2)} x_h. \end{aligned}$$

Iterando n volte, possiamo scrivere

$$x_i = P_i(\tau_C \leq n) + \sum_{h\in T} p_{ih}^{(n)} x_h, \tag{2.8}$$

ma h è transiente, quindi per il teorema 2.3.1, sappiamo che $\sum_{n\geq 1} p_{ih}^{(n)} < +\infty$ e la condizione necessaria per la convergenza delle serie ci assicura che $p_{ih}^{(n)} \longrightarrow 0$ per $n \to +\infty$ per ogni $h \in T$ fissato. Poiché $0 \leq x_h \leq 1$ per ogni $h \in T$ ed il numero degli stati transienti è finito, passando al limite in (2.8) possiamo concludere

$$x_i = P_i(\tau_C < +\infty) + 0 = \lambda_i$$

□

Esempi 2.5.1 :

1. **Catene di nascita e morte.** *Sia $E = \{0, 1, \ldots, m\}$, sul quale è definita una CM tramite la seguente matrice di transizione*

$$\mathcal{P} = \begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & \ldots & 0 \\ q_1 & r_1 & p_1 & 0 & 0 & \ldots & 0 \\ 0 & q_2 & r_2 & p_2 & 0 & \ldots & 0 \\ \vdots & \vdots & & & & & \\ 0 & 0 & 0 & \ldots & q_{m-1} & r_{m-1} & p_{m-1} \\ 0 & 0 & 0 & \ldots & 0 & q_m & r_m \end{pmatrix}$$

Vengono chiamate catene di nascita e morte perché si può pensare che X_n rappresenti la numerosità di una popolazione al tempo n, la cui massima dimensione è m e ad ogni istante può nascere o morire al più un individuo con probabilità dipendenti dalla dimensione della popolazione. Le p_i sono le probabilità di nascita, mentre le q_i sono le probabilità di morte, chiaramente $q_0 = 0 = p_m$ e per ogni $i \in E$, $q_i + r_i + p_i = 1$.

Osserviamo che la rovina del giocatore è un caso particolare di catena di nascita e morte con

$$r_0 = 1, \quad r_i = 0 \quad \forall\, 1 \leq i \leq m-1, \quad r_m = 1, \quad p_0 = 0, q_m = 0.$$

Con questa notazione lo stato 0 *è assorbente se e soltanto se* $p_0 = 0$*, così come* m *lo è se e soltanto se* $q_m = 0$*. In questo caso tutti gli altri stati sono transienti.*

Vogliamo calcolare la probabilità di assorbimento in $C = \{0\}$

$$\lambda_i = P_i(X_n = 0, \text{ per qualche } n) \qquad 0 \le i \le m.$$

Sappiamo che $\lambda_0 = 1$ *e* $\lambda_m = 0$ *e dal teorema 2.5.1, per* $i = 1, \ldots, m-1$*, il seguente sistema deve essere soddisfatto*

$$\begin{cases} \lambda_1 = \sum_{j \in C} p_{1j} + \sum_{j \in T} p_{1j}\lambda_j = q_1 + r_1\lambda_1 + p_1\lambda_2 \\ \lambda_2 = q_2\lambda_1 + r_2\lambda_2 + p_2\lambda_3 \\ \vdots \\ \lambda_k = q_k\lambda_{k-1} + r_k\lambda_k + p_k\lambda_{k+1} \\ \vdots \\ \lambda_{m-1} = q_{m-1}\lambda_{m-2} + r_{m-1}\lambda_{m-1} \end{cases}$$

Ricordando che $r_i = 1 - q_i + p + i$ *e risolvendo il precedente sistema a cascata, otteniamo*

$$\begin{cases} 1 = \lambda_1 - \dfrac{p_1}{q_1}(\lambda_2 - \lambda_1) \\ \vdots \\ \lambda_k = \lambda_{k+1} + \dfrac{p_{k+1}}{q_{k+1}}(\lambda_{k+1} - \lambda_{k+2}) \\ \vdots \\ \lambda_{m-2} = \Big(1 + \dfrac{p_{m-1}}{q_{m-1}}\Big)\lambda_{m-1} \end{cases}$$

ovvero sottraendo la seguente dalla precedente

$$\lambda_k - \lambda_{k+1} = \frac{q_k}{p_k}(\lambda_{k-1} - \lambda_k), \qquad 1 \le k \le m-1,$$

iterando si ottiene

$$\lambda_k - \lambda_{k+1} = \frac{q_k}{p_k}\frac{q_{k-1}}{p_{k-1}} \cdots \frac{q_1}{p_1}(\lambda_0 - \lambda_1) := \gamma_k(1 - \lambda_1), \tag{2.9}$$

d'altra parte sommando da 0 ad $m-1$

$$1 = \lambda_0 - \lambda_m = \sum_{k=0}^{m-1}(\lambda_k - \lambda_{k+1}) = (1 - \lambda_1)\sum_{k=0}^{m-1}\gamma_k$$

e risolvendo per λ_1 *possiamo concludere che*

$$\lambda_1 = 1 - \frac{1}{\sum_{k=0}^{m-1} \gamma_k},$$

dove per convenzione abbiamo posto $\gamma_0 = 1$.

Analogamente sommando (2.9) da i *ad* $m-1$ *e risolvendo per* λ_i *si ottiene*

$$\lambda_i = \frac{\sum_{k=i}^{m-1} \gamma_k}{\sum_{k=0}^{m-1} \gamma_k}.$$

Le precedenti formule danno quindi le probabilità di assorbimento per una catena di nascita e morte e nel caso particolare della rovina del giocatore, dove $r_i = 0$, $p_i = p$, $q_i = 1-p$, $\forall\, i$, *si ha* $\gamma_k = (\frac{1-p}{p})^k$ *e*

$$\lambda_i = \frac{\sum_{k=i}^{m-1} (\frac{1-p}{p})^k}{\sum_{k=0}^{m-1} (\frac{1-p}{p})^k}$$

e se $p = \frac{1}{2}$,

$$\lambda_i = \frac{m-i}{m} = \frac{a+b-i}{a+b}, \quad \lambda_a = \frac{a}{a+b}$$

ovvero la probabilità di vincita del giocatore A è uguale a quanto dista relativamente il giocatore B dalla posta in gioco.

2. *Si consideri la CM con il seguente grafo*

Si tratta di una CM irriducibile a stati finiti, dunque tutti gli stati sono persistenti positivi.

Vogliamo calcolare

$$P_1(\textit{arrivare in 6 senza passare da 5 })$$

come possiamo usare quanto sappiamo delle probabilità di assorbimento se vi è un'unica classe chiusa?

Poichè quando arriviamo in 5 o 6 la dinamica della nostra catena perde di interesse, ai fini della nostra domanda considerare la CM precedente è equivalente a considerare la CM modificata con grafo

dove abbiamo reso gli stati 5 e 6 assorbenti. Indichiamo con λ_i le probabilità di assorbimento in 6 per la catena modificata. Dunque

$$P_1(\textit{arrivare in 6 senza passare da 5 }) = \lambda_1$$

Sapendo che $\lambda_6 = 1$ e $\lambda_5 = 0$ ed impostando il sistema (2.7), otteniamo

$$\begin{cases} \lambda_1 &= \frac{1}{2}\lambda_2 + \frac{1}{2}\lambda_4 \\ \lambda_2 &= \frac{1}{3}\lambda_1 + \frac{1}{3}\lambda_3 \\ \lambda_3 &= \frac{1}{2} + \frac{1}{2}\lambda_2 \\ \lambda_4 &= \lambda_1 \end{cases}$$

che ha soluzione $\lambda_1 = \lambda_2 = \lambda_4 = \frac{1}{3}$, $\lambda_3 = \frac{2}{3}$.

Un secondo quesito di interesse è sapere in media quanto tempo una CM impiega per essere assorbita in una classe chiusa.

In generale, fissata una classe chiusa C (non necessariamente irriducibile), sappiamo che possiamo calcolare le probabilità di assorbimento in C tramite il sistema lineare del teorema precedente. Siamo interessati a calcolare $E_i(\tau_C)$, dove $\tau_C = \inf\{n > 0 : X_n \in C\}$. Si deve però fare attenzione, infatti la v.a. τ_C può prendere valori in $\{1, \ldots, +\infty\}$ e per ogni i

$$P_i(\tau_C = +\infty) = P_i(X_n \in C', \text{ per qualche classe chiusa } C' : C \cap C' = \emptyset)$$

e quindi può risultare infinita con probabilità positiva, ovvero non integrabile.

L'unica maniera per assicurare l'integrabilità della v.a. τ_C è considerando la classe chiusa

$$C = C_1 \cup \cdots \cup C_K,$$

in questo caso per ogni $i \in T$, $P_i(\tau_C < +\infty) = \lambda_i = 1$ ed ha quindi senso calcolarne la media.

Vale allora il seguente teorema

Teorema 2.5.2 : *Sia $\{X_n\}$ una CM con spazio degli stati E finito e sia C la classe chiusa costituita da tutti gli stati persistenti. Poniamo $T_i = E_i(\tau_C)$, dove $\tau_C = \inf\{n > 0 : X_n \in C\}$, allora $T_i = 1$ per ogni $i \in C$, mentre per gli stati transienti il seguente sistema è verificato*

$$T_i = 1 + \sum_{r \in T} p_{ir} T_r, \qquad i \in T \tag{2.10}$$

Dimostrazione: Utilizzando la stessa notazione del teorema 2.5.1, abbiamo per $i \in T$

$$T_i = E_i(\tau_C) = \sum_{n=1}^{+\infty} n P_i(\tau_C = n) = \sum_{n=1}^{+\infty} n g_i^{(n)}$$

ed inoltre sappiamo che per $n \geq 2$

$$g_i^{(n)} = \sum_{r \in T} p_{ir} g_r^{(n-1)}.$$

Dunque

$$\begin{aligned}
T_i &= \sum_{n=1}^{+\infty} n g_i^{(n)} = g_i^{(1)} + \sum_{n=2}^{+\infty} n g_i^{(n)} = g_i^{(1)} + \sum_{n=2}^{+\infty} n \sum_{r \in T} p_{ir} g_r^{(n-1)} \\
&= g_i^{(1)} + \sum_{r \in T} p_{ir} \sum_{n=2}^{+\infty} (n-1+1) g_r^{(n-1)} \\
&= g_i^{(1)} + \sum_{r \in T} p_{ir} \sum_{n=2}^{+\infty} (n-1) g_r^{(n-1)} + \sum_{r \in T} p_{ir} \sum_{n=2}^{+\infty} g_r^{(n-1)} \\
&= g_i^{(1)} + \sum_{r \in T} p_{ir} T_r + \sum_{r \in T} p_{ir} \lambda_r = g_i^{(1)} + \sum_{r \in T} p_{ir} T_r + \lambda_i - g_i^{(1)} = 1 + \sum_{r \in T} p_{ir} T_r,
\end{aligned}$$

dove abbiamo potuto scambiare e dividere le somme poichè lo spazio degli stati è finito ed abbiamo usato il fatto che $\lambda_i = 1$ per ogni $i \in T$, grazie alla definizione di C. □

Esempio 2.5.1 : *Consideriamo la CM rappresentata dal seguente grafo*

È facilmente verificabile che $\{3,4,5\}$ *è l'insieme degli stati transienti e vi sono due classi chiuse ed irriducibili distinte* $C_1 = \{1,2\}$, $C_2 = \{6,7\}$. *Calcoliamo le probabilità di assorbimento in* C_1

$$\begin{cases} \lambda_3 & = \frac{1}{3} + \frac{1}{3}\lambda_4 + \frac{1}{3}\lambda_5 \\ \lambda_4 & = \frac{1}{2} + \frac{1}{2}\lambda_5 \\ \lambda_5 & = \frac{1}{4}\lambda_3 + \frac{1}{4}\lambda_4, \end{cases}$$

la cui soluzione è $(\lambda_3, \lambda_4, \lambda_5) = (\dfrac{2}{3}, \dfrac{2}{3}, \dfrac{1}{3})$. *Se andiamo a calcolare le probabilità di assorbimento in* C_2 *tramite il sistema*

$$\begin{cases} \eta_3 & = \frac{1}{3}\eta_4 + \frac{1}{3}\eta_5 \\ \eta_4 & = \frac{1}{2}\eta_5 \\ \eta_5 & = \frac{1}{2} + \frac{1}{4}\eta_3 + \frac{1}{4}\eta_4, \end{cases}$$

queste risultano ovviamente complementari delle precedenti $(\eta_3, \eta_4, \eta_5) = (\dfrac{1}{3}, \dfrac{1}{3}, \dfrac{2}{3})$. *Infine impostiamo il sistema per i tempi medi di assorbimento in* $C = C_1 \cup C_2$

$$\begin{cases} T_3 & = 1 + \frac{1}{3}T_4 + \frac{1}{3}T_5 \\ T_4 & = 1 + \frac{1}{2}T_5 \\ T_5 & = 1 + \frac{1}{4}T_3 + \frac{1}{4}T_4, \end{cases}$$

troviamo la soluzione $(T_3, T_4, T_5) = (\dfrac{43}{18}, \dfrac{37}{18}, \dfrac{19}{9})$.

Se calcoliamo i tempi medi di assorbimento in $C = \{0, m\}$ nel caso della rovina del giocatore con $p = \frac{1}{2}$, si ottiene che $T_i = i(m-i)$, da cui $T_a = a(m-a) = a(a+b-a) = ab$ ed il tempo medio di assorbimento è massimo per $i = \frac{a+b}{2}$

2.6 Esercizi

1. Dare un esempio di CM a stati finiti per la quale il tempo medio di assorbimento in una classe chiusa ed irriducibile (a partire da uno stato transiente) può essere infinito.

2. Calcolare i tempi medi di assorbimento in $C = \{0, m\}$ nel caso della rovina del giocatore con $p = \frac{1}{2}$.

3. Il diagramma di flusso di un programma di calcolatore è rappresentato dalla seguente figura.

   Ad ogni istante di tempo il programma può passare da uno stato i ad un qualsiasi altro stato cui i sia collegato, secondo le probabilità indicate in figura.

   (a) L'evoluzione del programma può essere descritta da una catena di Markov, scriverne la matrice di transizione e classificarne gli stati.

   (b) Qual è il tempo medio di esecuzione del programma, partendo da 1? È un tempo maggiore o minore che se si partisse da 3?

4. Sia $\{X_n\}_{n\geq 1}$ una successione di v.a. che esprime il numero di clienti in fila ad uno sportello al tempo n. A causa di problemi di spazio, non vi possono mai essere più di 4 clienti in fila. Inoltre se l'operatore riesce ad esaurire la fila, chiude lo sportello, poichè deve recarsi ad altro servizio. All'inizio vi sono 2 clienti in fila. Assumendo che le X_n formano una catena di Markov con la

seguente matrice di transizione

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

(a) Disegnare il grafo corrispondente e classificare gli stati della catena.

(b) Calcolare il tempo medio per chiudere lo sportello, partendo da una fila con due clienti.

5. Una postazione di servizio può processare al massimo 4 lavori per minuto, se ne deve processare di più si ferma. Sia X_n il numero di lavori in fila per essere serviti al minuto n e supponiamo che sia una catena di Markov con la seguente matrice delle probabilità di transizione

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(a) Classificare gli stati e disegnare il grafo associato alla catena.

(b) Calcolare il tempo medio affinchè la postazione si fermi sapendo che si è partiti da un solo lavoro in fila.

6. Le 5 fasi di una lavorazione industriale rappresentano gli stati di una catena di Markov che viene identificatata dalla seguente matrice di transizione, assumendo che 4 e 5 sono le fasi finali.

$$\begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(a) Disegnare il grafo corrispondente e classificare gli stati della catena.

(b) Assumendo che inizialmente si parta dalla fase 1, calcolare $P(X_2 = 4)$.

(c) Calcolare i tempi medi per completare un manufatto.

7. Un topo viene posto in un labirinto rappresentato dal seguente schema.

Ad ogni vertice il topo sceglie con uguale probabilita' una delle direzioni possibili e puo' compiere solo un passo alla volta (cioè ad ogni tempo può andare solo nei vertici adiacenti), tranne che nel vertice 4, dove è situata l'uscita dal labirinto. Sia X_n la posizione del topo al tempo n. X_n rappresenta una catena di Markov.

   (a) Disegnare il grafo corrispondente a questa catena e scrivere la matrice di transizione.
   (b) Classificare gli stati.
   (c) Calcolare la probabilità per il topo di uscire dal labirinto partendo dal vertice 1.
   (d) Calcolare i tempo medio, a partire da ogni posizione, che il topo impiega per uscire dal labirinto.

8. Calcolare i tempi medi di assorbimento per l'esempio 2 di pagina 11 degli appunti.

9. Si consideri la catena di Markov X_n con la seguente matrice di transizione

$$\begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

   (a) Determinare tutte le probabilità invarianti.
   (b) Trovare i tempi medi di ritorno in ciascuno stato.

10. Data la CM con il seguente grafo:

(a) Determinare il periodo di ciascuno stato.

(b) Considerando la catena definita da $Y_n = X_{2n}$, si determinino le classi chiuse ed irriducibili e gli stati transienti di Y_n.

(c) Per la catena Y_n, calcolare in quale classe chiusa irriducibile è più probabile essere assorbiti, partendo da 4.

2.7 Probabilità invarianti

Una domanda naturale che sorge quando si hanno dei sistemi dinamici rappresentati da una catena di Markov è cosa succede se si lascia andare il tempo ad infinito, ovvero se il sistema raggiunge una qualche situazione di equilibrio

Definizione 2.7.1 : *Sia $\{X_n\}_n$ una CM con spazio degli stati E e matrice di transizione $\mathcal{P} = (p_{ij})_{i,j \in E}$, sia inoltre $\pi = \{\pi_j\}_{j \in E}$ una misura su E. La misura π è detta* **invariante** *o* **stazionaria** *per la CM $\{X_n\}_n$ se*

$$\pi_j = \sum_{i \in E} p_{ij} \pi_i, \quad 0 \leq \pi_j \leq 1, \quad \sum_{j \in E} \pi_j = 1$$

ovvero in notazione operatoriale (matriciale)

$$\pi^T = \pi^T \mathcal{P}, \quad 0 \leq \pi_j \leq 1, \quad \sum_{j \in E} \pi_j = 1$$

Occorre chiedersi dunque quando esiste una misura invariante, se è unica e in che senso fornisce informazioni per il comportamento asintotico della catena di Markov ovvero se ci permette di calcolare

$$\lim_{n \to +\infty} P(X_n = j) = \pi_j^n, \qquad \lim_{n \to +\infty} P_i(X_n = j) = \lim_{n \to +\infty} p_{ij}^{(n)}$$

Osservazioni 2.7.1 :

1. *Se π è una probabilità stazionaria per la CM con matrice di transizione $\mathcal{P}$, allora*

$$\pi^T = \pi^T \mathcal{P} = \pi^T \mathcal{P}\mathcal{P} = \pi^T \mathcal{P}^2 = \cdots = \pi^T \mathcal{P}^n, \qquad \forall\, n,$$

dunque se si partisse al tempo iniziale da una distribuzione invariante ($\pi^0 = \pi$) allora necessariamente la catena di Markov rimarrebbe in questa configurazione a tutti i tempi

$$\pi_i^n = P(X_n = i) = ((\pi^0)^T \mathcal{P})_i = \pi_i, \qquad \forall\, n$$

In altre parole si può dire che una distribuzione stazionaria rappresenta una configurazione di equilibrio per il sistema dinamico descritto dalla CM.

2. *Se j è uno stato transiente, abbiamo dimostrato che $\lim_{n \to +\infty} p_{ij}^{(n)} = 0$, per ogni $i \in E$. Conseguentemente, se esiste una probabilità invariante, si deve avere*

$$\pi_j = \sum_{i \in E} \pi_i p_{ij}^{(n)} = \lim_{n \to +\infty} \sum_{i \in E} \pi_i p_{ij}^{(n)} = 0$$

ovvero deve essere nulla sugli stati transienti. Notiamo che siamo potuti passare al limite sotto il segno di serie, grazie alla positività e sommabilità degli addendi.

Di seguito forniamo un primo teorema di esistenza per le misure invarianti. Chiaramente tutti i risultati che sono stati e che verranno esposti per catene irriducibili possono essere riformulati per classi chiuse irriducibili.

Teorema 2.7.1 : *Sia $\{X_n\}_n$ una catena di Markov irriducibile, persistente positiva con spazio degli stati E. Allora esiste un'unica probabilità invariante π data da*

$$\pi_j = \frac{1}{E_j(\tau_j)}, \qquad \forall\, j \in E.$$

Dimostrazione: Scelto un arbitrario stato j, definiamo le successioni dei tempi di $k-$mo ritorno in j

$$\tau_j^{(1)} = \tau_j, \ldots, \tau_j^{(k)} = \min\{n > \tau_j^{(k-1)} \,:\, X_n = j\}$$

e dei tempi di interritorno in j

$$r_j^1 = \tau_j^{(1)}, \ldots, r_j^k = \tau_j^{(k)} - \tau_j^{(k-1)}.$$

Vogliamo dimostrare che $\{r_j^k\}_{k\geq 1}$ è una successione di v.a. indipendenti sotto qualsiasi P_i.

Perciò presa una qualsiasi $k-$pla $n_1, \ldots, n_k \in \mathbb{N}$,

$$\begin{aligned}
& P_i(r_j^1 = n_1, \ldots, r_j^k = n_k) = P_i(\tau_j^{(1)} = n_1, \ldots, \tau_j^{(k)} - \tau_j^{(k-1)} = n_k) \\
= \;& P_i(\tau_j^{(1)} = n_1, \tau_j^{(2)} = n_1 + n_2, \ldots, \tau_j^{(k)} = n_1 + n_2 + \cdots + n_k) \\
= \;& P_i(X_1 \neq j, .., X_{n_1-1} \neq j, X_{n_1} = j, .., X_{n_1+n_2-1} \neq j, X_{n_1+n_2} = j, .., X_{n_1+..+n_k} = j) \\
= \;& P(X_{n_1+\cdots+n_k} = j, X_{n_1+\cdots+n_k-1} \neq j, \ldots, X_{n_1+\cdots+n_{k-1}+1} \neq j | X_{n_1+\cdots+n_{k-1}} = j) \\
\cdot \;& P(X_{n_1+\cdots+n_{k-1}} = j, X_{n_1+\cdots+n_{k-1}-1} \neq j, \ldots, X_{n_1+\cdots+n_{k-2}+1} \neq j | X_{n_1+\cdots+n_{k-2}} = j) \\
& \vdots \\
\cdot \;& P(X_{n_1} = j, X_{n_1-1} \neq j, \ldots X_1 \neq j | X_0 = i) \\
= \;& \rho_{ij}^{(n_1)} \rho_{jj}^{(n_2)} \ldots \rho_{jj}^{(n_k)},
\end{aligned}$$

dove l'ultimo passaggio è stato realizzato grazie alla proprietà di Markov. Nell'ultimo prodotto solo il primo fattore dipende da i, quindi se saturiamo rispetto ad r_j^1 sotto qualsiasi P_i

$$P_i(r_j^2 = n_1, \ldots, r_j^k = n_k) = P_i(r_j^2 = n_1) \ldots P_i(r_j^k = n_k) = \rho_{jj}^{(n_2)} \ldots \rho_{jj}^{(n_k)},$$

che dimostra l'indipendenza e l'identica distribuzione delle v.a. $\{r_j^k\}_{k\geq 2}$

Prendendo $i = j$ si ha anche l'identica distribuzione a partire da $k = 1$

$$P_j(r_j^k = n) = P_j(r_j^1 = n).$$

Calcolando la media di r_j^1 otteniamo

$$E_j(r_j^1) = \sum_{n=1}^{+\infty} nP_j(r_j^1 = n) = \sum_{n=1}^{+\infty} nP_j(\tau_j = n) = E_j(\tau_j) = m_j$$

e possiamo applicare la legge dei grandi numeri

$$\lim_{m\to+\infty} \frac{\sum_{k=1}^m r_j^k}{m} = E_j(\tau_j) = m_j.$$

D'altra parte, se definiamo la v.a. $N_n(j) =$ il numero di visite nello stato j nei primi n passi, allora per costruzione si ha

$$\tau_j^{N_n(j)} \le n < \tau_j^{N_n(j)+1}$$

conseguentemente

$$\frac{\tau_j^{N_n(j)}}{N_n(j)} \le \frac{n}{N_n(j)} < \frac{\tau_j^{N_n(j)+1}}{N_n(j)+1}\frac{N_n(j)+1}{N_n(j)} \tag{2.11}$$

e grazie alla persistenza dello stato j, si ha necessariamente che $N_n(j) \longrightarrow +\infty$ per $n \to +\infty$ con probabilità $P_i = 1$ per ogni i. Usando un'estensione della legge dei grandi numeri (vedere [B]) possiamo affermare

$$\frac{\tau_j^{N_n(j)}}{N_n(j)} = \frac{r_j^1 + \cdots + r_j^{N_n(j)}}{N_n(j)} \longrightarrow m_j$$

e sfruttando le disuguaglianze (2.11) possiamo concludere che per $n \to +\infty$

$$\frac{n}{N_n(j)} \longrightarrow m_j$$

sotto P_j. In realtà lo stesso risultato vale sotto qualsiasi P_i, infatti la successione $\{r_j^k\}_{k\ge2}$ è i.i.d. sotto qualsiasi P_i, dunque per la stessa legge dei grandi numeri

$$\frac{r_j^2 + \cdots + r_j^{N_n(j)}}{N_n(j)-1} \longrightarrow E_i(r_j^2) = \alpha$$

dove questa media è indipendente da i per quanto detto prima. Poichè per $i = j$, r_j^k ha la stessa distribuzione di r_j^1, non si può che avere $\alpha = m_j$. D'altra parte sotto qualsiasi P_i

$$\frac{r_j^1 + \cdots + r_j^{N_n(j)}}{N_n(j)} = \frac{r_j^1}{N_n(j)} + \frac{r_j^2 + \cdots + r_j^{N_n(j)}}{N_n(j)-1}\frac{N_n(j)-1}{N_n(j)} \longrightarrow \alpha = m_j$$

Quindi sotto $4P_i$ per ogni i il reciproco $\dfrac{N_n(j)}{n}$ tende a $\dfrac{1}{m_j}$ ed inoltre verifica $\dfrac{N_n(j)}{n} \leq 1$ per costruzione, perciò applicando il teorema della convergenza dominata si ottiene

$$E_i\left(\frac{N_n(j)}{n}\right) \longrightarrow \frac{1}{m_j}, \qquad \text{per} \quad n \to +\infty. \tag{2.12}$$

Se la catena è persistente positiva il limite (2.12 è ben definito ed è diverso da zero (viceversa se fosse persistente nulla o transiente si può concludere che il limite è zero).

Notiamo che possiamo riesprimere $N_n(j) = \sum_{k=1}^{n} \mathbf{1}_{\{X_k=j\}}$ e dunque

$$E_i\left(\frac{N_n(j)}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} P_i(X_k = j) = \frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)}$$

Rimane da dimostrare che $\{\pi_j\}_j = \{\frac{1}{m_j}\}_j$ è una probabilità invariante per la catena di Markov.

1. $\pi_j = \dfrac{1}{m_j} \geq 0$;

2. $\sum_{j\in E} \pi_j = \sum_{j\in E} \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)} = \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n}\sum_{j\in E} p_{ij}^{(k)} = 1$, dove nell'ultima uguaglianza abbiamo sfruttato il teorema della convergenza dominata per scambiare le somme ed il limite;

3. Invarianza: scelto un qualsiasi $h \in E$ si ha
$$\begin{aligned}\sum_{i\in E} \pi_i p_{ij} &= \sum_{i\in E} p_{ij} \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} p_{hi}^{(k)} = \lim_{n\to+\infty} \frac{1}{n}\sum_{i\in E} p_{ij}\sum_{k=1}^{n} p_{hi}^{(k)} \\ &= \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n}\sum_{i\in E} p_{hi}^{(k)} p_{ij} = \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} p_{hj}^{(k+1)} = \pi_j;\end{aligned}$$

4. Unicità: sia $\bar{\pi}$ una seconda probabilità invariante, allora si ha
$$\bar{\pi}_j = \sum_{i\in E} \bar{\pi}_i p_{ij} \quad \Rightarrow \quad \bar{\pi}_j = \sum_{i\in E} \bar{\pi}_i p_{ij}^{(k)}$$
da cui prendendone una media aritmetica si ottiene
$$\bar{\pi}_j = \sum_{i\in E} \frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)} \bar{\pi}_i \longrightarrow \pi_j \sum_{i\in E} \pi_i = \pi_j \qquad \text{per} \quad n \to +\infty.$$

$\square$

Osserviamo che la precedente dimostrazione può essere ripetuta (fino a pagina 38) anche nel caso di una CM irriducibile persistente nulla. In questo caso peró i limiti saranno divergenti ad infinito, che implica che per ogni j $\pi_j = 0$ e dunque non può rappresentare una probabilità in quanto non si somma ad 1. D'altra parte questo dimostra che, qualora esista una distribuzione invariante (per qualsiasi catena), essa deve essere nulla sugli stati persistenti nulli, così come lo avevamo concluso per gli stati transienti.

Qui di seguito presentiamo una versione più semplice del teorema di esistenza ed unicità, nel caso in cui $|E| < +\infty$.

Teorema 2.7.2 : *Teorema di Markov-Kakutani*

Sia $\{X_n\}_n$ *una catena di Markov con spazio degli stati E finito. Allora esiste una probabilità invariante* π*, che è anche unica se la catena è irriducibile.*

Dimostrazione: Sia $|E| = m$ e $\mathcal{P}$ la matrice $m \times m$ di transizione associata alla catena. Consideriamo l'insieme delle misure di probabilità su E, ovvero

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^m : \sum_{i=1}^{m} x_i = 1\,,\ 0 \leq x_i \leq 1\}$$

L'insieme $\mathcal{C}$ è un sottoinsieme chiuso e limitato si $\mathbb{R}^m$ ed è quindi compatto.

Sia $\mathbf{v} \in \mathcal{C}$ e definiamo la successione di vettori

$$\mathbf{v}^n = \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{v} \cdot \mathcal{P}^k$$

Allora $\mathbf{v}^n \in \mathcal{C}$, infatti

1. $v_i^n \geq 0$ poiché $v_v^n = \dfrac{1}{n}\displaystyle\sum_{k=0}^{n-1}\sum_{j=1}^{m} v_j p_{ji}^{(k)} \geq 0$;

2. è ancora una probabilità poiché

$$\begin{aligned}
\sum_{i=1}^{m} v_i^n &= \sum_{i=1}^{m}\frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=1}^{m} v_j p_{ji}^{(k)} = \frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=1}^{m}\sum_{i=1}^{m} v_j p_{ji}^{(k)} \\
&= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=1}^{m} v_j \sum_{i=1}^{m} p_{ji}^{(k)} = \frac{1}{n}\sum_{k=0}^{n-1}\sum_{j=1}^{m} v_j \\
&= \frac{1}{n}\sum_{k=0}^{n-1} 1 = 1
\end{aligned}$$

3. l'insieme $\mathcal{C}$ è un compatto quindi si può estrarre una sottosuccessione convergente da $\{\mathbf{v}^n\}_n$, che denotiamo con $\{\mathbf{v}^{n_k}\}_k$ e con π il suo limite.

Per compatezza, necessariamente $\pi \in \mathcal{C}$ ed è dunque una misura di probabilità, dobbiamo dimostrare che è invariante per $\mathcal{P}$

$$\begin{aligned} \mathbf{v}^{n_k} - \mathbf{v}^{n_k} \cdot \mathcal{P} &= \frac{1}{n_k} \sum_{h=0}^{n_k-1} \mathbf{v} \cdot \mathcal{P}^h - \frac{1}{n_k} \sum_{h=0}^{n_k-1} \mathbf{v} \cdot \mathcal{P}^h \cdot \mathcal{P} \\ &= \frac{1}{n_k} \sum_{h=0}^{n_k-1} \left[\mathbf{v} \cdot \mathcal{P}^h - \mathbf{v} \cdot \mathcal{P}^{(h+1)}\right] = \frac{1}{n_k}(\mathbf{v} - \mathbf{v} \cdot \mathcal{P}^{n_k}). \end{aligned}$$

Ogni componente del precedente prodotto righe per colonne è limitata da 1, perciò spingendo k ad infinito si ottiene per ogni i

$$(\frac{1}{n_k}(\mathbf{v} - \mathbf{v} \cdot \mathcal{P}^{n_k}))_i \longrightarrow 0$$

d'altro canto

$$\mathbf{v}^{n_k} - \mathbf{v}^{n_k} \cdot \mathcal{P} \longrightarrow \pi - \pi \cdot \mathcal{P}$$

concludendo la dimostrazione. □

Riassumendo le conclusioni dei precedenti teoremi sappiamo che, se esiste una misura stazionaria, su ogni stato transiente o persistente nullo questa deve essere nulla, conseguentemente per una catena irriducibile transiente o persistente nulla non può esistere una misura invariante poiché

$$\sum_{j \in E} \pi_j = 0$$

e quindi non soddisfa gli assiomi della probabilità.

Viceversa se abbiamo una catena di Markov irriducibile persistente positiva esiste un'unica probabilità invariante data da

$$\pi_j = \frac{1}{E_j(\tau_j)}, \qquad j \in E$$

e quindi questo vale anche per ogni catena irriducibile con spazio degli stati finito.

Conseguentemente, una volta decomposto un qualsiasi spazio degli stati nell'insieme dei transienti T e nelle sue classi chiuse ed irriducibili $C_1, \ldots, C_K$ il teorema di esistenza implica tutte le misure invarianti devono essere nulle su T e su tutti gli stati persistenti nulli e che su ogni classe chiusa ed irriducibile persistente positiva è concentrata un'unica probabilità invariante $\pi(C_1), \ldots, \pi(C_K)$.

Allora tutte le misure invarianti relative allo spazio E, dovendo verificare $\sum_{j\in E}\pi_j = 1$, non potranno essere altro che combinazioni convesse delle $\pi(C_1), \ldots, \pi(C_K)$.

Forniamo ora alcuni esempi per illustrare le precedenti conclusioni

Esempi 2.7.1 :

1. *Si consideri la CM con matrice di transizione*

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

   Classificando gli stati è immediato constatare che $\{5\}$ *è l'unico stato assorbente, mentre tutti gli altri sono transienti.*

   Dunque la probabilità invariante deve essere concentrata sull'unico stato persistente positivo, ovvero

$$\pi = (0, 0, 0, 0, 1).$$

2. *Sia data la CM con grafo*

   la cui matrice di transizione associata è

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

essendo una catena finita deve esistere almeno una distribuzione stazionaria. Impostiamo il sistema $\pi^T = \pi^T \mathcal{P}$

$$(2.13)\qquad \begin{cases} \pi_1 = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{5}\pi_3 \\ \pi_2 = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{5}\pi_3 \\ \pi_3 = \frac{1}{5}\pi_3 \\ \pi_4 = \frac{1}{5}\pi_3 + \frac{1}{3}\pi_4 + \frac{1}{3}\pi_5 \\ \pi_5 = \frac{1}{5}\pi_3 + \frac{2}{3}\pi_4 + \frac{2}{3}\pi_5 \\ \sum_{i=1}^{5} \pi_i = 1 \\ 0 \leq \pi_i \leq 1 \qquad \forall\, i \end{cases}$$

la cui soluzione è data da $\pi_1 = \pi_2$, $\pi_3 = 0$ $\pi_4 = \frac{1}{2}\pi_5$, poiché in realtà il sistema è costituito da due sottosistemi separati. Dagli ultimi due vincoli si ottiene infine

$$\pi_1 = -\frac{3}{4}\pi_5 + \frac{1}{2}, \qquad 0 \leq \pi_5 \leq \frac{2}{3}.$$

L'insieme delle soluzioni può essere quindi scritto come

$$\Big(-\frac{3}{4}\lambda + \frac{1}{2}, -\frac{3}{4}\lambda + \frac{1}{2}, 0, \frac{1}{2}\lambda, \lambda \Big), \qquad 0 \leq \lambda \leq \frac{2}{3}$$

e per $\lambda = 0$ otteniamo il vettore $(\frac{1}{2}, \frac{1}{2}, 0, 0, 0)$, mentre per $\lambda = \frac{2}{3}$ il vettore $(0, 0, 0, \frac{1}{3}, \frac{2}{3})$. Effettuando la sostituzione di parametro $\alpha = \frac{3}{2}\lambda$ ci si rende conto che un'espressione alternativa dell'insieme delle soluzioni è

$$\Big(\tfrac{1}{2}(1-\alpha), \tfrac{1}{2}(1-\alpha), 0, \tfrac{1}{3}\alpha, \tfrac{2}{3}\alpha\Big), \qquad 0 \leq \alpha \leq 1, \ \ \textit{ovvero}$$

$$(2.14)\qquad (1-\alpha)\Big(\tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0\Big) + \alpha\Big(0, 0, 0, \tfrac{1}{3}, \tfrac{2}{3}\Big), \qquad 0 \leq \alpha \leq 1.$$

Alternativamente, anziché impostare direttamente il sistema (2.13), possiamo classificare gli stati riconoscendo le due classi chiuse ed irriducibili $\{1, 2\}$, $\{4, 5\}$ e lo stato transiente $\{3\}$. Il teorema di esistenza ed unicità implica che $\pi_3 = 0$ e che su ognuna delle classi esiste un'unica misura invariante, date dai due sistemi ridotti

$$\begin{cases} \pi_1 = \pi_2 \\ \pi_1 + \pi_2 = 1 \end{cases}, \qquad \begin{cases} \pi_4 = \frac{1}{2}\pi_5 \\ \pi_4 + \pi_5 = 1 \end{cases}$$

con soluzioni $(\pi_1, \pi_2) = (\frac{1}{2}, \frac{1}{2})$ e $(\pi_4, \pi_5) = (\frac{1}{3}, \frac{2}{3})$. Immergendo quest'ultime in $\mathbb{R}^5$ e facendone la combinazione convessa otteniamo direttamente la rappresentazione (2.14).

Vi è un caso particolare quando determinare una probabilità stazionaria risulta particolarmente semplice.

Definizione 2.7.2 : *Una matrice di transizione $\mathcal{P}$ si dice* **bistocastica** *se*

$$\sum_{i \in E} p_{ij} = 1$$

ovvero anche le colonne si sommano ad 1.

Un caso particolare di matrice bistocastica è chiaramente quando $\mathcal{P}$ è simmetrica.

Nel caso di una matrice bistocastica, se lo spazio degli stati E è finito, allora la probabilità invariante è data da

$$\pi_j = \frac{1}{|E|}, \qquad \forall\, j \in E$$

infatti

$$\sum_{i \in E} \pi_i p_{ij} = \frac{1}{|E|} \sum_{i \in E} p_{ij} = \frac{1}{|E|} = \pi_j.$$

2.8 Esercizi

1. Dare la definizione di probabilità invariante per una catena di Markov.

2. Enunciare e dimostrare il teorema di Markov-Kakutani.

3. Sia $\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$ una matrice di transizione per una catena di Markov X_n. Calcolare approssimativamente $E(X_n)$, per n grande.

4. Si consideri la catena di Markov X_n con la seguente matrice delle probabilità di transizione
$$\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}$$
   (a) Disegnare il grafo associato e classificare gli stati.
   (b) Calcolare tutte le probabilità invarianti.

5. Si consideri la catena di Markov X_n a tre stati tale che, in un passo, da 1 si possa andare solo in 2, da 2 si possa andare in 2 od in 3 con uguale probabilità e da 3 si possa andare solo in 1. Inizialmente ogni stato è ugualmente probabile.

   (a) Calcolare $P(X_0 = 1, X_2 = 3)$ e $P(X_1 = 1, X_3 = 3)$.

   (b) Dire se esistono, ed eventualmente calcolare, tutte le distribuzioni invarianti della catena.

6. In un macchinario vi sono 5 relais, anche se esso puo' funzionare anche con 3. Ad ogni tempo, ciascun relais ha probabilita' 1/2 di guastarsi, ma grazie al modo operativo della macchina non se ne possono guastare piu' di 2 contemporaneamente. Inoltre, se il numero di relais e' minore od uguale a due, un meccanismo aggiunge automaticamente due relais con probabilita' 1/2.

   (a) SeX_n indica il numero di relais presenti al tempo n, giustificare che e' una Catena di Markov e scriverne la matrice di transizione.

   (b) Classificare gli stati ed individuare le classi chiuse ed irriducibili.

   (c) Calcolare le eventuali distribuzioni invarianti.

7. Sia $\{X_n\}_{n\in\mathbb{N}}$ una CM a stati in $S = \{0, 1, 2, 3, 4\}$ con le seguenti probabilità di transizione diverse da zero:

$$\begin{array}{rl} p_{01} = \frac{1}{2}, & p_{02} = \frac{1}{2} \\ p_{13} = \frac{1}{4}, & p_{14} = \frac{3}{4} \\ p_{23} = \frac{1}{4}, & p_{24} = \frac{3}{4} \\ p_{30} = 1 & p_{40} = 1 \end{array}$$

   (a) Calcolare $p_{00}^{(n)}$ per $n \geq 1$ ed il periodo di ciascuno stato.

   (b) Determinare la distribuzione, a partire da 0, del tempo di primo arrivo in 0, τ_0.

   (c) Trovare le eventuali distribuzioni stazionarie.

8. Un' urna contiene 2 palline rosse ed un numero X_0 compreso tra 0 e 3 di palline nere. Si effettuano delle estrazioni dall'urna. Ad ogni estrazione se esce una pallina nera, la si mette da parte, invece se esce una rossa, la si reinserisce nell'urna insieme ad una pallina nera (si tenga presente che al massimo si hanno tre palline nere). Sia X_n il numero di palline nere presenti nell'urna dopo l'$n-$ma estrazione.

   (a) Riconoscere che $\{X_n\}$ è una CM e trovarne le probabilità di transizione.

   (b) Studiare il carattere degli stati e trovare le eventuali distribuzioni invarianti.

(c) Sia $X_0 = 2$. Un giocatore scommette che la catena raggiunga lo stato 3 prima dello stato 0. Con quale probabilità vince?

9. La successione di v.a. X_n a valori in $\{0, 1, 2\}$ è definita nel modo seguente:

$$X_0 = 1$$

Per ogni $n \geq 1$ si lancia una moneta equa

$$X_{n+1} = \begin{cases} (X_n + 1)\mathrm{mod}3 & \text{se esce testa} \\ 1 & \text{altrimenti} \end{cases}$$

(a) Dimostrare che si tratta di una CM.

(b) Calcolare la probabilità dell'evento $\{X_1 = 1, X_3 = 1\}$.

(c) Calcolare le probabilità di transizione e quelle invarianti.

2.9 Comportamento asintotico

Vogliamo ora rispondere al terzo quesito formulato in precedenza sul comportamento asintotico di una catena di Markov. In particolare vogliamo capire sotto quali condizioni risulta vero

$$\lim_{n\to+\infty} p_{ij}^{(n)} = \pi_j, \qquad \lim_{n\to+\infty} \pi_j^n = \pi_j.$$

Osserviamo immediatamente che il primo limite implica il secondo, infatti

$$\lim_{n\to+\infty} \pi_j^n = \lim_{n\to+\infty} \sum_{h\in E} \pi_h^0 p_{hj}^{(n)} = \sum_{h\in E} \pi_h^0 \lim_{n\to+\infty} p_{hj}^{(n)} = \sum_{h\in E} \pi_h^0 \pi_j = \pi_j,$$

dove le operazioni di somma e limite si sono potute scambiare grazie alla positività di tutti gli addendi.

Per individuare le condizioni necessarie a garantire i precedenti limiti, dobbiamo introdurre qualche altra nozione.

Definizione 2.9.1 : *Sia E lo spazio degli stati, allora per ogni $i \in E$ definiamo* **periodo** *di i il MCD$\{n : p_{ii}^{(n)} > 0\}$*

Osservazione 2.9.1 : *Se per uno stato i risulta* $p_{ii} > 0$ *allora chiaramente il periodo di i è 1; non è vero però il viceversa infatti si abbia la seguente matrice di transizione* 3×3

$$\mathcal{P} = \begin{pmatrix} 0 & p_{12} & 0 \\ p_{21} & 0 & p_{23} \\ p_{31} & p_{32} & 0 \end{pmatrix}$$

allora disegnando il grafo è facile rendersi conto che $p_{11}^{(2)} > 0$ *e* $p_{11}^{(3)} > 0$, *ma 2 e 3 sono coprimi e dunque il periodo dello stato 1 è 1 anche in questo caso.*

È naturale chiedersi se nel caso delle catene irriducibili possiamo dire qualcosa in più.

Proposizione 2.9.1 : *In una CM irriducibile tutti gli stati hanno lo stesso periodo.*

Dimostrazione: Siano i e j due stati arbitrari e sia t il periodo di i. Vogliamo dimostrare che se $p_{jj}^{(n)} > 0$ per qualche n, allora $t|n$ (t divide n).

La catena è irriducibile, quindi esistono due indici n_0 ed n_1 tali che $p_{ij}^{(n_0)} > 0$ e $p_{ji}^{(n_1)} > 0$. Conseguentemente

$$p_{ii}^{(n_0+n_1)} = \sum_{k \in E} p_{ik}^{(n_0)} p_{kj}^{(n_1)} \leq p_{ij}^{(n_0)} p_{ji}^{(n_1)} > 0$$

perciò, per definizione di MCD, $t|n_0 + n_1$.

D'altra parte se n_2 è un indice tale che $p_{jj}^{(n_2)} > 0$, allora si ha anche

$$p_{ii}^{(n_0+n_1+n_2)} \geq p_{ij}^{(n_0)} p_{jj}^{(n_2)} p_{ji}^{(n_1)} > 0,$$

da cui $t|n_0 + n_1 + n_2$. Insieme con la precedente conclusione, quest'ultima divisibilità implica $t|n_2$, ovvero il periodo di i deve dividere il periodo di j.

I ruoli di i e j possono essere scambiati quindi anche il periodo di j deve dividere il periodo di i, in altre parole i due stati hanno lo stesso periodo. □

Quanto detto per una catena di Markov irriducibile può essere riespresso per ogni classe chiusa ed irriducibile.

Definizione 2.9.2 : *Una CM* $\{X_n\}_n$ *si dice* **aperiodica** *se tutti gli stati hanno periodo 1.*

Procediamo a dimostrare il cosiddetto teorema ergodico per le catene di Markov. Per fare questo abbiamo bisogno di presentare prima due lemmi.

Lemma 2.9.1 : *Siano $n_1, \dots n_k$ naturali coprimi, allora per n sufficientemente grande è possibile scrivere*

$$n = \sum_{h=1}^{k} a_h n_h \tag{2.15}$$

con $a_h \in \mathbb{N}$ per $h = 1, \dots, k$.

Dimostrazione: Poichè $n_1, \dots n_k$ sono coprimi, è possibile trovare $b_1, \dots, b_k \in \mathbb{Z}$ tali che

$$1 = \sum_{h=1}^{k} b_h n_h.$$

Poniamo $N = \sum_{h=1}^{k} n_h$ e scelto $n \geq N$ scriviamo

$$n = qN + r, \qquad \text{con} \quad q, r \in \mathbb{N}, \quad 0 \leq r < N.$$

Allora

$$n = qN + r = q \sum_{h=1}^{k} n_h + r \cdot 1 = \sum_{h=1}^{k} q n_h + r \sum_{h=1}^{k} b_h n_h = \sum_{h=1}^{k} (q + r b_h) n_h.$$

Se scegliamo n tale che $\left[\dfrac{n}{N}\right] = q > N \max_h |b_h|$, allora i coefficienti

$$a_h := q + r b_h \geq q - |b_h| N > 0$$

□

Lemma 2.9.2 : *Sia $\{X_n\}_n$ una catena di Markov irriducibile ed aperiodica, allora presi $i, j \in E$ arbitrari esiste un indice $\bar{n}(i,j)$ tale che $p_{ij}^{(n)} > 0$ per ogni $n > \bar{n}(i,j)$.*

Dimostrazione: Essendo la catena aperiodica ed irriducibile, qualsiasi stato j ha periodo 1, allora devono esistere $k \geq 1$ naturali $n_1, \dots, n_k$ coprimi tra loro tali che $p_{jj}^{(n_h)} > 0$ per ogni $h = 1, \dots, k$.

Dal lemma precedente, esiste un naturale n_0 sufficientemente grande tale che scelto qualsiasi $\bar{n} \geq n_0$, questo può essere scritto come

$$\bar{n} = \sum_{h=1}^{k} a_h n_h, \qquad \text{con coefficienti naturali } a_h > 0, \quad \forall\, h = 1, \dots, k.$$

Allora

$$p_{jj}^{(\bar{n})} \geq \prod_{h=1}^{k} p_{jj}^{(a_h n_h)} \geq \prod_{h=1}^{k} p_{jj}^{(n_h)} \dots p_{jj}^{(n_h)} > 0,$$

dove il fattore è stato ripetuto a_h volte, ma la CM è irriducibile e quindi esiste un naturale $\tilde{n}_0$ tale che $p_{ij}^{(\tilde{n}_0)} > 0$.

Dunque preso $n^*(i,j) = n_0 + \tilde{n}_0$, si avrà che per ogni $n \geq n^*$, $p_{ij}^{(n)} \geq p_{ij}^{(\tilde{n}_0)} p_{jj}^{(\bar{n})} > 0$, poichè si può sempre scegliere $\bar{n} = n - \tilde{n}_0 \geq n_0$. □

Passiamo ora a dimostrare il risultato principale di questo paragrafo.

Teorema 2.9.1 : *Teorema ergodico 1*

Sia data una catena di Markov irriducibile, aperiodica, persistente positiva con spazio degli stati E e sia π la sua probabilità invariante. Allora

$$\lim_{n \to +\infty} p_{ij}^{(n)} = \pi_j, \qquad \forall\, i, j, \in E.$$

Dimostrazione: Con un leggero abuso di notazione indichiamo con $\mathcal{P} = (p_{ij})_{i,j \in E}$ la matrice di transizione associata alla CM, che sia essa di dimensione finita od infinita (ovvero un operatore).

Data la famiglia di probabilità condizionate p_{ij}, per il teorema di Kolmogorov, è possibile costruire uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ ed una successione di vettori aleatori bidimensionali $\{(X_n, Y_n)\}_n$ con $\{X_n\}_n$ indipendente da $\{Y_n\}_n$ e con distribuzioni condizionate

$$P(X_n = j | X_{n-1} = i) = p_{ij} = P(Y_n = j | Y_{n-1} = i), \qquad \forall\, n \in \mathbb{N}.$$

Quindi prese singolarmente, ognuna delle due successioni è una CM; vogliamo dimostrare che congiuntamente formano una CM bidimensionale con spazio degli stati $E \times E$.

Calcolando le candidate probabilità di transizione per $\{(X_n, Y_n)\}_n$ abbiamo per indipendenza

$$P((X_n, Y_n) = (k,h) | (X_{n-1}, Y_{n-1}) = (i,j)) = \frac{P(X_n = k, Y_n = h, X_{n-1} = i, Y_{n-1} = j)}{P(X_{n-1} = i, Y_{n-1} = j)}$$
$$= \frac{P(X_n = k, X_{n-1} = i) P(Y_n = h, Y_{n-1} = j)}{P(X_{n-1} = i) P(Y_{n-1} = j)} = p_{ik} p_{jh}.$$

Usando la proprietà di Markov per ognuna delle componenti e l'indipendenza delle successioni è diretto verificare che anche la successione congiunta verifica la proprietà di Markov e la matrice

$$\tilde{\mathcal{P}} = (\tilde{p}_{(i,j)(k,h)})_{(i,j),(k,h) \in E \times E} = (p_{ik} p_{jh})_{(i,j),(k,h) \in E \times E}$$

è effettivamente una matrice stocastica poiché

1. $\tilde{p}_{(i,j)(k,h)} \geq 0$ per ogni $(i,j), (k,h) \in E \times E$;

2. Vale

$$\sum_{(k,h)\in E\times E} \tilde{p}_{(i,j)(k,h)} = \sum_{k\in E}\sum_{h\in E} p_{ik}p_{jh} = \sum_{k\in E} p_{ik} \sum_{h\in E} p_{jh} = 1.$$

Analogamente si può verificare la relazione

$$\tilde{p}^{(n)}_{(i,j)(k,h)} = p^{(n)}_{ik} p^{(n)}_{jh}, \qquad \forall\, (i,j),(k,h) \in E\times E$$

Inoltre la catena accoppiata è aperiodica poiché lo era la catena di partenza, infatti se MCD$\{n : p^{(n)}_{ii} > 0\} = 1$ per ogni $i \in E$, lo stesso vale per il MCD $\{n : \tilde{p}^{(n)}_{(i,j)(i,j)} > 0\}$ per ogni $(i,j) \in E \times E$ poiché $\tilde{p}^{(n)}_{(i,j)(i,j)} = p^{(n)}_{ii} p^{(n)}_{jj}$.

Anche l'irriducibilità si trasferisce alla catena accoppiata, infatti per ogni scelta di indici $i,j,k,h \in E$, dal lemma 2.9.2 sappiamo che esistono due indici $n_1(i,k)$ ed $n_2(j,h)$ tali che $p^{(n)}_{ik} > 0$ per $n \geq n_1$ e $p^{(n)}_{jh} > 0$ per $n \geq n_2$, quindi prendendo $n_0(i,j,k,h) = \max(n_1(i,k), n_2(j,h))$ otteniamo

$$\tilde{p}^{(n)}_{(i,j)(k,h)} > 0, \qquad \forall\, n \geq n_0.$$

In altre parole due qualsiasi stati in $E \times E$ comunicano tra loro, cioè la catena accoppiata è irriducibile.

Infine se la catena originale è persistente positiva, deve esserlo anche la catena $\{(X_n, Y_n)\}_n$. Siano infatti

$$\tau_{(i,j)} = \inf\{n :\ (X_n, Y_n) = (i,j)\}, \quad \tau^X_i = \inf\{n :\ X_n = i\}, \quad \tau^Y_j = \inf\{n :\ Y_n = j\},$$

per dimostrare la persistenza positiva della catena calcoliamo direttamente $E_{(i,j)}(\tau_{(i,j)})$ tenendo a mente che separatamente sia X_n sia Y_n sono persistenti positive e dunque $E_i(\tau^X_i) < +\infty$ e $E_j(\tau^Y_j) < +\infty$. Per indipendenza delle sue successioni si ha

$$\begin{aligned}
P_{(i,j)}(\tau_{(i,j)} = k) &= P(X_0 = i, X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i, Y_0 = j, Y_k = j) \\
&+ P(X_0 = i, X_k = i, Y_0 = j, Y_1 \neq j, \dots, Y_{k-1} \neq j, Y_k = j) \\
&= P_i(\tau^X_i = k) p^{(k)}_{jj} + P_j(\tau^Y_j = k) p^{(k)}_{ii}
\end{aligned}$$

e siccome $p^{(k)}_{jj}$, $p^{(k)}_{ii} \leq 1$ si ha

$$\begin{aligned}
E_{(i,j)}(\tau_{(i,j)}) &= \sum_{k=1}^{+\infty} k P_{(i,j)}(\tau_{(i,j)} = k) \leq \sum_{k=1}^{+\infty} k P_i(\tau^X_i = k) + \sum_{k=1}^{+\infty} k P_j(\tau^Y_j = k) \\
&= E_i(\tau^X_i) + E_j(\tau^Y_j) < +\infty.
\end{aligned}$$

Riassumendo abbiamo dimostrato che la catena $\{(X_n, Y_n)\}_n$ è di Markov, irriducibile, aperiodica, persistente positiva; possiamo dunque applicare il teorema di esistenza ed unicità e concludere che esiste un'unica probabilità invariante. Vogliamo verificare che essa è data da $\pi(i,j) = \pi_i \pi_j$, $(i,j) \in E \times E$.

1. $\pi_i\pi_j \geq 0$, per ogni $(i,j) \in E \times E$;
2. $\sum_{(i,j)\in E\times E} \pi_i\pi_j = \sum_{i\in E} \pi_i \sum_{j\in E} \pi_j = 1$;
3. Invarianza:
$$\sum_{(i,j)\in E\times E} \pi_i\pi_j\tilde{p}_{(i,j)(k,h)} = \sum_{(i,j)\in E\times E} \pi_i\pi_j p_{ik}p_{jh} = \sum_{i\in E} \pi_i p_{ik}\pi_i \sum_{j\in E} \pi_j p_{jh} = \pi_k\pi_h$$

Per unicità abbiamo trovato la misura invariante.

Non ci rimane che dimostrare la tesi.

Poiché tutti gli stati della catena accoppiata sono persistenti positivi, scelto (i_0, i_0), per ogni altro (i,j) si ha

$$P_{(i,j)}((X_n, Y_n) = (i_0, i_0), \text{ i.s. }) = 1$$

e dunque $P_{(i,j)}(\tau_{(i_0,i_0)} < +\infty) = 1$. D'altra parte, per $m \leq n$ e per ogni (k,h) si ha per la proprietà di Markov

$$\begin{aligned}
&P_{(i,j)}((X_n, Y_n) = (k,h),\, \tau_{(i_0,i_0)} = m) \\
&= P_{(i,j)}((X_n,Y_n) = (k,h), (X_m,Y_m) = (i_0,i_0), (X_{m-1},Y_{m-1}) \neq (i_0,i_0), .., (X_1,Y_1) \neq (i_0,i_0)) \\
&= \tilde{p}^{(n-m)}_{(i_0,i_0)(k,h)} P_{(i,j)}((X_m, Y_m) = (i_0,i_0), (X_{m-1}, Y_{m-1}) \neq (i_0,i_0), \ldots, (X_1, Y_1) \neq (i_0,i_0)) \\
&= \tilde{p}^{(n-m)}_{(i_0,i_0)(k,h)} P_{(i,j)}(\tau_{(i_0,i_0)} = m) = p^{(n-m)}_{i_0,k} p^{(n-m)}_{i_0,h} P_{(i,j)}(\tau_{(i_0,i_0)} = m),
\end{aligned}$$

perció sommando

$$\begin{aligned}
P_{(i,j)}(X_n = k, \tau_{(i_0,i_0)} = m) &= \sum_{h\in E} P_{(i,j)}((X_n, Y_n) = (k,h),\, \tau_{(i_0,i_0)} = m) \\
&= \sum_{h\in E} p^{(n-m)}_{i_0,k} p^{(n-m)}_{i_0,h} P_{(i,j)}(\tau_{(i_0,i_0)} = m) \\
&= p^{(n-m)}_{i_0,k} P_{(i,j)}(\tau_{(i_0,i_0)} = m).
\end{aligned}$$

Analogamente

$$P_{(i,j)}(Y_n = h, \tau_{(i_0,i_0)} = m) = p^{(n-m)}_{i_0,h} P_{(i,j)}(\tau_{(i_0,i_0)} = m).$$

Se specifichiamo le precedenti relazioni per $k = h$ e sommiamo per $m = 1, \ldots, n$, si ottiene

$$P_{(i,j)}(X_n = k, \tau_{(i_0,i_0)} \leq n) = P_{(i,j)}(Y_n = k, \tau_{(i_0,i_0)} \leq n)$$

e quindi vale anche

$$\begin{aligned}
P_{(i,j)}(X_n = k) &= P_{(i,j)}(X_n = k, \tau_{(i_0,i_0)} \leq n) + P_{(i,j)}(X_n = k, \tau_{(i_0,i_0)} > n) \\
&= P_{(i,j)}(Y_n = k, \tau_{(i_0,i_0)} \leq n) + P_{(i,j)}(X_n = k, \tau_{(i_0,i_0)} > n) \\
&\leq P_{(i,j)}(Y_n = k) + P_{(i,j)}(\tau_{(i_0,i_0)} > n).
\end{aligned}$$

D'altronde scambiando i ruoli di X ed Y si può anche ottenere

$$P_{(i,j)}(X_n = k) = P_{(i,j)}(Y_n = k) + P_{(i,j)}(\tau_{(i_0,i_0)} > n),$$

da cui

$$|P_{(i,j)}(Y_n = k) - P_{(i,j)}(X_n = k)| \leq P_{(i,j)}(\tau_{(i_0,i_0)} > n),$$

ma $P_{(i,j)}(X_n = k) = P(X_n = k|X_0 = i, Y_0 = j) = P(X_n = k|X_0 = i) = p_{ik}^{(n)}$ per indipendenza, ovvero la precedente relazione ci dice

$$|p_{ik}^{(n)} - p_{jk}^{(n)}| \leq P_{(i,j)}(\tau_{(i_0,i_0)} > n) \longrightarrow 0 \quad \text{per } n \to +\infty,$$

per la persistenza positiva dello stato (i_0, i_0). Quindi $\lim\limits_{n\to+\infty} |p_{ik}^{(n)} - p_{jk}^{(n)}| = 0$, che implica

$$\lim_{n\to+\infty} |\pi_k - p_{jk}^{(n)}| = \lim_{n\to+\infty} |\sum_{i\in E} \pi_i p_{ik}^{(n)} - \sum_{i\in E} \pi_i p_{jk}^{(n)}| = \lim_{n\to+\infty} |\sum_{i\in E} \pi_i (p_{ik}^{(n)} - p_{jk}^{(n)})| = 0$$

grazie all'M-test di Weierstrass (lemma 2.9.3), da cui la tesi. $\square$

Lemma 2.9.3 : *M-test di Weierstrass*
Se per una successione a doppio indice $\{x_{nk}\}_{n,k\in\mathbb{N}}$ si ha che per ogni k fissato esistono x_k e $M_k \geq 0$ tali che

$$\lim_{n\to=+\infty} x_{nk} = x_k, \quad , |x_k| \leq M_k, \quad \sum_k M_k < +\infty,$$

allora

$$\lim_{n\to+\infty} \sum_k x_{nk} = \sum_k x_k$$

Dimostrazione:

$$|\sum_k x_{nk} - \sum_k x_k| \leq \sum_{k\leq k_0} |x_{nk} - x_k| + 2\sum_{k>k_0} M_k \longrightarrow 0$$

scegliendo opportunamente k_0 dalla definizione di limite. $\square$

Si può dare una forma alternativa del teorema ergodico, sotto una condizione che a volte è più facile da verificare.

Definizione 2.9.3 : *Si dice che una matrice di transizione $\mathcal{P}$ su $E \times E$ è* **regolare** *se esiste un indice $n \in \mathbb{N}$ tale che $p_{ij}^{(n)} > 0$ per ogni $i, j \in E$. Una CM è detta regolare se lo è la sua matrice di transizione.*

Diamo subito un criterio di regolarità almeno nel caso di spazio degli stati finito.

Proposizione 2.9.2 : *Si abbia una CM spazio degli stati* $|E| < +\infty$*e matrice di transizione* $\mathcal{P}$*, se tutti gli stati comunicano tra loro ed esiste* $h \in E$ *tale che* $p_{hh} > 0$*, allora la catena è regolare.*

Dimostrazione: per ogni scelta di $i, j \in E$, sappiamo che esiste un indice $n(i, j)$ tale che $p_{i,j}^{(n(i,j))} > 0$. Essendo E finito, possiamo prendere $n = \max_{i,j \in E} n(i, j)$, vogliamo dimostrare che $\mathcal{P}^{2n}$ ha senz'altro tutti gli elementi positivi. Infatti si ha per ogni $i, j \in E$

$$p_{ij}^{(2n)} \geq p_{ih}^{(n(i,h))} p_{hh} \dots p_{hh} p_{hj}^{(n(h,j))} > 0,$$

dove p_{hh} è stato preso $2n - n(i, h) - n(h, j)$ volte. □

Cerchiamo ora di capire le relazioni tra tutte le nozioni introdotte.

1. Per definizione se una catena è regolare deve essere irriducibile.

2. Una catena può essere irriducibile ma non regolare: presentiamo due esempi.

   (a) Si abbia la CM con matrice di transizione

   $$\mathcal{P} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix},$$

   disegnandone il grafo è immediato vedere che la CM è irriducibile. Questa matrice non è aperiodica (verificare che il periodo è 2) e non è neanche regolare, poichè per ogni $n \in \mathbb{N}$, $\mathcal{P}^{2n+1} = \mathcal{P}$, mentre

   $$\mathcal{P}^{2n} = \mathcal{P}^2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

   quindi non esiste alcuna potenza tale che $\mathcal{P}^n$ abbia tutti gli elementi diversi da 0.

   (b) Supponiamo che $E = \mathbb{N}$ e di avere una CM rappresentata dal seguente grafo

Di nuovo, la CM è irriducibile e quindi tutti gli stati devono avere lo stesso carattere. D'altra parte se consideriamo le probabilità di transizione $p^{(n)}_{2^i+1,2^i+1}$ queste sono non nulle soltanto se $n \geq 2^i+1$. Dunque non vi potrà mai essere una potenza della matrice di transizione con elementi tutti non nulli.

3. Una catena può essere aperiodica, ma non irriducibile, si consideri

$$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

   allora tutti gli stati hanno periodo 1 (verificare), ma vi sono due classi chiuse ed irriducibili.

4. Irriducibile non implica aperiodico, basti vedere l'esempio 2(a).

5. Una catena aperiodica non è necessariamente regolare, per esempio potrebbe non essere irriducibile.

6. Una catena regolare deve essere irriducibile ed aperiodica. Infatti è irriducibile per definizione. Vogliamo dimostrare che preso uno stato i arbitrario, questo ha periodo 1. Poichè la catena è regolare, esiste un indice n tale che $p^{(n)}_{ii} > 0$. Inoltre la CM è irriducibile, allora deve esistere almeno uno stato h_0 tale che $p_{h_0 i} > 0$. Se per assurdo si ha $p_{hi} = 0$ per ogni $h \in E$, poichè $\sum_{h\in E} p_{ih} = 1$, deve esistere un h_1 tale che $p_{ih_1} > 0$, ovvero i comunica con h_1, ma non è vero il viceversa e si ha una contraddizione con l'irriducibilità della catena.

   Possiamo quindi concludere che

$$p^{(n+1)}_{ii} \geq p^{(n)}_{ih_0} p_{h_0 i} > 0,$$

   ma n ed $n+1$ sono sicuramente coprimi e dunque il periodo di i è 1.

7. Se $|E| < +\infty$ una CM irriducibile ed aperiodica è anche regolare. Infatti avendo un numero finito di stati possiamo applicare il Lemma 2.9.2 a tutti loro e prendere $n = \max_{i,j\in E} n(i,j)$.

   Invece se $|E| = +\infty$ può esistere una catena aperiodica ed irriducibile che non sia regolare, basta vedere la CM dell'esempio 2(b).

Possiamo infine enunciare

Teorema 2.9.2 : *Sia $\{X_n\}_n$ una Catena di Markov persistente positiva e regolare con spazio degli stati E e sia π la sua probabilità invariante. Allora*

$$\lim_{n\to+\infty} p_{ij}^{(n)} = \pi_j, \qquad \forall\, i, j, \in E.$$

Dimostrazione: vedere punto 6 precedente. □

In conclusione i due teoremi sono equivalenti quando lo spazio degli stati è finito, mentre le condizioni del teorema 2.9.2 (e quindi il teorema è più debole) implicano quelle del teorema 2.9.1 se $|E| = +\infty$.

2.10 Esercizi

1. Dare la definizione di catena di Markov regolare ed enunciare il criterio di regolarità (per catene di Markov finite).

2. Un ispettore bancario visita ripetutamente ogni mese 5 istituti bancari, partendo dalla sua sede (che indichiamo con 1). Sia X_n la posizione dell'ispettore al tempo n e supponiamo che individui una catena di Markov con la seguente matrice di transizione.

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

   (a) Disegnare il grafo corrispondente e classificare gli stati della catena.
   (b) Calcolare tutte le probabilità invarianti.
   (c) Calcolare approssimativamente, per n grande, $P(X_n = 5, X_{n-1} = 6 | X_0 = 1)$.

3. Il numero di pezzi lavorati in un'officina all'ora n è rappresentato da una catena di Markov X_n con spazio degli stati $E = \{0, 1, 2, 3, 4\}$ e matrice di transizione

$$\begin{pmatrix} \frac{3}{8} & \frac{1}{2} & \frac{1}{8} & 0 & 0 \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{4} & \frac{1}{8} & \frac{3}{8} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{3}{8} & \frac{5}{8} \end{pmatrix}$$

(a) Disegnare il grafo corrispondente, classificare gli stati della catena e dire se la catena è regolare.

(b) Calcolare tutte le probabilità invarianti.

(c) Calcolare approssimativamente, per n grande, $E(X_n)$.

4. Sia $\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$ una matrice di transizione per una catena di Markov X_n. Calcolare approssimativamente $E(X_n)$, per n grande.

5. Sia X_n la catena di Markov su $\{1, 2, 3\}$ con la seguente matrice di transizione

$$\begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

(a) Trovare le classi chiuse ed irriducibili e dire se la catena è regolare.

(b) Calcolare le probabilità invarianti.

(c) Calcolare approssimativamente $E(X_n)$ per n grande.

6. Cinque macchinari sono collegati da percorsi secondo il seguente schema.

Un carrello di servizio passa, per la manutenzione, da una macchina all'altra con uguale probabilità, seguendo i percorsi tracciati.

Sia X_n la posizione del carrello al tempo n. Assumendo che questa sia una catene di Markov

(a) Dire se è regolare

(b) Calcolare tutte le eventuali probabilità invarianti.
(c) Calcolare approssimativamente $E(X_n)$.

7. Si consideri la catena di Markov X_n a tre stati tale che, in un passo, da 1 si possa andare solo in 2, da 2 si possa andare in 2 od in 3 con uguale probabilità e da 3 si possa andare solo in 1. Inizialmente ogni stato è ugualmente probabile.
   (a) Calcolare $P(X_2 = 1)$ e $P(X_2 = 1, X_4 = 3)$, $P(X_0 = 1, X_2 = 3)$ e $P(X_1 = 1, X_3 = 3)$.
   (b) Dire se la catena è regolare, giustificando la risposta.
   (c) Dire se esistono, ed eventualmente calcolare, tutte le distribuzioni invarianti della catena.
   (d) Per n grande, calcolare approssimativamente
   $$P(X_n = 1) \quad \text{e} \quad P(X_n = 1, X_{n+2} = 3).$$
   e $E_i(X_n X_{n+2})$, per $i = 1, 2, 3$ ed $E(X_n X_{n+2})$.

8. Dare un esempio, o dire se non è possibile, di una CM:
   (a) irriducibile;
   (b) aperiodica;
   (c) irriducibile , ma non aperiodica;
   (d) aperiodica, ma non irriducibile;
   (e) irriducibile ed aperiodica;
   (f) regolare;
   (g) regolare, ma non irriducibile;
   (h) regolare, ma non aperiodica;
   (i) aperiodica ed irriducibile, ma non regolare

9. Dare un esempio di CM di periodo 3 e di periodo 4.

10. Dare un esempio di CM persistente positiva, persistente nulla e transiente.

11. Dare un esempio di una CM per la quale non esista alcuna probabilità invariante.

12. Dare un esempio di CM per la quale la probabilità invariante esiste, ma non vale il teorema ergodico.

2.11 Esempi ed applicazioni

Definizione 2.11.1 : *Una distribuzione di probabilità π su uno spazio E si dice* **reversibile** *per una matrice di transizione $\mathcal{P}$ su E se vale*

$$\pi_i p_{ij} = \pi_j p_{ji}, \qquad \forall\, i, j \in E \tag{2.16}$$

la precedente equazione è detta del bilancio dettagliato.

Osservazione 2.11.1 : *Se π è reversibile per $\mathcal{P}$, allora è anche invariante. (verificare)*

Abbiamo il seguente teorema che ci permette di costruire catene di Markov che abbiano come probabilità invariante una distribuzione preassegnata. Questa tecnica è particolarmente utile in problemi di simulazione.

Teorema 2.11.1 : *Sia E uno spazio degli stati con $|E| < +\infty$. Se π, distribuzione di probabilità su E, non coincide con la distribuzione uniforme, allora esiste sempre una matrice di transizione $\mathcal{P}$ ed una catena di Markov ad essa associata $\{X_n\}_n$, regolare che ha π come distribuzione reversibile (e quindi invariante).*

Dimostrazione: Senza perdere di generalità assumiamo che $\pi_i > 0$ per ogni $i \in E$. Se così non fosse potremmo semplicemente restringere lo spazio degli stati. Sia $\mathcal{Q}$ una qualsiasi matrice di transizione $|E| \times |E|$ irriducibile e simmetrica (e quindi bisocastica) e si definisca una nuova matrice $\mathcal{P}$ nella seguente maniera

$$p_{ij} = \begin{cases} q_{ij}, & \pi_j \geq \pi_i, \quad i \neq j \\[2ex] q_{ij}\dfrac{\pi_j}{\pi_i}, & \pi_j < \pi_i, \quad i \neq j \\[2ex] 1 - \displaystyle\sum_{j \neq i} p_{ij}, & \qquad\qquad i = j \end{cases}$$

La precedente definizione non sarebbe chiaramente possibile (non distinguerebbe tra i due casi) qualora la distribuzione π fosse uniforme.

Anche la matrice $\mathcal{P}$ cosí definita è una matrice di transizione (verificare). Inoltre la distribuzione π è reversibile e quindi invariante per $\mathcal{P}$. Infatti per $i \neq j$ se $\pi_i = \pi_j$, essendo $\mathcal{Q}$ simmetrica la relazione di reversibilità (2.16) è automaticamente soddisfatta, per $\pi_i < \pi_j$, vale

$$\pi_i p_{ij} = \pi_i q_{ij} = \pi_i q_{ji} = \pi_i p_{ji} \frac{\pi_j}{\pi_i}$$

se invece $\pi_i > \pi_j$ si ha analogamente

$$\pi_i p_{ij} = \pi_i q_{ij} \frac{\pi_j}{\pi_i} = \pi_j q_{ij} = \pi_j q_{ji} = \pi_j p_{ji}.$$

Per $i = j$ la (2.16) è di nuova ovvia.

Infine possiamo dimostrare che, quando π_i non è costante in i, allora $\mathcal{P}$ è addirittura regolare.

Prima di tutto, grazie all'irriducibilità di $\mathcal{Q}$ è possibile trovare una coppia di indici i_0, j_0 tali che $q_{i_0 j_0} > 0$ e $\pi_{j_0} < \pi_{i_0}$. Per assicurarsi dell'esistenza di tali indici, consideriamo l'insieme

$$M = \{i \in E \,:\, \pi_i = \max_{j \in E} \pi_j\}$$

che è senz'altro non vuoto poichè E è finito. Poiché $\mathcal{Q}$ è irriducibile, devono esistere due indici $i_0 \in M$ e $j_0 \in M^c$ che connettono i due insiemi, ovvero $q_{i_0 j_0} > 0$ e per costruzione non si può che avere $\pi_{i_0} > \pi_{j_0}$ (la disuguaglianza stretta dovuta al fatto che la distribuzione non è uniforme).

Con questa scelta di indici si ha

$$\begin{aligned}
p_{i_0 i_0} &= 1 - \sum_{j \neq i_0} p_{i_0 j} = 1 - \sum_{j \neq i_0, j_0} p_{i_0 j} - p_{i_0 j_0} = 1 - \sum_{j \neq i_0, j_0} p_{i_0 j} - q_{i_0 j_0} \frac{\pi_{j_0}}{\pi_{i_0}} \\
&\geq 1 - \sum_{j \neq i_0, j_0} q_{i_0 j} - q_{i_0 j_0} \frac{\pi_{j_0}}{\pi_{i_0}} = 1 - \sum_{j \neq i_0} q_{i_0 j} + q_{i_0 j_0} (1 - \frac{\pi_{j_0}}{\pi_{i_0}}) \\
&= q_{i_0 i_0} + q_{i_0 j_0} (1 - \frac{\pi_{j_0}}{\pi_{i_0}}) \geq q_{i_0 j_0} (1 - \frac{\pi_{j_0}}{\pi_{i_0}}) > 0,
\end{aligned}$$

ovvero la catena è regolare essendo già irriducibile.

Il precedente teorema dovuto a Metropolis, Rosenbluth, Rosenbluth, Teller, Teller, fornisce quindi un algoritmo, detto di Metropolis per costruire una CM con distribuzione invariante prefissata. Questo algoritmo trova importanti applicazioni nel campo della simulazione, quando per esempio si volgiono calcolare funzionali dipendenti da una distribuzione di probabilità non completamente nota. Infatti in questo caso si applica l'algoritmo di Metropolis e il teorema ergodico per approssimare il funzionale dipendente da π con lo stesso funzionale calcolato sulla distribuzione marginale della catena a tempo grande.

Esempio 2.11.1 : *Su $E = \{1, 2, 3, 4\}$ si ha la distribuzione $(\frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4})$ e vogliamo costruire la matrice di transizione dell'algoritmo di Metropolis. Prendiamo*

$$\mathcal{Q} = \begin{pmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{4} & \frac{1}{12} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$$

Poiché $\pi_1 < \pi_j$ per $j = 2, 3, 4$, la prima riga di $\mathcal{P}$ coincide con la prima riga di $\mathcal{Q}$. Inoltre $\pi_2 = \pi_4 < \pi_3$, dunque

$$\begin{cases} p_{21} = q_{21}\frac{\pi_1}{\pi_2} = 0 \\ p_{23} = q_{23} = \frac{1}{4} \\ p_{24} = q_{24} = \frac{1}{4} \\ p_{22} = 1 - \frac{1}{2} = \frac{1}{2} \end{cases}, \qquad \begin{cases} p_{31} = q_{31}\frac{\pi_1}{\pi_3} = \frac{2}{3}\frac{1}{3} = \frac{1}{9} \\ p_{32} = q_{32}\frac{\pi_2}{\pi_3} = \frac{1}{4}\frac{2}{3} = \frac{1}{6} \\ p_{34} = q_{34}\frac{\pi_4}{\pi_3} = \frac{1}{12}\frac{2}{3} = \frac{1}{18} \\ p_{33} = 1 - \frac{4}{18} = \frac{7}{9} \end{cases}, \qquad \begin{cases} p_{41} = q_{41}\frac{\pi_1}{\pi_4} = 0 \\ p_{42} = q_{42} = \frac{1}{4} \\ p_{43} = q_{43} = 0 \\ p_{33} = 1 - \frac{1}{4} = \frac{3}{4}. \end{cases}$$

Riassumendo

$$\mathcal{P}\begin{pmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{9} & \frac{1}{6} & \frac{1}{18} & \frac{7}{9} \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$$

Vediamo ora un'applicazione dell'algoritmo di Metropolis, chiamata simulated annealing, che può tornare utile anche in meccanica statistica.

Esempio 2.11.2 : *Simulated annealing*

Su uno spazio degli stati E finito sia definita una distribuzione di probabilità parametrizzata, determinata da una funzione Hamiltoniana $H : E \longrightarrow \mathbb{R}$, tramite

$$\pi_i^\epsilon = \frac{\mathrm{e}^{-\frac{H(i)}{\epsilon}}}{Z^\epsilon}, \qquad \epsilon > 0,$$

dove Z^ϵ è l'opportuna costante di normalizzazione affinchè si abbia una probabilità (potrebbe essere non nota).

A partire da una qualsiasi matrice di transizione irriducibile e simmetrica, applicando l'algoritmo di Metropolis otterremo una famiglia di matrici di transizione $\mathcal{P}^\epsilon$ ognuna con π^ϵ come distribuzione invariante, dove

$$p_{ij}^\epsilon = \begin{cases} q^{ij} & H(j) \leq H(i),\ j \neq i \\ q_{ij}\mathrm{e}^{-\frac{H(i)-H(j)}{\epsilon}} & H(j) > H(i),\ j \neq i \\ 1 - \sum_{j\neq i} p_{ij} & i = j \end{cases}$$

sono i suoi elementi. Dunque si sceglie lo stato j a partire da i con transizione q_{ij} se $H(j) \leq H(i)$, altrimenti si effettua la transizione con probabilità $q_{ij}\mathrm{e}^{-\frac{H(i)-H(j)}{\epsilon}}$ e viene rifiutata con probabilità $1 - \sum_{j\neq i} p_{ij}$.

Come visto, non occorre conoscere la costante Z^ϵ, inoltre potremmo avere una conoscenza parziale della funzione H. Per esempio potrbbe essere particolarmente *difficile determinare i suoi punti di minimo, che denotiamo con $i_1, \dots, i_m$. Sui punti di minimo, se riscriviamo la nostra distribuzione come*

$$\pi_i^\epsilon = \frac{\mathrm{e}^{-\frac{H(i)}{\epsilon}}}{\mathrm{e}^{-\frac{H(i_1)}{\epsilon}} + \cdots + \mathrm{e}^{-\frac{H(i_m)}{\epsilon}} + \sum_{j \neq i_1,..,i_m} \mathrm{e}^{-\frac{H(j)}{\epsilon}}}$$

poiché gli indici nella sommatoria hanno tutti $H(j) \geq H(i_k)$ ($k = 1, \dots, m$), per $i = i_1, \dots, i_m$ si ha che

$$\pi_i^\epsilon \longrightarrow \frac{1}{m}, \qquad \text{per } \epsilon \to 0$$

ovvero si tende alla distribuzione uniforme concentrata sui punti di minimo. Perciò per determinare numericamente i punti di minimo di H si può costruire la CM secondo l'algoritmo di Metropolis, implementarla e sfruttare il teorema ergodico osservando le simulazioni della catena a tempi grandi.

Per chiudere il capitolo, presentiamo alcuni esempi guida di catene a stati numerabili, che sono alla base dello studio di processi Markoviani più complessi.

Esempio 2.11.3 : **Catene di nascita e morte**
Consideriamo su $E = \mathbb{N}$ la matrice di transizione

$$\mathcal{P} = \begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ q_1 & r_1 & p_1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & q_2 & r_2 & p_2 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ 0 & 0 & 0 & \dots & q_{m-1} & r_{m-1} & p_{m-1} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{pmatrix}$$

che può essere vista come il limite della catena di nascita e morte su una popolazione al massimo di m individui già studiata nell'esempio 2.5.1(a). In altre parole, possiamo avere una popolazione che non ha limiti di dimensione ma dove ad ogni istante può nascere o morire al più un individuo.

Prima di tutto vogliamo stabilire sotto quali condizioni la catena definita da $\mathcal{P}$ risulta persistente o transiente. Notiamo che se $p_i, q_i > 0$ $\forall i$, allora la CM è irriducibile e tutti gli stati devono avere lo stesso carattere. Ci poniamo in questo caso, poiché altrimenti potremmo tagliare una parte dello spazio degli stati e ridurci alla classe chiusa ed irriducibile di dimensione infinita.

Se, come al solito, indichamo con $\tau_i = \inf\{n : X_n = i\}$ e con $\rho_{ij} = P_i(\tau_j < +\infty)$ le probabilità di arrivo negli stati, ricordiamo che vale l'equazione di rinnovo

$$\rho_{ij} = p_{ij} + \sum_{h \neq j} p_{ih}\rho_{hj}$$

che per $i = j = 0$ diventa

$$\rho_{00} = p_{00} + p_{01}\rho_{10} = r_0 + p_0\rho_{10} = 1 \qquad \Longleftrightarrow \qquad \rho_{10} = 1.$$

Per dimostrare che $\rho_{10} = 1$, fondamentalmente ci riconduciamo al caso finito e quindi passiamo al limite. Infatti per ogni $m \in \mathbb{N}$ senz'altro abbiamo

$$P_1(\tau_0 < +\infty) \geq P_1(\tau_0 < \tau_m) = P(\textit{arrivo in } 0 \textit{ prima che in } m) \tag{2.17}$$

Per calcolare questa probabilità, possiamo modificare il tratto della catena fino allo stato m

nella catena ad m stati con 0 ed m stati assorbenti

Per quanto abbiamo visto per le catene di nascita e morte a stati finiti

$$P_1(\tau_0 < \tau_m) = \lambda_1^{\{0\}} = 1 - \frac{1}{\sum_{i=0}^{m-1} \gamma_i},$$

con $\gamma_0 = 1$ e $\gamma_i = \dfrac{q_1 \cdots q_i}{p_1 \cdots p_i}$. La disuguaglianza (2.17) vale per ogni m, dunque

$$P_1(\tau_0 < +\infty) \geq \lim_{m\to+\infty} P_1(\tau_0 < \tau_m) = \lim_{m\to+\infty} \left[1 - \frac{1}{\sum_{i=0}^{m-1} \gamma_i}\right]$$

che risulta uguale ad 1 se e soltanto se

$$\sum_{i=0}^{+\infty} \gamma_i = +\infty. \tag{2.18}$$

D'altra parte, poiché per arrivare in m a partire da 0 occorrono almeno $m - 1$ passi, abbiamo

$$\begin{aligned} P_1(\tau_0 < +\infty) &= \lim_{m\to+\infty} P_1(\tau_0 < m-1) = \lim_{m\to+\infty} P_1(\tau_0 < \tau_m) \\ &= 1 - \frac{1}{a+1} \qquad \Longleftrightarrow \sum_{i=1}^{+\infty} \gamma_i = a < +\infty. \end{aligned}$$

Quindi abbiamo dimostrato che la catena è persistente se e soltanto se $\sum_{i=0}^{+\infty} \gamma_i = +\infty$.
In particolare se $p_i = p$ *e* $q_i = q$ *per ogni* $i \in \mathbb{N}$, *la catena è transiente se e soltanto se* $\frac{q}{p} < 1$.

Vediamo ora sotto quali condizioni la catena di nascita e morte ammette la distribuzione invariante. Impostando il sistema $\pi^T = \pi^T \mathcal{P}$ *si ha*

$$\begin{cases} \pi_0 = \pi_0 r_0 + \pi_1 q_1 \\ \vdots \\ \pi_k = p_{k-1}\pi_{k-1} + r_k \pi_k + q_{k+1}\pi_{k+1} \\ \vdots \\ \sum_{k=0}^{+\infty} \pi_k = 1 \end{cases}$$

da cui

$$\begin{cases} q_1\pi_1 - (1 - r_0)\pi_0 = q_1\pi_1 - p_0\pi_0 = 0 \\ \vdots \\ q_{k+1}\pi_{k+1} - p_k\pi_k = q_k\pi_k - p_{k-1}\pi_{k-1} \end{cases}$$

quindi per ricorrenza si ottiene

$$q_{k+1}\pi_{k+1} - p_k\pi_k = 0 \quad \Rightarrow \quad \pi_{k+1} = \frac{p_k}{q_{k+1}}\pi_k.$$

Sostituendo ogni equazione nella successiva si ottiene intermini di π_0

$$\begin{cases} \pi_1 = \frac{p_0}{q_1}\pi_0 \\ \pi_2 = \frac{p_1}{q_2}\pi_1 = \frac{p_1 p_0}{q_2 q_1}\pi_0 \\ \vdots \\ \pi_k = \frac{p_{k-1}\cdots p_0}{q_k \cdots q_1}\pi_0 = \frac{p_0}{p_k}\frac{1}{\gamma_k}\pi_0 \\ \sum_{k=0}^{+\infty} \pi_k = \pi_O\Big(1 + \sum_{k=1}^{+\infty} p_0 \frac{1}{p_k\gamma_k}\Big) \end{cases}$$

che possiede una soluzione non triviale se e soltanto se la serie $\sum_{k=1}^{+\infty} p_0 \frac{1}{p_k\gamma_k}$ *converge.*

Riassumendo

$$
\begin{aligned}
&\sum_{i=0}^{+\infty} \gamma_i < +\infty \quad \Rightarrow \quad \textit{catena transiente}\\
&\sum_{i=0}^{+\infty} \gamma_i = +\infty, \quad \sum_{k=1}^{+\infty} p_0 \frac{1}{p_k \gamma_k} = +\infty \Rightarrow \quad \textit{catena persistente nulla}\\
&\sum_{i=0}^{+\infty} \gamma_i = +\infty, \quad \sum_{k=1}^{+\infty} p_0 \frac{1}{p_k \gamma_k} < +\infty \Rightarrow \quad \textit{catena persistente positiva}
\end{aligned}
$$

Analizziamo un'altra catena a stati numerabili piuttosto comune

Esempio 2.11.4 : *Sia $E = \mathbb{N}$ e consideriamo la catena di Markov con matrice di transizione*

$$
\mathcal{P} = \begin{pmatrix}
q_1 & p_1 & 0 & 0 & 0 & . & . & . & . & . \\
q_2 & 0 & p_2 & 0 & 0 & . & . & . & . & . \\
q_3 & 0 & 0 & p_3 & 0 & . & . & . & . & . \\
\vdots & & & & & & & & & \\
q_n & 0 & 0 & 0 & 0 & . & . & p_n & 0 & . \\
\vdots & & & & & & & & &
\end{pmatrix}
$$

Disegnandone il grafo associato, è immediato vedere che se $p_i > 0$ per ogni i e $q_i = 0$ solo per un numero finito di indici $i \geq 2$, allora la catena è irriducibile ed aperiodica ($q_1 > 0$), dunque gli stati hanno tutti lo stesso carattere.

D'altra parte essa non è regolare, poichè $p_{ii}^{(n)} > 0$ solo per $n \geq i+1$.

Per stabilire il carattere della catena applichiamo il teorema 2.3.2 con $j_0 = 1$, ottenendo il sistema

$$
\begin{cases}
x_2 = p_2 x_3 \\
x_3 = p_3 x_4 \\
\vdots \\
x_k = p_k x_{k+1} \\
\vdots \\
0 \leq x_i \leq 1
\end{cases}.
$$

Per sostituzione si ottiene che per ogni $k \geq 2$ la soluzione è data da

$$
x_k = \frac{x_2}{p_2 p_3 \cdots p_{k-1}}
$$

che è non triviale e limitata superiormente da 1 se e soltanto se per ogni k esiste un $\epsilon > 0$ tale che

$$\prod_{j=2}^{k-1} p_j > \epsilon, \quad \textit{ovvero} \quad \prod_{j=2}^{k-1}(1 - q_j) > \epsilon. \tag{2.19}$$

Passando ai logaritmi si verifica che (2.19) è valida se la serie $\sum_{j\geq 2} q_j$ converge. Perciò se $\sum_{j\geq 2} q_j < +\infty$, gli stati sono tutti transienti ed il teorema ergodico non vale, in quanto non esiste alcuna distribuzione invariante.

Se invece $\sum_{j\geq 2} q_j = +\infty$, gli stati sono tutti persistenti e per cercare la probabilità stazionaria impostiamo il sistema $\pi^T = \pi^T \mathcal{P}$.

$$\begin{cases} \pi_1 = \sum_{i=1}^{+\infty} q_i \pi_i \\ \pi_2 = p_1 \pi_1 \\ \pi_3 = p_2 \pi_2 \\ \vdots \\ \pi_k = p_{k-1}\pi_{k-1} \\ \vdots \\ \sum_{i=1}^{+\infty} \pi_i = 1 \end{cases},$$

dal quale concludiamo che la soluzione può essere scritta come

$$\pi_k = p_{k-1}p_{k-2}\cdots p_1\pi_1, \qquad \forall\, k \geq 2.$$

Sostituendo nell'ultima condizione, si ottiene che

$$\pi_1\Big(1 + \sum_{i=2}^{+\infty} p_1 \cdots p_{i-1}\Big) = 1$$

e la soluzione è non nulla se e soltanto se $\sum_{i=2}^{+\infty} p_1 \cdots p_{i-1} < +\infty$, nel qual caso tutti gli stati sono persistenti positivi.

Per concludere questo capitolo, presentiamo un'ultimo esempio che fornisce l'accenno di una parte molto ampia della probabilità che è chiamata Teoria delle code e che ha applicazioni nella programmazione dei servizi.

Esempio 2.11.5 : File d'attesa

Un certo servizio è fornito a degli utenti e man mano che questi arrivano vengono serviti. Per ogni utente è necessario un certo tempo di servizio ed, a seconda della frequenza degli arrivi, è possibile che si formi una fila d'attesa di dimensione variabile (virtualmente illimitata). Tutte le variabili che esprimono i tempi di servizio, di arrivo etc. non sono deterministiche, bensì aleatorie, perciò un primo possibile modo di modellizzare una fila d'attesa è il seguente.

Siano $\{W_n\}_n$ e $\{Z_n\}_n$ due successioni indipendenti di v.a. i.i.d. che rappresentano

$$\begin{aligned} W_n &= \text{numero di clienti in arrivo al tempo } n \\ Z_n &= \text{numero di clienti serviti al tempo } n, \end{aligned}$$

dove, per ogni n, $W_n \sim p_W$ e $Z_n \sim p_Z$. Indichiamo con X_n il numero di clienti in fila al tempo n.

Allora, condizionatamente a $X_{n-1} = k$ la v.a.

$$X_n|(X_{n-1} = k) = k + W_n - Z_n$$

ovvero $X_n|X_{n-1}$ ha la stessa legge di $W_n - Z_n$, che ha densità

$$\begin{aligned} p_{W_n-Z_n}(k) &= P(W_n - Z_n = k) = \sum_{m=0}^{+\infty} P(W_n - Z_n = k|W_n = m)P(W_n = m) \\ &= \sum_{m=0}^{+\infty} P(Z_n = m-k)P(W_n = m) = \sum_{m=0}^{+\infty} p_Z(m-k)p_W(m). \end{aligned}$$

È chiaro che $\{X_n\}$ è una catena di Markov poichè

$$X_n = X_{n-1} + \xi_n, \qquad \xi_n = W_n - Z_n \quad \text{i.i.d., indipendenti da} \quad X_1 \dots X_{n-1}$$

e ci possiamo chiedere quale sia la matrice di transizione associata

$$P(X_n = j|X_{n-1} = i) = P(X_{n-1} + \xi_n = j|X_{n-1} = i) = P(\xi_n = j-i|X_{n-1} = i) = P(\xi_n = j-i)$$

per $j > 0$ e

$$P(X_n = 0|X_{n-1} = i) = P(W_n - Z_n \leq i).$$

Un caso semplice si ha quando si suppone $W \sim$Bernouilli(α) e $Z \sim$Bernouilli(β), da cui

$$W_Z = \begin{cases} -1 & (1-\alpha)\beta \\ 0 & \alpha\beta + (1-\alpha)(1-\beta) \\ 1 & \alpha(1-\beta) \end{cases}$$

ovvero le tre probabilità di transizione p_{ii-1}, p_{ii}, p_{ii+1}, essendo tutte le altre nulle.

Ci troviamo di nuovo nel caso di una catena di nascita e morte con $p_i = p = \alpha(1-\beta)$ e $q_i = q = \beta(1-\alpha)$, perciò

$$\begin{aligned}
&\textit{se} \quad \frac{q}{p} < 1 \quad (\beta < \alpha) \Rightarrow \textit{catena transiente} \\
&\textit{se} \quad \frac{q}{p} = 1 \quad (\beta = \alpha) \Rightarrow \textit{catena persistente nulla} \\
&\textit{se} \quad \frac{q}{p} > 1 \quad (\beta > \alpha) \Rightarrow \textit{catena persistente positiva,}
\end{aligned}$$

ma $E(W) = \alpha$ e $E(Z) = \beta$, quindi quando si programma un servizi, per non far esplodere la coda è necessario che in media si servano più utenti di quanti ne arrivano.

Studiare esempio 5.35 del libro di testo.

2.12 Esercizi di riepilogo

1. In un ufficio vi sono due sportelli. Ad intervalli di tempo unitari arriva una persona, con probabilità $\frac{1}{2}$, oppure non arriva nessuno. Nell'intervallo di tempo successivo (prima dell'eventuale arrivo di un altro utente), ognuno dei due sportelli, indipendentemente dall'altro, se occupato completa il servizio il servizio con probabilità $\frac{1}{4}$ oppure non lo completa.

   Una persona che arriva sceglie lo sportello libero, se ve n'è solo uno, ne sceglie uno a caso se sono liberi entrambi o va via senza mettersi in fila, se sono entrambi occupati. Inizialmente l'ufficio è vuoto e indichiamo con X_n e Y_n il numero di utenti rispettivamente al primo ed al secondo sportello al tempo n.

   (a) Dando per scontato che $\{(X_n, Y_n)\}$ sia una CM omogenea, scriverne la matrice di transizione.

   (b) Qual è la probabilità che ci sia un tempo in cui l'ufficio è nuovamente vuoto. Quant'è il tempo medio necessario affinchè l'ufficio si svuoti per la prima volta.

   (c) Per n grande, calcolare approssimativamente il numero medio di persone presenti nell'ufficio al tempo n

   (d) Per n grande, calcolare approssimativamente la probabilità che la catena sia in $(1, 1)$ al tempo n e che vi rimanga anche ai tempi $n + 1$ e $n + 2$.

   (e) Dire se X_n è una CM omogenea, giustificandone la risposta.

2. Per la catena di nascita e morte a stati numerabili, trovare quali condizioni danno la ricorrenza nulla o positiva, se $p_i = p$, $q_i = q$, per ogni $i \in \mathbb{N}$.

3. Due squadre, composte ognuna con un capitano ed un giocatore, tirano con l'arco ad un bersaglio con due settori concentrici. I giocatori hanno tutti la stessa abilità e ciascuno colpisce il settore centrale con probabilità $\frac{1}{6}$, il settore esterno con probabilità $\frac{1}{3}$ oppure manca il bersaglio.

   Il gioco si svolge secondo le seguenti regole: se il giocatore che sta tirando colpisce il settore centrale continua a tirare, se colpisce quello esterno passa il turno all'altro giocatore della stessa squadra, altrimenti passa l'arco al capitano della squadra avversaria. Al tempo zero inizia uno dei due capitani scelto a caso.

   Indicando con 1 e 2 il capitano e d il giocatore della squadra A e con 3 e 4 il capitano e giocatore della squadra B, sia X_n il giocatore che sta tirando al tempo n.

   (a) Dando per scontato che X_n sia una CM, scrivere la matrice di transizione associata.

   (b) Calcolare la probabilità che al tempo 0 tiri la squadra A e che continui anche al tempo successivo.

   (c) Calcolare approssimativamente che al tempo 1000 stia tirando il capitano della squadra B.

4. Si consideri la seguente matrice di transizione su $E = \{1, 2, 3, 4, 5\}$

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

   (a) Trovare le classi chiuse ed irriducibili e dire se la catena è regolare

   (b) Calcolare le probabilità invarianti.

   (c) Qual è la probabilità, partendo da 1, di arrivare prima o poi in $\{4, 5\}$?

5. Data la matrice di transizione

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

   (a) Trovare tutte le eventuali distribuzioni invarianti.

   (b) Calcolare $P_3(\exists\, n \geq 1 : X_n = 1)$ e $P_4(\exists\, n \geq 1 : X_n = 1)$.

6. Sia X_n la CM con la matrice di transizione

$$\begin{pmatrix} \frac{1}{6} & \frac{5}{6} & 0 \\ \frac{1}{6} & \frac{5}{6} & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

   (a) Studiare il carattere degli stati e trovare le eventuali distribuzioni invarianti.

   (b) Qual è la probabilità di non raggiungere mai lo stato 3?

   (c) Sia $X_0 = 1$ con probabilità 1. Supponiamo di lanciare una moneta equa, indipendente dalla catena, ogni volta che $X_n = 3$. Qual è la probabilità di avere almeno una testa nei primi tre lanci?

7. Una catena di Markov ha matrice di transizione

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

   (a) Disegnare il grafo associato alla catena e classificare gli stati.

   (b) Calcolare tutte le distribuzioni invarianti.

   (c) Calcolare approssimativamente $E(X_n X_{n-1})$.

8. Dimostrare che una catena di Markov regolare è anche irriducibile ed aperiodica. In quale caso vale anche l'implicazione contraria?

9. Supponendo di avere una catena di Markov regolare a valori in $\{1, 2, 3, ...\}$, scrivere un' approssimazione per n grande per $E(X_{n+1}/X_n)$.

10. Una catena di Markov ha spazio degli stati $E = \{1, 2, 3, 4, 5, 6, 7\}$ e la seguente matrice di transizione
$$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{3} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & 0 & \frac{1}{5} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$
    (a) Classificare gli stati e disegnare il grafo associato.
    (b) Per ogni classe chiusa ed irriducibile, calcolare le probabilità di assorbimento a partire da ogni stato transiente.

11. Una catena di Markov ha spazio degli stati $E = \{1, 2, 3, 4, 5, 6\}$ e la seguente matrice di transizione
$$\begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$
    (a) Classificare gli stati e disegnare il grafo associato.
    (b) Calcolare tutte le probabilità invarianti.

12. Un team di tre persone $(1, 2, 3)$ svolge un progetto che, per essere approvato, deve passare al vaglio di uno dei due revisori dell'ufficio $(4, 5)$. Ad ogni tempo il progetto è in mano ad uno solo dei componenti o dei revisori e viene passato dall'uno all'altro secondo il seguente schema.

Ogni componente ha uguale probabilità di passare il progetto ad un suo collega od ad un revisore con il quale è collegato. Il revisore 5 rimanda il progetto al componente che l'ha trasmesso con probabilità 1/6, lo trattiene con probabilità 1/6, altrimenti dà l'approvazione finale con probabilità 1/2. Invece il revisore 4 rimanda il progetto al componente che l'ha trasmesso con provabilità 1/3 , lo tiene con probabilità 1/6 e dà l'approvazione finale con probabilità 1/6.

(a) Sia X_n = posizione del progetto al tempo n. Dando per scontato che definisce una CM, scriverne la matrice di transizione e classificare gli stati.

(b) Qual è la probabilità che il progetto passi al revisore 5 prima di essere visto dal revisore 4?.

13. Sia $\{X_n\}$ una CM rappresentata dalla seguente matrice di transizione

$$\begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}$$

(a) Calcolare il periodo di ciascuno stato.

(b) Calcolare tutte le probabilità invarianti.

(c) Se possibile, calcolare approsimativamente per n grande $E(X_n/X_{n-1})$.

14. Sia data la seguente matrice di transizione:

$$\begin{pmatrix} \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$$

(a) Trovare tutte le probabilità invarianti.

(b) Calcolare la probabilità di raggiungere l'insieme $\{1, 4\}$ e la probabilità di raggiungere l'insieme $\{3, 5\}$ partendo dallo stato 2.

15. Sia $\{X_n\}$ una CM con matrice di transizione

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & \frac{2}{3} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

(a) Classificare gli stati.

(b) dire se esistono e calcolare tutte le probabilità invarianti.

Capitolo 3

TEORIA DELLA MISURA

3.1 Algebre e misure

In questo capitolo introduciamo alcune nozioni fondamentali di teoria della misura che saranno necessarie nel seguente capitolo per poter comprendere le medie condizionate e le martingale.

Definizione 3.1.1 : *Sia Ω un insieme e $\mathcal{A}$ una famiglia di insiemi contenuti in esso. Allora $\mathcal{A}$ è un'* **algebra** *se verifica*

(a) per ogni $A \in \mathcal{A}$ anche $A^c \in \mathcal{A}$;

(b) se $A, B \in \mathcal{A}$ allora $A \cup B \in \mathcal{A}$.

Chiaramente una $\sigma-$algebra è anche un'algebra.

Esempi 3.1.1 :

1. *Sia $\Omega = \mathbb{Z}$, verificare che la famiglia*

$$\mathcal{A} = \{A \subseteq \mathbb{Z} : \ A \, o \, A^c \text{ è finito}\}$$

è un'algebra.

2. *Sia $\Omega = \mathbb{R}$, verificare che*

$$\mathcal{A} = \{\bigcup_{i=1}^{n} (a_i, b_i] \ : \ n \in \mathbb{N}^*, \ a_i < b_i \in \mathbb{R}\}$$

è un'algebra.

3. *Se* $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \mathcal{A}_3 \subseteq \dots$ *è una successione crescente di* $\sigma-$*algebre, allora* $\mathcal{A} = \bigcup_{i=1}^{+\infty} \mathcal{A}_i$ *è un'algebra, ma potrebbe non essere una* $\sigma-$*algebra. Perchè ?*

Definizione 3.1.2 *Sia* $\mathcal{A}$ *un'algebra su* Ω*, una funzione* $\mu : \mathcal{A} \longrightarrow \mathbb{R}$ *è detta una* **misura** *su* $\mathcal{A}$ *se*

1. $\mu(A) \geq \mu(\emptyset) = 0$*, per ogni* $A \in \mathcal{A}$*;*
2. *se* $\{A_i\}_i \subseteq \mathcal{A}$ *e* $A_i \cap A_j = \emptyset$ *per ogni* $i \neq j$*, allora*

$$\mu(\bigcup_{i=1}^{+\infty} A_i) = \sum_{i=1}^{+\infty} \mu(A_i).$$

La terna $(\Omega, \mathcal{A}, \mu)$ *è chiamata spazio di misura.*

Una misura è detta $\sigma-$**finita** *se esiste una successione di insiemi* $\{A_n\}_n \subset \mathcal{A}$ *tale che per ogni* n $\mu(A_n) < +\infty$ *e* $\bigcup_{n=1}^{+\infty} A_n = \Omega$*.*

La seguente proposizione fornisce le proprietà della misura.

Proposizione 3.1.1 : *Sia* $(\Omega, \mathcal{A}, \mu)$ *uno spazio di misura. Allora*

1. *per ogni* $A, B \in \mathcal{A}$ *tali che* $A \subseteq B$ *si ha* $\mu(A) \leq \mu(B)$*;*
2. *per ogni successione* $A_1 \subseteq A_2 \subseteq \dots$ *(* $A_1 \supseteq A_2 \supseteq \dots$*) di insiemi in* $\mathcal{A}$

$$\lim_{i \to +\infty} \mu(A_i) = \mu(\bigcup_{i=1}^{+\infty} A_i), \qquad (\lim_{i \to +\infty} \mu(A_i) = \mu(\bigcap_{i=1}^{+\infty} A_i)$$

3. *(subadditività) per ogni* $\{A_n\}_n \subseteq \mathcal{A}$*,* $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$ *e*

$$\mu(\bigcup_{i=1}^{+\infty} A_i) \leq \sum_{i=1}^{+\infty} \mu(A_i)$$

Dimostrazione: poichè $A \subseteq B$ si ha $B = A \cup (B \cap A^c)$, unione disgiunta. Dalla seconda proprietà della misura

$$\mu(B) = \mu(B) + \mu(B \cap A^c)$$

da cui la prima affermazione è dimostrata.

Per la seconda, costruiamo la seguente successione di insiemi in $\mathcal{A}$

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \dots, B_k = A_k \setminus A_{k-1}, \dots,$$

Questi sono insiemi disgiunti, per inclusione $A_k = B_1 \cup \cdots \cup B_k$ e $\bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i$. Quindi

$$\begin{aligned}\lim_{k\to+\infty} \mu(A_k) &= \lim_{k\to+\infty} \mu(B_1 \cup \cdots \cup B_k) = \lim_{k\to+\infty} \sum_{n=1}^{k} \mu(B_n) \\ &= \sum_{n=1}^{+\infty} \mu(B_n) = \mu(\bigcup_{n=1}^{+\infty} B_n) = \mu(\bigcup_{n=1}^{+\infty} A_n).\end{aligned}$$

Infine per dimostrare la terza affermazione, notiamo la famiglia di insiemi così definita

$$B_1 = A_1, \dots, B_k = A_k \cap A_{k-1}^c \cap \cdots \cap A_1^c$$

è disgiunta e $A_1 \cup \cdots \cup A_n = B_1 \cup \cdots \cup B_n$, dunque per le due precedenti parti si ha

$$\begin{aligned}\mu(\bigcup_{i=1}^{+\infty} A_i) &= \lim_{n\to+\infty} \mu(\bigcup_{i=1}^{n} A_i) = \lim_{n\to+\infty} \mu(\bigcup_{i=1}^{n} B_i) \\ &= \lim_{n\to+\infty} \sum_{i=1}^{n} \mu(B_i) \leq \lim_{n\to+\infty} \sum_{i=1}^{n} \mu(A_i) = \sum_{i=1}^{+\infty} \mu(A_i)\end{aligned}$$

completando la dimostrazione. □

Data un'algebra $\mathcal{A}$, denotiamo con $\sigma(\mathcal{A})$ la più piccola $\sigma-$algebra contenente $\mathcal{A}$, ovvero

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{F}\supset\mathcal{A}} \mathcal{F}, \quad \mathcal{F}\,\sigma - \text{algebra}$$

ed è detta la $\sigma-$algebra generata da $\mathcal{A}$.

Il seguente teorema ci dice che per assegnare una misura su una $\sigma-$algebra è sufficiente assegnarla sull'algebra che la genera

Teorema 3.1.1 : *Teorema di Caratheodory*

Sia μ una misura $\sigma-$finita su un'algebra $\mathcal{A}$ di insiemi di Ω, allora μ ha un'unica estensione su $\sigma(\mathcal{A})$.

Per una dimostrazione vedere [B].

In particolare il teorema di Caratheodory si può generalizzare considerando una semialgebra anzichè un'algebra.

Definizione 3.1.3 *Una collezione di insiemi su* Ω, $\mathcal{S}$, *è detta una semialgebra se*

(a) per ogni $S, T \in \mathcal{S}$ *allora* $S \cap T \in \mathcal{S}$*;*

(b) per ogni $S \in \mathcal{S}$, *allora* $S^c = \bigcup_{i=1}^n T_i$ *per qualche* n *e* $T_i \in \mathcal{S}$ *disgiunti.*

Esempio 3.1.1 : *in* $(0, 1]$ *la famiglia*

$$\mathcal{R} = \{(a, b] : a < b \in \mathbb{R}\}$$

costituisce una semialgebra (verificare). Ogni insieme aperto può essere scritto in termini di intervalli aperti ed ogni intervallo aperto può essere scritto in termini di elementi di $\mathcal{R}$, *la* $\sigma-$*algebra dei Boreliani su* $\mathbb{R}$ *è generata da* $\mathcal{R}$. *Dunque, per il teorema di Caratheodory, ogni volta che vogliamo definire una misura su* $\mathcal{B}(\mathbb{R})$ *allora è sufficiente assegnarla agli insiemi di* $\mathcal{R}$.

La misura di Lebesgue su $\mathbb{R}$ *è quella che pone*

$$\mu((a, b]) = b - a.$$

Analogamente su $\mathbb{R}$ *la seguente famiglia è una semialgebra (verificare)*

$$\mathcal{R} = \{(-\infty, x] : x \in \mathbb{R}\}.$$

3.2 Integrazione

In questo paragrafo vogliamo introdurre le notazioni e le proprietà principali della teoria dell'integrazione astratta rispetto ad una qualsiasi misura $\sigma-$finita.

Consideriamo uno spazio misurabile $(\Omega, \mathcal{F})$, con $\mathcal{F}$ $\sigma-$algebra e sia μ una misura $\sigma-$finita definita su di esso.

Definizione 3.2.1 : *un'applicazione* $\phi : \Omega \longrightarrow \mathbb{R}$ *è detta una* **funzione semplice** *se esistono* $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{R}$ *e* $A_1, \ldots, A_n \in \mathcal{F}$ *disgiunti tali che*

$$\phi(\omega) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(\omega), \qquad \omega \in \Omega$$

con $\mu(A_i) < +\infty$ *per ogni* $i = 1, \ldots, n$. *Indichiamo con* $\mathcal{S}$ *l'insieme delle funzioni semplici.*

Le funzioni semplici sono, per costruzione, funzioni misurabili rispetto a $\mathcal{F}$, poichè per ogni $B \in \mathcal{B}(\mathbb{R})$

$$\phi^{-1}(B) = \bigcup_{j=0}^{k} A_{i_j} \in \mathcal{F}$$

per qualche $k \leq n$ con la convenzione che $A_0 = \emptyset$.

Per funzioni semplici definiamo l'integrale rispetto alla misura μ come

$$\int_\Omega \phi d\mu = \sum_{i=1}^{n} a_i \mu(A_i) \tag{3.1}$$

In quel che segue, nella notazione di integrale spesso ometteremo Ω a pedice.

Nel caso in cui $\Omega = \mathbb{R}$ e μ sia esattamente la misura di Lebesgue sulla retta reale, allora questa nozione coincide con quella di somme di Riemann, quando si considerano gli insiemi A_i come intervalli.

D'ora in poi diremo che una proprietà $\mathcal{I}$ è verificata μ quasi ovunque (q.o.) se $\mu(\{\omega : \mathcal{I}$ non è vera $\}) = 0$. Per esempio, prese ϕ e ψ misurabili

$$\psi \leq \phi \qquad q.o. \text{ se} \quad \mu(\{\omega : \psi(\omega) > \phi(\omega)\}) = 0.$$

Nel seguente teorema elenchiamo le principali proprietà dell'integrale di funzioni semplici.

Teorema 3.2.1 : *Per ogni* $\phi, \psi \in \mathcal{S}$, $a, b \in \mathbb{R}$ *valgono*

1. *se* $\phi \geq 0$ *q.o., allora* $\displaystyle\int \phi d\mu \geq 0$ *(monotonia);*

2. $a\phi + b\psi \in \mathcal{S}$ e $\int (a\phi + b\psi) d\mu = a \int \phi d\mu + b \int \psi d\mu$ *(linearità);*

3. *se* $\psi = \phi$ *q.o., allora*
$$\int \psi d\mu = \int \phi d\mu;$$

4. $|\phi| \in \mathcal{S}$ *e*
$$\Big| \int \phi d\mu \Big| \leq \int |\phi| d\mu.$$

Dimostrazione: dimostriamo solamente la terza proprietà poichè le altre si dimostrano in maniera analoga. Siano

$$\phi = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}, \quad \psi = \sum j = 1^m b_j \mathbf{1}_{B_j}$$

con $A_1, \ldots, A_n, B_1, \ldots, B_m \in \mathcal{F}$, $a_1, \ldots, a_n, b_1, \ldots, b_m \in \mathbb{R}$.

Consideriamo $C_{ij} = A_i \cap B_j$ e definiamo

$$A_0 = \bigcup_j B_j \setminus \bigcup_i A_i, \quad B_0 = \bigcup_i A_i \setminus \bigcup_j B_j$$

allora abbiamo

$$a\phi + b\psi = \sum_{i=0}^{n} \sum_{j=0}^{m} (aa_i + bb_j) \mathbf{1}_{C_{ij}}, \qquad a_0 = b_0 = 0$$

gli insiemi C_{ij} sono disgiunti a due a due, quindi si ottiene

$$\begin{aligned} \int (a\phi + b\psi) d\mu &= \sum_{i=0}^{n} \sum_{j=0}^{m} (aa_i + bb_j) \mu(C_{ij}) = \sum_{i=0}^{n} \sum_{j=0}^{m} aa_i \mu(C_{ij}) + \sum_{i=0}^{n} \sum_{j=0}^{m} bb_j \mu(C_{ij}) \\ &= a \sum_{i=0}^{n} a_i \mu(A_i) + b \sum_{j=0}^{m} b_j \mu(B_j) = \int \phi d\mu + \int \psi d\mu \end{aligned}$$

che conclude la dimostrazione. □

Vogliamo generalizzare la definizione di integrale ad una classe di funzioni più ampia. Consideriamo una funzione $f : \Omega \longrightarrow \mathbb{R}$ limitata e misurabile e tale che esista un insieme $C \subseteq \Omega$ di misura finita e con $f = 0$ su C^c. Se $|f| \leq M$ per qualche costante $M > 0$, denotiamo per $n \in \mathbb{N}$

$$\begin{aligned} C_k^n &= \{\omega \in C : (k-1)\frac{M}{n} < f(\omega) \leq k\frac{M}{n}\}, \qquad -n \leq k \leq n \\ \psi_n(\omega) &= \sum_{k=-n}^{n} k\frac{M}{n} \mathbf{1}_{C_k^n}(\omega), \qquad \phi_n(\omega) = \sum_{k=-n}^{n} (k-1)\frac{M}{n} \mathbf{1}_{C_k^n}(\omega) \end{aligned}$$

notiamo che, per costruzione, gli insiemi C_k^n sono tutti misurabili, ψ e ϕ sono funzioni semplici e

$$
\begin{aligned}
&\phi_n(\omega) \leq f(\omega) \leq \psi(\omega), \qquad \forall\, \omega \in \Omega,\, n \in \mathbb{N} \\
&\phi_n(\omega) = 0 = \psi_n(\omega), \qquad \forall\, \omega \in C^c \\
&\psi_n(\omega) - \phi_n(\omega) = \frac{M}{n}\mathbf{1}_C \Rightarrow \int (\psi_n - \phi_n)d\mu = \frac{M}{n}\mu(C) \longrightarrow 0 \quad \text{per}\, n \to +\infty
\end{aligned}
$$

poichè la misura di C è finita. Dunque per ogni $n \in \mathbb{N}$

$$\inf_{\psi \geq f} \int \psi d\mu - \frac{M}{n}\mu(C) \leq \int \psi_n d\mu - \frac{M}{n}\mu(C) = \int \phi_n d\mu \leq \sup_{\phi \leq f} \int \phi d\mu$$

dove ψ e ϕ sono funzioni semplici. Siccome vale naturalmente la disuguaglianza contraria, si ottiene

$$\sup_{\phi \leq f} \int \phi d\mu = \inf_{\psi \geq f} \int \psi d\mu$$

e questo numero verrà indicato con $\int f d\mu$, definendo quindi l'integrale per tutte le funzioni limitate e misurabili con supporto finito.

Abbiamo costruito l'integrale tramite approssimazione di funzioni semplici ed esattamente nella stessa maniera possiamo dimostrare che continuano a valere le stesse proprietà

Proposizione 3.2.1 : *Sia $C \in \mathcal{F}$ tale che $\mu(C) < +\infty$ e siano f, g due funzioni a valori reali limitate e misurabili tali che $f = g = 0$ su C^c. Allora*

1. *se $f \geq 0$ q.o., allora $\int f d\mu \geq 0$;*

2. *per ogni $a, b \in \mathbb{R}$, $\int (af + bg)d\mu = a \int f d\mu + b \int g d\mu$.*

Poichè se abbiamo una funzione misurabile e limitata f anche $f\nVdash_C$ lo è allora si può definire $\int f\mathbf{1}_C d\mu$ e per approssimazione si può dimostrare

$$\int_C f d\mu = \int f\mathbf{1}_C d\mu$$

La definizione di integrale può essere estesa ad una classe ancora più ampia di funzioni. Per $f : \Omega \longrightarrow \mathbb{R}$ misurabile non negativa, definiamo

$$\int f d\mu = \sup\{\int h d\mu\,,\, 0 \leq h \leq f,\, h\, \text{limitata misurabile con}\; \mu(\{h(x) > 0\}) < +\infty\}$$

abbiamo dunque il seguente

Proposizione 3.2.2 : *sia* $\{C_n\}_n$ *una successione crescente verso* Ω *di insiemi in* $\mathcal{F}$ *di misura finita, allora per* $n \to +\infty$

$$\int_{C_n} (f \wedge n) d\mu \uparrow \int f d\mu.$$

La precedente proposizione fornisce il metodo per calcolare l'integrale. Le stesse proprietà di monotonia e linearità valgono anche per questa estensione.

Infine diciamo che una funzione misurabile è **integrabile** se

$$\int |f| d\mu < +\infty.$$

Allora se definiamo la parte positiva e la parte negativa di f

$$f^+(\omega) = \max(f(\omega), 0), \qquad f^-(\omega) = \max(-f(\omega), 0),$$

queste sono funzioni misurabili non negative e per le quali l'integrale è definito e

$$f(\omega) = f^+(\omega) - f^-(\omega), \qquad |f(\omega)| = f^+(\omega) + f^-(\omega)$$

e si può quindi definire l'integrale

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

per il quale valgono le solite proprietà di monotonia e linearità.

Esempi 3.2.1 : *prendiamo come spazio misurabile* $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R})$

(a) Se prendiamo μ *uguale alla misura di Lebesgue e come classe di funzioni integrabili, le funzioni continue su un compatto, allora la precedente definizione di integrale coincide con quella di integrale di Riemann.*

(b) sia $F : \mathbb{R} \longrightarrow \mathbb{R}$ *una funzione non decrescente limitata continua a destra con limiti a sinistra e prendiamo* μ *come la misura indotta da* F*, ovvero*

$$\begin{aligned} &\mu((a,b]) = F(b) - F(a), \qquad \forall\, a, b \in mathbbR \quad \textit{di continuità per } F \\ &\mu(a) = F(a) - F(a-) \qquad \forall\, a \in \mathbb{R} \quad \textit{di discontinuità di } F. \end{aligned}$$

Se prendiamo come classe delle funzioni integrabili le funzioni continue e limitate su $\mathbb{R}$*, allora la precedente definizione coincide con la definizione dell'integrale di Riemann - Stiltijes*

$$\int f(x) dF(x) = \sup_{\pi_n} \sum_{x_i \in \pi_n} f(x_i)(F(x_i) - F(x_{i-1}))$$

dove il sup è preso su tutte le possibili partizioni crescenti di $\mathbb{R}$.

(c) *Se nel precedente punto scegliamo F differenziabile, per il teorema del valor medio $(F(x_i) - F(x_{i-1})) = F'(\xi_i)(x_i - x_{i-1})$, dove $x_{i-1} \leq \xi_i \leq x_i$., allora μ è una misura assolutamente continua (vedremo in seguito la definizione precisa) rispetto alla misura di Lebesgue e l'integrale diventa*

$$\begin{aligned}\int f(x)dF(x) &= \sup_{\pi_n} \sum_{x_i \in \pi_n} f(x_i)(F(x_i) - F(x_{i-1})) \\ &= \sup_{\pi_n} \sum_{x_i \in \pi_n} f(x_i)F'(\xi_i)(x_i - x_{i-1}) = \int f(x)F'(x)dx\end{aligned}$$

Esempi di questo genere sono dati dalle distribuzioni Gaussiane, esponenziali etc.

(d) *Se consideriamo un qualsiasi spazio di probabilità $(\Omega, \mathcal{F}, P)$, la precedente è la definizione corretta di media di una v.a. X*

$$E(X) = \int X dP$$

Ulteriori proprietà dell'integrale

1. Se $f = g$ $\mu-$ q.o., allora $\int f d\mu = \int g d\mu$.

2. **Disuguaglianza di Jensen**: sia $\mu(\Omega) = 1$ e siano $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ una funzione convessa, $f : \Omega \longrightarrow \mathbb{R}$ una funzione misurabile tali che f e $\phi(f)$ sono integrabili, allora

$$\phi\left(\int f d\mu\right) \leq \int \phi(f) d\mu. \tag{3.2}$$

Dimostrazione: Indichiamo con $c = \int f d\mu$ e con $l(x)$ la funzione lineare $ax + b$ scelta in maniera tale che $l(c) = \phi(c)$ e $l(x) \leq \phi(x)$ per ogni $x \neq c$. Questa funzione esiste grazie alla convessità di ϕ. Allora

$$\begin{aligned}\int \phi(f)d\mu &\geq \int l(f)d\mu = \int (af + b)d\mu = a\int f d\mu + b\mu(\Omega) \\ &= ac + b = l(c) = \phi(c) = \phi\left(\int f d\mu\right) \qquad \square\end{aligned}$$

La disuguaglianza di Jensen vale al contrario per funzioni concave.

Come conseguenza della disuguaglianza di Jensen abbiamo che

$$(\int |f| d\mu)^p \leq \int |f|^p d\mu, \qquad \forall\, p \geq 1$$

In generale, sullo spazio di tutte le funzioni per le quali è definito l'integrale rispetto a μ, definiamo i seguenti sottospazi per $p \geq 1$

$$\begin{aligned} L^p(\mu) &= \{f : ||f||_p = (\int |f|^p d\mu\Big)^{\frac{1}{p}} < +\infty\} \\ L^\infty(\mu) &= \{f : ||f||_\infty = \text{essup}|f| < +\infty\} \\ \text{essup} f &= \inf\{y > 0 : \mu(\{|f| > y\}) = 0\} \end{aligned}$$

Questi spazi sono in realtà classi di equivalenza rispetto al coincidere $\mu-$q.o. delle funzioni. Inoltre $||\cdot||_p$, $p \in [1, \infty]$ costituisce una norma su L^p (verificare). Rispetto a queste norme, sono valide le seguenti importanti disuguaglianze

3. **Disuguaglianza di Hölder**: siano $p, q \in (1, +\infty)$ tali che $\dfrac{1}{p} + \dfrac{1}{q} = 1$ (coniugati), allora

$$\int |fg| d\mu \leq ||f||_p ||g||_q, \qquad f \in L^p(\mu),\ g \in L^q(\mu) \tag{3.3}$$

$$\int |fg| d\mu \leq ||f||_\infty ||g||_1, \qquad f \in L^\infty(\mu),\ g \in L^1(\mu). \tag{3.4}$$

Dimostrazione: La seconda disuguaglianza è immediata poichè $|f| \leq ||f||_\infty$ $\mu-$q.o. Per esponenti coniugati abbiamo che se $||f||_p = 0$ o $||g||_q = 0$, la disuguaglianza è trivialmente vera. Imponiamo dunque $||f||_p > 0, ||g||_q > 0$, senza perdere di generalità si può supporre che $||f||_p = ||g||_q = 1$, altrimenti è sufficiente considerare le funzioni normalizzate $\tilde{f} = \dfrac{f}{||f||_p}$, $\tilde{g} = \dfrac{g}{||g||_q}$ e svolgere la dimostrazione per esse.

Per $y > 0$ fissato, consideriamo la funzione

$$\phi(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy, \qquad x \geq 0,$$

allora $\phi'(x) = x^{p-1} - y$ e $\phi''(x) = (p-1)x^{p-2} \geq 0$, da cui si ha un punto di minimo per $x_0 = y^{\frac{1}{p-1}}$ e $x_0^p = y^{\frac{p}{p-1}} = y^q$, con minimo

$$\phi(x_0) = \frac{y^q}{p} + \frac{y^q}{q} - y^{\frac{1}{p-1}-1} = y^q - y^q = 0.$$

Perciò $\phi(x) \geq 0$ per ogni $x \geq 0, y \geq 0$, se scegliamo $x = |f|$ e $y = |g|$ e la dimostrazione è completata per monotonia e linearità dell'integrale. $\square$

Per $p = q = 2$ (che sono coniugati) si ottiene la disuguaglianza di Cauchy-Schwarz.

4. **Disuguaglianza di Minkowski**: siano $f, g \in L^p(\mu)$, $p \in [1 + \infty]$ allora

$$||f + g||_p \leq ||f||_p + ||g||_p \tag{3.5}$$

Dimostrazione: Per $p \geq 1$ basta notare che

$$|f + g|^p \leq 2^p(|f|^p + |g|^p)$$

da cui $f + g \in L^p(\mu)$ e quindi applicare la disuguaglianza di Hölder alle funzioni $|f|\,|f+g|^{p-1}$ e $|g|\,|f+g|^{p-1}$. Per funzioni in $L^\infty(\mu)$ la disuguaglianza segue dalla definizione di essup. □

Osservazione 3.2.1 : *Nel caso in cui $\mu(\Omega) < +\infty$ gli spazi $L^p(\mu)$ rispettano la relazione di inclusione*

$$L^\infty(\mu) \subseteq L^q(\mu) \subseteq L^p(\mu), \qquad \forall q \geq p \geq 1,$$

infatti dalla disuguaglianza di Jensen abbiamo per $\frac{q}{p} > 1$

$$\begin{aligned} \int |f|^q d\mu &= \int \mu(\Omega)|f|^{\frac{q}{p}p} \frac{d\mu}{\mu(\Omega)} = \int (\mu(\Omega)^{\frac{p}{q}}|f|^p)^{\frac{q}{p}} \frac{d\mu}{\mu(\Omega)} \\ &\geq \Big(\int (\mu(\Omega)^{\frac{p}{q}}|f|^p) \frac{d\mu}{\mu(\Omega)} \Big)^{\frac{q}{p}} = \mu(\Omega) \frac{1}{\mu(\Omega)^{\frac{q}{p}}} \Big(\int |f|^p d\mu \Big)^{\frac{q}{p}} \end{aligned}$$

e prendendo la potenza $\frac{1}{q}$ di entrambi i lati della disuguaglianza si ottiene

$$||f||_p \leq \mu(\Omega)^{\frac{1}{q}-\frac{1}{p}}||f||_q$$

ovvero se $f \in L^q(\mu)$ allora $f \in L^p(\mu)$. Per il caso infinito si ha

$$\int |f|^p d\mu \leq ||f||_\infty \mu(\Omega) < +\infty.$$

3.3 Passaggio al limite sotto il segno di integrale

In quest'ultima sezione presentiamo i principali teoremi che consentono il passaggio al limite nella teoria generale della misura.

Teorema 3.3.1 : **Teorema della convergenza monotona**

Sia $\{f_n\}$ una successione crescente di funzioni misurabili, integrabili rispetto a μ non negative convergente puntualmente ad una funzione $f \in L^1(\mu)$ misurabile, ovvero

$$f_n(\omega) \leq f_{n+1}(\omega) \leq \ldots \uparrow f(\omega), \qquad \forall\, \omega \in \Omega,$$

allora

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

Dimostrazione: Per monotonia dell'integrale

$$0 \leq \int f_n d\mu \leq \int f_{n+1} d\mu \leq \cdots \leq \int f d\mu < +\infty$$

questa è una successione reale monotona e limitata, da cui deve esistere

$$\lim_{n\to+\infty} \int f_n d\mu \leq \int f d\mu.$$

Per dimostrare l'uguaglianza è sufficiente mostrare che per ogni partizione finita di Ω, $\{A_i\}$, vale

$$\sum_i \inf_{\omega\in A_i} f(\omega)\mu(A_i) \leq \lim_{n\to+\infty} \int f_n d\mu.$$

Poniamo $\inf_{\omega\in A_i} f(\omega) = \nu_i$ e si scelgano numeri reali ϵ, $u_i < \nu_i$ e $u_i = 0$ se $\nu_i = 0$, tali che

$$\epsilon < \sum_i u_i \mu(A_i).$$

Se $\omega \in A_i$ per la convergenza puntuale delle funzioni si ha $f_n(\omega) \uparrow f(\omega) \geq \nu_i$, da cui $f_n(x) \geq u_i$ per n sufficientemente grande.

Se denotiamo $A_{i_n} = \{\omega \in A_i \,:\, f_n(\omega) \geq u_i\}$, per quanto detto prima $A_{i_n} \uparrow A_i$ per $n \to +\infty$ e per le proprietà della misura

$$\mu(A_{i_n}) \uparrow \mu(A_i), \qquad n \to +\infty.$$

Abbiamo dunque

$$\begin{aligned}\int f_n d\mu &\geq \sum_i \inf_{\omega\in A_{i_n}} f_n(\omega)\mu(A_{i_n}) + \sum_i \inf_{\omega\in A_i\setminus A_{i_n}} f_n(\omega)\mu(A_i \setminus A_{i_n}) \\ &\geq \sum u_i \mu(A_{i_n}) \longrightarrow \sum_i u_i \mu(A_i) > \epsilon\end{aligned}$$

e per l'arbitrarietà degli u_i, passando al limite si ha la tesi. □

Lemma 3.3.1 : **Lemma di Fatou**

Sia $\{f_n\}$ *una successione di funzioni misurabili, integrabili rispetto a* μ *non negative, allora*

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

Dimostrazione: definiamo $g_n = \inf_{k \geq n} f_k$, allora

$$f_n \geq g_n, \quad \forall\, n \quad \text{e} \quad 0 \leq g_n \uparrow \liminf_n f_n \quad \text{per} \quad n \to +\infty.$$

Applicando il teorema della convergenza monotona si ottiene

$$\int f_n d\mu \geq \int g_n d\mu \longrightarrow \int \liminf_n f_n d\mu \quad \text{per} \quad n \to +\infty \qquad \square$$

Osserviamo che il lemma di Fatou si può anche enunciare per il limsup ottenendo la disuguaglianza contraria.

Teorema 3.3.2 : **Teorema della convergenza dominata**

Sia $\{f_n\}$ *una successione di funzioni misurabili tale che* $|f_n| \leq g$ *per ogni* n *per qualche funzione* g *misurabile ed integrabile rispetto a* μ *ed inoltre si supponga che esiste* f *misurabile tale che* $f_n(\omega) \to f(\omega)$ *per quasi ogni* $\omega \in \Omega$. *Allora* f_n, f *sono integrabili per ogni* n *e*

$$\int f_n d\mu \longrightarrow \int f d\mu \quad \textit{per} \quad n \to +\infty.$$

Dimostrazione: Poichè per ogni n $|f_n| \leq g$ necessariamente $|f| \leq g$ e quindi per la monotonia dell'integrale tutte le funzioni risultano integrabili.

Poniamo $h_n = |f_n - f|$ ed $h = 2g$. Per ipotesi $h_n \longrightarrow 0$ e $0 \leq h_n \leq h$, quindi applicando il lemma di Fatou si ha

$$\int h d\mu = \int \liminf_n (h - h_n) d\mu \leq \liminf_n \int (h - h_n) d\mu = \int h d\mu - \limsup_n \int h_n d\mu$$

ma $\limsup_n \int h_n d\mu \longrightarrow 0$ sempre per il Lemma di Fatou applicato con il limsup. $\square$

Completiamo questa sezione con una definizione che utilizzeremo spesso nel prossimo capitolo.

Definizione 3.3.1 : *Sia dato uno spazio misurabile* $(\Omega, \mathcal{F})$ *e due misure* μ_1, μ_2 *su esso definite.*

(a) μ_1, μ_2 *si dicono* **mutualmente singolari**, *e si scrive* $\mu_1 \perp \mu_2$, *se esiste un insieme* $A \in \mathcal{F}$ *tale che*

$$\mu_1(A) = 0, \qquad \mu_2(A^c) = 0$$

(b) *Si dice che* μ_1 *è* **assolutamente continua rispetto a** μ_2, *e si scrive* $\mu_1 << \mu_2$ *se per ogni* $A \in \mathcal{F}$ *tale che* $\mu_2(A) = 0$ *anche* $\mu_1(A) = 0$.

(c) *se* $\mu_1 << \mu_2$ *e* $\mu_2 << \mu_1$, *le due misure si dicono* **equivalenti**

Capitolo 4

MEDIE CONDIZIONATE

4.1 Preliminari

In questa sezione presntiamo alcuni teoremi fondamentali in teoria della misura che occorreranno nelle definizioni dei prossimi capitoli. Iniziamo dando una definizione di essup alternativa a quella introdotta nel precedente capitolo, ma che permette di comprendere che l'estremo superiore essenziale è una generalizzazione del concetto di estremo superiore.

Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio di misura finito e sia $\{f_\alpha\}_{\alpha \in \mathcal{A}}$ una famiglia di funzioni misurabili non negative.

Definizione 4.1.1 : *Se esiste una funzione g tale che*

(i) $f_\alpha \leq g$ *q.o. per ogni* $\alpha \in \mathcal{A}$*;*

(ii) se $f_\alpha \leq h$ *q.o. per ogni* $\alpha \in \mathcal{A}$ *allora* $g \leq h$ *q.o.,*

allora si definisce $g = essup f_\alpha$*, l'estremo superiore essenziale della famiglia* $\{f_\alpha\}_{\alpha \in \mathcal{A}}$.

L'estremo superiore essenziale di una famiglia è unico a meno di insiemi di misura nulla.

Esempio 4.1.1 : *In generale l'estremo superiore è differente dall'estremo superiore essenziale. Come esempio, consideriamo la famiglia di funzioni misurabili sullo spazio* $([0,1], \mathcal{B}([0,1]), Lebesgue)$*, descritta da* $f_t = \mathbf{1}_{\{t\}}$*, si ha che* $g = essup_{t \in [0,1]} f_t = 0$*, ma* $\sup_{t \in [0,1]} f_t = 1$.

Se la famiglia è costituita da funzioni continue oppure se è numerabile, allora l'estremo superiore e l'estremo superiore essenziale coincidono.

Vale il seguente teorema di esistenza per l'estremo superiore essenziale

Teorema 4.1.1 : *Se si ammette che l'estremo superiore essenziale di una famiglia* $\{f_\alpha\}_{\alpha\in\mathcal{A}}$*, dove* $f_\alpha \geq 0$ *q.o. per ogni* $\alpha in \mathcal{A}$*, sia una funzione a valori nella retta reale estesa, allora esso esiste sempre.*

Dimostrazione: **1° caso**: La famiglia $\{f_\alpha\}_{\alpha\in\mathcal{A}}$ è limitata.

In questo caso, esiste una costante M tale che $0 \leq f_\alpha \leq M$, $\forall\, \alpha \in \mathcal{A}$ e poniamo

$$\lambda = \sup_n \int_\Omega \max_{\substack{1\leq i\leq n \\ \alpha_i\in\mathcal{A}}} f_{\alpha_i} d\mu.$$

Necessariamente si ha $\lambda \leq M\mu(\Omega)$.

Poichè λ è un estremo superiore, deve esistere una famiglia di funzioni $\{g_i\}_{i\in\mathbb{N}}$ tale che $g_1 \leq g_2 \leq \cdots \leq g_n$, dove $g_n = \max\limits_{1\leq i\leq n} f_{\alpha_i}$ per qualche $\{\alpha_i\}$ e per monotonia esiste

$$\int_\Omega g_n d\mu \longrightarrow \lambda \text{ per } n \to \infty.$$

Vogliamo dimostrare che $g = \lim\limits_{n\to\infty} g_n$ verifica le due proprietà dell'estremo superiore essenziale.

(i) Supponiamo per assurdo che esista un indice α_0 tale che $f_{\alpha_0} > g$ su un qualche insieme H di misura positiva. Senza perdere di generalità possiamo supporre che $H = \Omega$, infatti se così non fosse potremmo considerare la famiglia di funzioni $\{f_\alpha|_H\}$ e ripetere la dimostrazione che segue. Allora per ogni $\max\limits_{\alpha_0,\ldots,\alpha_n} f_{\alpha_i}$ contenente f_{α_0}, si ha anche $\max\limits_{\alpha_0,\ldots,\alpha_n} f_{\alpha_i} > g$ q.o..

D'altra parte se prendiamo la successione dei g_n convergente a g, si ha anche che

$$g_n \vee f_{\alpha_0} \to g \vee f_{\alpha_0}, \quad \text{ per } n \to +\infty$$

(con $x \vee y$ denotiamo $\max(x, y)$).

Per costruzione, $\int_\Omega (g_n \vee f_{\alpha_0}) d\mu \leq \lambda$ per ogni $n \in \mathbb{N}$ e dunque utilizzando il teorema della convergenza monotona si ha anche $\int_\Omega (g \vee f_{\alpha_0}) d\mu \leq \lambda$, mentre dalla nostra ipotesi si ha $\int_\Omega (g \vee f_{\alpha_0}) d\mu = \int_\Omega f_{\alpha_0} d\mu > \lambda$, ovvero una contraddizione.

(ii) Per assurdo, sia h una funzione tale che $f_\alpha \leq h$ per ogni $\alpha \in \mathcal{A}$ q.o. e $h < g$ q.o.

Per definizione di estremo superiore essenziale, deve esistere una f_{α_0} tale che $f_{\alpha_0} \geq h$ su un insieme di misura positiva, che è chiaramente una contraddizione.

2° caso: La famiglia $\{f_\alpha\}_{\alpha\in\mathcal{A}}$ è arbitraria.

In questo caso è sufficiente definire $h_\alpha = \arctan f_\alpha$. La nuova famiglia è limitata e dunque, grazie a quanto dimostrato prima, deve esistere $h = essuph_\alpha$. Grazie all'invertibilità della funzione arcotangente, si può definire $g = \tan h$, che verifica le proprietà dell'estremo superiore essenziale. $\square$

Utilizzeremo la nozione di essup per definire delle probabilità assolutamente continue rispetto a P.

Come esempio, se $\{\Omega, \mathcal{F}, P\}$ è uno spazio di probabilità completo (cioè ogni sottoinsieme di un insieme di misura nulla è misurabile ed ha misura nulla), se consideriamo un insieme $B \in \mathcal{F}$, tale che $P(B) > 0$ allora la probabilità condizionata

$$Q(A) = \frac{P(A \cap B)}{P(B)} \qquad \forall\, A \in \mathcal{F}$$

è una probabilità $Q << P$ sullo spazio misurabile $\{\Omega, \mathcal{F}\}$.

Introduciamo ora il teorema che è alla base della definizione di media condizionata.

Teorema 4.1.2 : *(Radon - Nikodym)*
Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio di misura finita e sia $\nu << \mu$ una misura finita sullo stesso spazio. Allora esiste una funzione f, $\mathcal{F}$-misurabile e tale che

$$\nu(A) = \int_A f d\mu \qquad \textit{per ogni} \quad A \in \mathcal{F} \tag{4.1}$$

e si scrive $f = \dfrac{d\nu}{d\mu}$.

Dimostrazione: Se $\nu(\Omega) = 0$, la conclusione è immediata (basta prendere $f \equiv 0$), sia quindi $\nu(\Omega) > 0$. Possiamo assumere senza perdere di generalità (a meno di normalizzazioni) che $\mu(\Omega) = 1$.

Dividiamo la dimostrazione in due passi.

1° passo: Sia ϵ tale che $\epsilon < \nu(\Omega)$, vogliamo dimostrare che esiste un $\Gamma \in \mathcal{F}$ tale che $\mu(\Gamma) > 0$ e $\nu(E) > \epsilon\mu(E)$ per ogni $E \subseteq \Gamma$ di misura positiva (μ o ν). Consideriamo la famiglia di insiemi

$$\mathcal{E} = \{\, E \in \mathcal{F} \,:\, \nu(E) \leq \epsilon\mu(E)\,\}.$$

Nessun insieme H tale che $\mu(H) = \mu(\Omega) = 1$ è in $\mathcal{E}$; infatti questo implica che $\mu(H^c) = 0$ e dunque per assoluta continuità anche $\nu(H^c) = 0$, perciò

$$\epsilon\mu(H) = \epsilon < \nu(\Omega) = \nu(H) + \nu(H^c) = \nu(H).$$

Se $\mathcal{E} = \emptyset$, possiamo scegliere $\Gamma = \Omega$ ed il primo passo è concluso.

Se invece $\mathcal{E}$ è non vuoto sia $(1 >)\gamma_1 = \sup\limits_{E \in \mathcal{E}} \mu(E)$, per definizione di estremo superiore deve esistere $E_1 \in \mathcal{E}$ tale $\mu(E_1) \geq \dfrac{1}{2}\gamma_1$.

Procedendo analogamente definiamo

$$\gamma_2 = \sup_{E \in \mathcal{E}} \{ \mu(E) \, : \, E \subseteq E_1^c \}.$$

Per costruzione $\gamma_2 \leq \gamma_1$ e, sempre per definizione di estremo superiore, esiste un insieme $E_2 \subseteq E_1^c$ tale che $\mu(E_2) \geq \frac{1}{2}\gamma_2$, ma $\nu(E_2) \leq \epsilon\mu(E_2)$.

Continuando nella stessa maniera, otteniamo una successione di insiemi disgiunti $\{E_i\}$ ed una di numeri reali $\{\gamma_i\}$ tali che

$$\gamma_n = \sup_{E \in \mathcal{E}} \Big\{ \mu(E) \, : \, E \subseteq \Big(\bigcup_{i=1}^{n-1} E_i \Big)^c \Big\} \longrightarrow 0.$$

La convergenza a zero è data dal fatto

$$1 > \mu(E_1 \cup \cdots \cup E_n) > \frac{1}{2}(\gamma_1 + \cdots + \gamma_n),$$

quindi la serie è convergente. Poniamo $F = \bigcup_{i \geq 1} E_i$, allora

$$\nu(F) = \sum_{i \geq 1} \nu(E_i) \leq \epsilon \sum_{i \geq 1} \mu(E_i) = \epsilon\mu(F),$$

dunque $F \in \mathcal{E}$ e necessariamente $\mu(F) < 1$, da cui $\Gamma = F^c$ deve avere misura positiva.

Inoltre, se $A \subseteq F^c$ e $\mu(A) > 0$, allora si ha $\nu(A) > \epsilon\mu(A)$. Infatti $A \subseteq \Gamma = \bigcap_{i \in \mathbb{N}} E_i^c$ implica che $A \subseteq E_i^c$ per ogni $i \in \mathbb{N}$, per cui se si avesse anche che $\nu(A) \leq \epsilon\mu(A)$ (ovvero $A \in \mathcal{E}$) ciò per costruzione implicherebbe

$$\mu(A) \leq \gamma_i \quad \text{per ogni } i \in \mathbb{N}$$

ovvero $\mu(A) = 0$ e conseguentemente (per assoluta continuità) $\nu(A) = 0$, che dà luogo ad una contraddizione.

Dimostrata l'esistenza dell'insieme Γ, poniamo $f(\omega) = \epsilon \mathbf{1}_\Gamma(\omega)$. Per costruzione $\mu(\Gamma) \geq \mu(E_1^c) > 1 - \gamma_1 > 0$, dunque

$$0 < \epsilon\mu(\Gamma) = \int_\Omega f d\mu,$$

e d'altra parte per quanto detto prima $\int_A f d\mu \leq \nu(A)$ per ogni $A \in \mathcal{F}$. Poniamo allora $f_\nu = \text{essup}\{g \, : \, \int_A g d\mu \leq \nu(A), \forall A \in \mathcal{F}\}$. Essendo g misurabile, l'insieme è non vuoto.

$\mathbf{2^o}$ passo: Vogliamo dimostrare che f_ν verifica la tesi del teorema.

Per costruzione, la funzione f è misurabile rispetto ad $\mathcal{F}$, non negativa ed integrabile. Inoltre non può che verificare l'uguaglianza (4.1).

Se, per qualche $A \in \mathcal{F}$, fosse $\displaystyle\int_A f_\nu > \nu(A)$, per la costruzione dell'estremo superiore essenziale, fatta nel lemma precedente, dovrebbe esistere una funzione $g_{n_0} = \max\{g_{\alpha_0}, \ldots, g_{\alpha_{n_0}}\}$, per qualche famiglia di indici, tale che

$$\int_A g_{n_0} d\mu > \nu(A),$$

che porta naturalmente ad una contraddizione.

Se invece $\displaystyle\int_A f_\nu d\mu < \nu(A)$ per qualche insieme A, allora si può definire

$$m(B) = \nu(A \cap B) - \int_{A\cap B} f_\nu d\mu \quad \forall\, B \in \mathcal{F}.$$

In particolare $m(\Omega) > 0$ e $m << \mu$, quindi per quanto sopra dimostrato esiste una funzione $\alpha \geq 0$ tale che $\displaystyle\int_\Omega \alpha d\mu > 0$ e

$$\int_B \alpha d\mu \leq m(B), \qquad \text{per ogni } \ B \in \mathcal{F}.$$

Inoltre si ha per ogni $B \in \mathcal{F}$

$$\begin{aligned} \int_B (f_\nu + \alpha) d\mu &= \nu(A \cap B) - \int_{A\cap B} f_\nu d\mu + \int_B f_\nu d\mu \\ &= \nu(A \cap B) + \int_{B \cap A^c} f\nu d\mu \leq \nu(A \cap B) + \nu(A^c \cap B) = \nu(B) \end{aligned}$$

e per la definizione di essup, si deve avere $f_\nu + \alpha \leq f_\nu$, una contraddizione poichè $\displaystyle\int_\Omega \alpha d\mu > 0$. □

Per quanto dimostrato, si può capire che una funzione di densità di probabilità può essere vista come la derivata di Radon-Nikodym di una probabilità assolutamente continua rispetto alla data.

Il teorema precedente ed il seguente si possono generalizzare alle misure $\sigma-$finite, che però qui non impiegheremo.

Teorema 4.1.3 : *(Lemma di decomposizione di Lebesgue)*

Siano μ e ν due misure finite su $(\Omega, \mathcal{F})$. Allora si può scrivere

$$\nu = \nu_a + \nu_s \quad \textit{dove} \quad \nu_a << \mu \quad \textit{e} \quad \nu_s \perp \mu$$

e la decomposizione è unica.

Dimostrazione: Se poniamo $a = \sup\{\nu(A), A \in \mathcal{F}, \mu(A) = 0\}$, per la finitezza delle misure, a è sicuramente finito (e quindi un massimo) e, per definizione di sup, deve esistere un insieme B tale che $\mu(B) = 0$ e $\nu(B) = a$. Si definiscano per ogni $F \in \mathcal{F}$

$$\nu_s(F) = \nu(F \cap B) \quad \text{e} \quad \nu_a(F) = (\nu - \nu_s)(F).$$

Allora si ha $\nu_s(\Omega) = \nu_s(B)$, dunque $\nu_s \perp \mu$ e $\nu_a << \mu$, infatti se per assurdo cosi' non fosse, si potrebbe trovare un insieme $H \subseteq B^c$, tale che $\nu_a(H) > 0$, ma $\mu(H) = 0$ e questo implicherebbe

$$\nu(A \cup B) = \nu(H) + \nu(B) > \nu(B) = a, \qquad \mu(H \cup B) = \mu(H) + \mu(B) = 0$$

che è chiaramente una contraddizione.

L'unicità si dimostra facilmente. Supponiamo che $\mu_s + \mu_a = \mu = \nu_s + \nu_a$ con μ_s e $\nu_s \perp \mu$ e $\mu_a, \nu_a << \mu$. Allora

$$\mu_s - \nu_s = \nu_a - \mu_a$$

e la prima misura è singolare rispetto a μ, mentre la seconda è assolutamente continua rispetto a μ, ma questo è possibile solo se entrambe sono la misura identicamente uguale a zero. $\square$

4.2 Definizioni e Proprietá

Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità completo, $X \geq 0$ una variabile aleatoria integrabile ed $\mathcal{A}$ una sotto $\sigma-$ algebra di $\mathcal{F}$. Su $\mathcal{A}$, definiamo la misura indotta da X come

$$\nu(A) = \int_A X dP \qquad \forall\ A \in \mathcal{A}.$$

La misura ν è finita, poichè $\nu(\Omega) = E(X) < +\infty$.

Per costruzione, $\nu << P|_{\mathcal{A}}$, quindi per il teorema di Radon-Nikodym, deve esistere una funzione $\mathcal{A}-$misurabile tale che

$$\nu(A) = \int_A f dP, \qquad \forall\ A \in \mathcal{A}$$

che necessariamente implica

$$\int_A f dP = \int_A X dP, \qquad \forall\ A \in \mathcal{A}.$$

Se la variabile aleatoria X è integrabile, ma a valori in $\mathbb{R}$, consideriamo le sue parti positiva X^+ e negativa X^-. Ognuna di queste v.a. induce una funzione $\mathcal{A}-$misurabile f^+ ed f^-, tali che

$$\int_A f^+ dP = \int_A X^+ dP, \qquad \int_A f^- dP = \int_A X^- dP, \qquad \forall\ A \in \mathcal{A}$$

e definiamo $f = f^+ - f^+$, che necessariamente è $\mathcal{A}-$misurabile e verifica

$$\int_A f dP = \int_A X dP, \qquad \forall\ A \in \mathcal{A}.$$

Definizione 4.2.1 : *Data una variabile aleatoria integrabile X ed una $\sigma-$algebra $\mathcal{A} \subseteq \mathcal{F}$, diciamo* **media condizionata di X data $\mathcal{A}$**, $E(X|\mathcal{A})$ *una qualsiasi variabile aleatoria $\mathcal{A}-$misurabile che verifichi*

$$\int_A E(X|\mathcal{A}) dP = \int_A X dP \qquad \forall\, A \in \mathcal{A} \tag{4.2}$$

Osservazione 4.2.1 : *La definizione di media condizionata ad una $\sigma-$algebra è data quasi ovunque, ovvero qualsiasi due funzioni f e g $\mathcal{A}-$misurabili, che verificano la (4.2), possono essere considerate come la media condizionata di X rispetto ad $\mathcal{A}$. D'altra parte, per quanto detto si ha*

$$\int_A (f-g) dP = 0 \qquad \forall\, A \in \mathcal{A}$$

e quindi anche sugli insiemi $\{\omega : f(\omega) > g(\omega)\}$ e $\{\omega : f(\omega) < g(\omega)\}$, che implica

$$\int_\Omega |f-g| dP = 0 \qquad \text{ovvero} \quad f = g \quad q.o.$$

Sulla base della definizione 4.2.1, possiamo introdurre anche la nozione di probabilità condizionata

Definizione 4.2.2 : *Per ogni $F \in \mathcal{F}$, definiamo la* **probabilità di F condizionata ad $\mathcal{A}$** *come*

$$P(F|A) = E(\mathbf{1}_F|\mathcal{A})$$

essendo $\mathbf{1}_F$ trivialmente integrabile.

La definizione di probabilità condizionata appena data naturalmente include la nozione di probabilità condizionata di un evento rispetto ad un altro. Infatti sia $B \in \mathcal{F}$ tale che $P(B) > 0$, allora la sotto $\sigma-$algebra generata da B è $\mathcal{A} = \{\Omega, \emptyset, B, B^c\}$ ed una qualsiasi funzione misurabile secondo $\mathcal{A}$ deve essere costante su B. Dunque $P(F|\mathcal{A})$ sarà quella variabile aleatoria tale che per ogni $\omega \in B$ è costante e si ha che

$$P(F \cap B) = \int_B \mathbf{1}_F dP = \int_B P(F|\mathcal{A}) dP = P(F|\mathcal{A}) P(B)$$

ovvero su B, $P(F|\mathcal{A}) = P(F|B)$.

Definizione 4.2.3 : *Siano X ed Y due variabili aleatorie a valori rispettivamente in $\mathbb{R}^n$ ed in $\mathbb{R}^d$ e sia X integrabile.*

Definiamo **media (probabilità) condizionata di X rispetto ad Y**

$$\begin{aligned} E(X|Y) &= E(X|\sigma(Y)) \\ P(F|Y) &= E(\mathbf{1}_F|\sigma(Y), \quad \forall F \in \mathcal{F} \end{aligned}$$

dove con $\sigma(Y)$ indichiamo la $\sigma-$algebra generata da Y.

Con l'aiuto del seguente teorema, la precedente definizione dà modo di chiarire anche cosa intendiamo per $E(X|Y=y)$ o $P(A|Y=y)$.

Teorema 4.2.1 : *Sia Y una v.a. a valori in $\mathbb{R}^d$ e sia Z una v.a. $\sigma(Y)-$misurabile a valori in $\mathbb{R}^n$. Allora esiste una funzione boreliana $\phi : \mathbb{R}^d \longrightarrow \mathbb{R}^n$ tale che $Z = \phi(Y)$.*

Dimostrazione: Possiamo considerare Z componente per componente e quindi non è restrittivo dimostrare il teorema solo per v.a. scalari.

Consideriamo l'insieme delle v.a. costituito da

$$\mathcal{L} = \{Z : Z = \phi(Y), \text{per qualche } \phi : \mathbb{R}^d \longrightarrow \mathbb{R},\, \mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}) - \textit{misurabile}\}.$$

Vogliamo dimostrare che $\mathcal{L} = \{Z : Z \,\text{è}\, \sigma(Y) - \text{misurabile}\}$.

È chiaro che $\mathcal{L} \subseteq \{Z : Z \,\text{è}\, \sigma(Y) - \text{misurabile}\}$ e dimostriamo quindi l'inclusione contraria.

Scelto qualsiasi $A \in \sigma(Y)$, la sua funzione indicatrice $\mathbf{1}_A$ appartiene necessariamente ad $\mathcal{L}$. Per definizione $A = \{\omega : Y(\omega) \in B\}$ per qualche $B \in \mathcal{B}(\mathbb{R}^d)$, quindi possiamo scegliere $\phi = \mathbf{1}_B$. Inoltre se Z_1 e $Z_2 \in \mathcal{L}$, allora $Z_1 = \phi_1(Y)$ e $Z_2 = \phi_2(Y)$, dunque

$$Z_1 + Z_2 = (\phi_1 + \phi_2)(Y) \in \mathcal{L},$$

che implica che tutte le v.a. semplici misurabili secondo $\sigma(Y)$ sono anche nella famiglia $\mathcal{L}$.

Poichè le funzioni Boreliane possono essere approssimate, secondo la convergenza puntuale, tramite funzioni semplici, è sufficiente dimostrare che $\mathcal{L}$ è chiusa rispetto alla convergenza puntuale.

Sia Z_n una successione in $\mathcal{L}$, convergente puntualmente a Z, allora $Z_n = \phi_n(Y)$ per un'opportuna successione di funzioni ϕ_n $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})-$misurabili.
Sia $B = \{y \in \mathbb{R}^d : \lim_{n\to\infty} \phi_n(y) \,\text{esiste}\}$. Si ha che $B \in \mathcal{B}(\mathbb{R}^d)$ e $\Omega = \{Y \in B\}$ definendo

$$\phi = \begin{cases} \lim_{n\to\infty} \phi_n(y) & y \in B \\ 0 & y \in B^c, \end{cases}$$

si ottiene che ϕ è una funzione boreliana e $Z = \phi(Y)$. $\square$

Osservazione 4.2.2 : *Poichè la media condizionata di una v.a. integrabile X rispetto ad Y è, per definizione, una v.a. $\sigma(Y)-$misurabile, dal teorema precedente si ottiene che $E(X|Y) = \phi(Y)$ per qualche opportuna ϕ Boreliana e dunque possiamo definire*

$$E(X|Y = y) = \phi(y).$$

Inoltre abbiamo che per ogni $F \in \sigma(Y)$ (ovvero $F = \{Y \in B\}$ per qualche B Boreliano)

$$\begin{aligned}\int_{\{Y\in B\}} XdP = \int_F XdP = \int_F E(X|Y)dP \\ = \int_F \phi(Y)dP = \int_{\{Y\in B\}} \phi(Y)dP = \int_B \phi(y)dF_Y(y),\end{aligned}$$

dove F_Y è la funzione di distribuzione di Y. Da qui possiamo ricostruire le formule di Bayes e delle probabilità totali, per esempio per ogni $A \in \mathcal{B}(\mathbb{R})$, abbiamo

$$P_X(A) = P(X \in A) = \int_R P(X \in A|Y = y)dF_Y(y).$$

La probabilità condizionata è una probabilità a tutti gli effetti così come la media condizionata è una media, come vediamo dalla prossima proposizione.

Teorema 4.2.2 : *Siano $X, Y, \{Y_n\}_{n\in\mathbb{N}}, Z$ v.a. integrabili, $a, b \in \mathbb{R}$ e sia $\mathcal{A} \subseteq \mathcal{F}$ una sotto $\sigma-$algebra. Allora q.o. valgono le seguenti proprietà:*

1. *Linearità:* $E(aX + bY|\mathcal{A}) = aE(X|\mathcal{A}) + bE(Y|\mathcal{A})$.
2. *Monotonia: Se $X \le Y$, allora $E(X\|\mathcal{A}) \le E(Y|\mathcal{A})$.*
3. *Teorema della convergenza monotona: Se $0 \le Y_n \uparrow Y$ q.o., allora*
$$E(Y_n|\mathcal{A}) \uparrow E(Y|\mathcal{A}) \qquad q.o.$$
4. $E(E(X|\mathcal{A})) = E(X)$.
5. *Se Z è $\mathcal{A}-$misurabile e $E(|ZY|), +\infty$, si ha*
$$E(ZY|\mathcal{A}) = ZE(Y|\mathcal{A}).$$
6. *Se $\mathcal{B} \subseteq \mathcal{A}$, allora*
$$E(E(X|\mathcal{A})|\mathcal{B}) = E(X|\mathcal{B}) = E(E(X|\mathcal{B})\mathcal{A}).$$

7. *Se $\sigma(Y)$ è indipendente da $\mathcal{A}$, allora $E(Y|\mathcal{A}) = E(Y)$.*

Dimostrazione: Per qualsiasi $A \in \mathcal{A}$

1. $\displaystyle\int_A E(aX+bY|\mathcal{A})dP = \int_A (aX+bY)dP = a\int_A XdP + b\int_A YdP$

   $\displaystyle = a\int_A E(X|\mathcal{A})dP + b\int_A E(Y|\mathcal{A})dP = \int_a [aE(X|\mathcal{A}) + bE(Y|\mathcal{A})]dP.$

2. $\displaystyle\int_A E(X|\mathcal{A}) = \int_A XdP \leq \int_A YdP = \int_A E(Y|\mathcal{A})dP.$

3. Da 2. abbiamo che $0 \leq Y_1 \leq Y_2 \leq \ldots \uparrow Y$ q.o. implica

$$0 \leq E(Y_1|\mathcal{A}) \leq E(Y_2|\mathcal{A}) \leq \cdots \leq E(Y|\mathcal{A}) \qquad \text{q.o.}$$

   Osserviamo che ogni $E(Y_i|\mathcal{A})$ è definita a meno di un insieme di misura nulla F_i, ma $F = \bigcup_{i=1}^{\infty} F_i$ è anch'esso di misura nulla, dunque possiamo definire univocamente ovunque tutte le medie condizionate. Dunque per monotonia $E(Y_n|\mathcal{A})(\omega) \to Z(\omega)$ per qualche Z $\mathcal{A}-$misurabile.

   D'altra parte, per ogni $A \in \mathcal{A}$

$$\begin{aligned} &0 \leq Y_1\mathbf{1}_A \leq Y_2\mathbf{1}_A \leq \ldots \uparrow Y\mathbf{1}_A \\ &0 \leq E(Y_1|\mathcal{A})\mathbf{1}_A \leq E(Y_2|\mathcal{A})\mathbf{1}_A \leq \ldots \uparrow Z\mathbf{1}_A \end{aligned}$$

   e per il teorema della convergenza monotona si ha

$$\begin{aligned} \int_A E(Y_n|\mathcal{A})dP &\longrightarrow \int_A ZdP \\ \int_A E(Y_n|\mathcal{A})dP = \int_A Y_n dP &\longrightarrow \int_A YdP = \int_A E(Y|\mathcal{A})dP. \end{aligned}$$

   Dunque $\displaystyle\int_A ZdP = \int_A E(Y|\mathcal{A})dP$, per ogni $A \in \mathcal{A}$, ovvero $Z = E(Y|\mathcal{A})$ q.o.

4. $\Omega \in \mathcal{A}$.

5. Supponiamo dapprima che $Z \geq 0$. Chiaramente $ZE(Y|\mathcal{A})$ è $\mathcal{A}-$misurabile.

   Prima di tutto verifichiamo l'identità per le funzioni semplici. Sia quindi

$$Z = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i} \qquad \text{con} \qquad a_i \in \mathbb{R},\, A_i \in \mathcal{A}.$$

Allora per ogni $A \in \mathcal{A}$, otteniamo

$$\begin{aligned}
\int_A ZYdP &= \int_A \sum_{i=1}^n a_i \mathbf{1}_{A_i} YdP = \sum_{i=1}^n a_i \int_A \mathbf{1}_{A_i} YdP \\
&= \sum_{i=1}^n a_i \int_{A\cap A_i} YdP = \sum_{i=1}^n a_i \int_{A\cap A_i} E(Y|\mathcal{A})dP \\
&= \int_A \sum_{i=1}^n a_i \mathbf{1}_{A_i} E(Y|\mathcal{A})dP = \int_A ZE(Y|\mathcal{A})dP.
\end{aligned}$$

Per Z misurabile, usiamo la densità delle funzioni semplici in $L^1(P)$ ed il punto 3.

6. $\mathcal{B} \subseteq \mathcal{A}$, perciò $E(X|\mathcal{B})$ è $\mathcal{B}-$misurabile e quindi anche $\mathcal{A}-$misurabile. Applicando il punto 5. con $Z = E(X|\mathcal{B})$ e $Y = 1$, otteniamo il primo lato dell'uguaglianza. Per l'altro lato, notiamo che $E(E(X|\mathcal{A})|\mathcal{B})$ è $\mathcal{B}-$misurabile e per ogni $A \in \mathcal{B} \subseteq \mathcal{A}$, si ha

$$\int_A E(E(X|\mathcal{A})\mathcal{B})dP = \int_A E(X|\mathcal{A})dP = \int_A XdP = \int_A E(X|\mathcal{B})dP.$$

7. $E(Y)$ è $\mathcal{A}-$misurabile in quanto funzione costante e per ogni $A \in \mathcal{A}$, abbiamo

$$\int_A E(Y|\mathcal{A})dP = \int_A YdP = \int_\Omega \mathbf{1}_A YdP = E(Y)P(A)$$

per indipendenza.

Per l'arbitrarietà dell'insieme A la tesi è dimostrata. □

Esempi 4.2.1 :

1. *Prendiamo come spazio di probabilità* $([0,1], \mathcal{B}([0,1]), Lebesgue)$ *e definiamo le sotto* $\sigma-$*algebre diadiche*

$$\mathcal{D}_n = \sigma(\{[\frac{j}{2^n}, \frac{j+1}{2^n}),\ n \in \mathbb{N},\ 0 \leq j \leq 2^n - 1\})$$

quindi una funzione per essere misurabile secondo $\mathcal{D}_n$*, deve essere costante su ogni intervallo del tipo* $[\frac{j}{2^n}, \frac{j+1}{2^n})$*. Sia ora* f *una qualsiasi funzione Boreliana, cos'è* $E(f|D_n)$*?*

Sappiamo che

$$E(f|D_n)(\omega) = \alpha_j \qquad \textit{costante per} \quad \omega \in [\frac{j}{2^n}, \frac{j+1}{2^n})$$

e che per la definizione di media condizionata si deve avere

$$\int_{[\frac{j}{2^n}, \frac{j+1}{2^n})} f d\lambda = \int_{[\frac{j}{2^n}, \frac{j+1}{2^n})} E(f|D_n) d\lambda = \alpha_j \frac{1}{2^n}$$

dunque

$$\alpha_j = 2^n \int_{[\frac{j}{2^n}, \frac{j+1}{2^n})} f d\lambda$$

ovvero la media della funzione sull'intervallino.

2. *Consideriamo* $([-\frac{1}{2}, \frac{1}{2}], \mathcal{B}([-\frac{1}{2}, \frac{1}{2}]), \lambda)$ *e le v.a.* $X(t) = t^2$, $Y(t) = t$. *La* $\sigma-$*algebra generata da* X *è* $\sigma(X) = \{$*insiemi Boreliani simmetrici rispetto allo* $0\}$, *quindi se una funzione è misurabile rispetto a* $\sigma(X)$ *deve essere simmetrica.*

   Allora $E(Y|X)$ *deve essere una funzione simmetrica tale che il suo integrale su qualsiasi insieme simmetrico è uguale all'integrale della funzione identica su qualsiasi insieme simmetrico, dunque si deve avere*

$$E(X|Y) = 0 \qquad q.o.$$

3. *Siano* X, Y *due v.a. indipendenti, allora per le proprietà prima dimostrate si ha*

$$E(3X + 2Y|Y) = 3E(X|Y) + 2E(Y|Y) = 3E(X) + Y.$$

4.3 Versioni Regolari

Esempio 4.3.1 : *Prendiamo una v.a. bidimensionale* (X, Y) *distribuita uniformemente nel triangolo* Δ *del piano cartesiano di vertici* $(0, 0)$, $(1, 0)$ *e* $(1, 2)$, *ovvero*

$$f(x) = \begin{cases} 1 & (x, y) \in \Delta \\ 0 & (x, y) \in \Delta^c. \end{cases}$$

Qual è la distribuzione di Y *dato* $X = \frac{1}{3}$*?*

Sappiamo che se abbiamo le funzioni di densità, allora

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \tag{4.3}$$

e quindi nel nostro caso si ottiene

$$
\begin{aligned}
f_X(x) &= \int_0^{2x} dy = 2x \\
f_{Y|X}(y|x) &= \frac{1}{2x}\mathbf{1}_{\{2x \leq y \leq 2\}}
\end{aligned}
$$

distribuzione uniforme su $[0, 2x]$*, dove* $x \in (0, 1]$ *è fissato.*

Come si ottiene la formula (4.3)?

Abbiamo già visto che tramite il teorema 4.2.1 riusciamo a dare un significato a $E(X|Y = y)$ adesso vogliamo capirne la sua distribuzione.

Teorema 4.3.1 : *Sia* X *una v.a. definita su* $(\Omega, \mathcal{F}, P)$ *e* $\mathcal{A} \subseteq \mathcal{F}$ *una sotto* $\sigma-$*algebra. Allora esiste una famiglia di v.a.* $\{F_t\}_{t\in\mathbb{R}}$ *tale che*

1. *Per ogni* $\omega \in \Omega$ *fissato, l'applicazione*

$$
\begin{aligned}
\phi_\omega : \quad \mathbb{R} \longrightarrow & \quad \mathbb{R} \\
t \longrightarrow & \quad F_t(\omega) = \phi_\omega(t)
\end{aligned}
$$

   è una funzione di ripartizione di probabilità.

2. *per ogni* $t \in \mathbb{R}$ *fissato, l'applicazione*

$$F_t(\omega) : \Omega \longrightarrow \mathbb{R}$$

   è uguale q.o. a $P(X \leq t|\mathcal{A})$ *(ovvero è una versione di a* $P(X \leq t|\mathcal{A})$*).*

Dimostrazione: Sia $r \in \mathbb{Q}$. Consideriamo una funzione $g_r(\omega)$ che sia q.o. uguale a $P(X \leq r|\mathcal{A})$ (ovvero una versione).

Vogliamo dimostrare che g_r è una funzione di ripartizione su $\mathbb{Q}$.

Per definizione $0 \leq g_r(\omega) \leq 1$ q.o.

Se $r < s \in \mathbb{Q}$, allora $\mathbf{1}_{\{X\leq r\}} \leq \mathbf{1}_{\{X\leq s\}}$ e , per le proprietà della media condizionata ciò implica

$$g_r(\omega) \leq g_s(\omega)$$

per quasi ogni ω.

Quindi se $A_{r,s} = \{\omega : g_r(\omega) > g_s(\omega)\}$, allora $P(A_{r,s}) = 0$ e vale $P(\bigcup_{r,s\in\mathbb{Q}} A_{r,s}) = 0$ in quanto unione numerabile.

Poniamo $B_1 = \{\omega : \lim_{\substack{r\to\infty \\ r\in\mathbb{Q}}} g_r(\omega) = 1\}$. Per $\omega \notin \bigcup_{r,s\in\mathbb{Q}} A_{r,s}$ e per una qualsiasi successione di razionali $\{r_i\}$ crescente ad infinito, avviene

$$g_{r_1}(\omega) \le g_{r_2}(\omega) \le \dots,$$

e dunque $\{\omega : \lim_{\substack{r\to\infty \\ r\in\mathbb{Q}}} g_r(\omega) = 1\} = \lim_{n\to\infty} g_n(\omega) = 1\}$.

D'altra parte q.o.

$$g_n(\omega) = P(X \le n|\mathcal{A}) = E(\mathbf{1}_{\{X\le n\}}|\mathcal{A})(\omega)$$

e $\mathbf{1}_{\{X\le n\}} \uparrow 1$ per $n \to +\infty$, dunque per il teorema della convergenza monotona

$$P(X \le n|\mathcal{A}) \uparrow E(1|\mathcal{A}) = 1 \qquad q.o.$$

che implica $P(B_1) = 1$ e $P(B_1^c) = 0$.

Analogamente possiamo dimostrare che l'insieme

$$B_0 = \{\omega : \lim_{\substack{r\to-\infty \\ r\in\mathbb{Q}}} g_r(\omega) = 0\}$$

ha probabilità 1 e quindi il suo complementare ha probabilità 0.

Definiamo ora

$$F_t(\omega) = \begin{cases} \lim_{\substack{r\downarrow t \\ r\in\mathbb{Q}}} g_r(\omega) & \omega \notin \bigcup_{r,s\in\mathbb{Q}} A_{r,s} \cup B_0^c \cup B_1^c \\ \int_{-\infty}^{t} \frac{1}{\pi(1+s^2)} ds & \text{altrimenti} \end{cases}$$

allora $F_t(\omega)$ è una versione regolare di $P(X \le t|\mathcal{A})(\omega)$.

Vogliamo dimostrare che $F_t(\omega)$ è una funzione di ripartizione per ogni ω fissato.

Infatti

(i) $F_t(\omega) = \lim_{\substack{r\downarrow t \\ r\in\mathbb{Q}}} g_r(\omega)$ q.o. ; $g_r(\omega$ è non decrescente sui razionali, dunque $F_t(\omega)$ è continua a destra. (l'altra lo è automaticamente)

(ii) Necessariamente $\lim_{t\to+\infty} F_t(\omega) = 1$ e $\lim_{t\to-\infty} F_t(\omega) = 0$.

(iii) grazie alla limitatezza della funzione devono esistere i limiti sinistri.

Infine $F_t(\omega)$ è una versione di $P(X \leq t|\mathcal{A})$, poichè lungo qualsiasi successione decrescente di razionali si ha

$$F_t(\omega) = \lim_{n\to+\infty} g_{r_n}(\omega),$$

per il teorema della convergenza monotona $g_{r_n} \downarrow E(\mathbf{1}_{\{x\leq t\}}|\mathcal{A})$ q.o. per $n \to +\infty$.

$\square$

Osservazione 4.3.1 : *Associata a $F_t(\omega)$, come definita nel precedente teorema, vi è una misura di probabilità su $\mathcal{B}(\mathbb{R})$ che indicheremo con μ_ω.*

Proposizione 4.3.1 : *Per qualsiasi insieme $B \in \mathcal{B}(\mathbb{R})$ fissato, si ha*

$$\mu_\omega(B) = P(X \in B|\mathcal{A})(\omega) \qquad q.o.$$

Dimostrazione: Consideriamo la famiglia di insiemi

$$\mathcal{H} = \{B \in \mathcal{B}(\mathbb{R}) : \mu_\omega(B) = P(X \in B|\mathcal{A})(\omega)\,\text{q.o.}\}.$$

Gli intervalli del tipo $(-\infty, t] \in \mathcal{H}$, $\forall t \in \mathbb{R}$. Per le proprietà della probabilità condizionata, anche gli intervalli del tipo $(s, t] \in \mathcal{H}$, $\forall s, t \in \mathbb{R}$, quindi per linearità tutta l'algebra da essi generata $\mathcal{P}$ è contenuta in $\mathcal{H}$.

Applicando il teorema della classe monotona si può dimostrare che

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{P}) \subseteq \mathcal{H}. \qquad \square$$

Proposizione 4.3.2 : *Scelta una qualsiasi funzione Boreliana $\phi : \mathbb{R} \longrightarrow \mathbb{R}$, abbiamo*

$$E(\phi(X)|\mathcal{A}) = \int_{\mathbb{R}} \phi(x) d\mu_\omega(x) \qquad q.o.\ \omega \in \Omega$$

Dimostrazione: Si dimostra prima per funzioni indicatrici e per funzioni semplici, quindi si passa al limite per funzioni misurabili. $\square$

Proposizione 4.3.3 : *(Disuguaglianza condizionale di Jensen)*

Sia X una v.a. a valori in $\mathbb{R}$ e $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ una funzione convessa, tali che X e $\phi(X)$ siano entrambe integrabili. Allora

$$\phi(E(X|\mathcal{A})) \leq E(\phi(X)|\mathcal{A})$$

Dimostrazione: Sia μ_ω la versione regolare della distribuzione condizionata ad $\mathcal{A}$, allora

$$\begin{aligned} E(\phi(X)|\mathcal{A}) &= \int_{\mathbb{R}} \phi(x) d\mu_\omega(x) \qquad \text{q.o. } \omega \in \Omega \\ E(X|\mathcal{A}) &= \int_{\mathbb{R}} x d\mu_\omega(x) \qquad \text{q.o. } \omega \in \Omega \end{aligned}$$

ed applicando la disuguaglianza di Jensen usuale, otteniamo la tesi. $\square$

Capitolo 5

MARTINGALE A TEMPO DISCRETO

In questo capitolo presntiamo la teoria delle martingale a tempo discreto. Questo tipo di struttura è importante poichè fornisce gli strumenti per affrontare lo studio di processi a tempo continuo quali il moto Browniano. Infatti le proprietà ed i teoremi relativi alla teoria delle martingale risultano più comprensibili nel contesto discreto e sono gli stessi che possono essere estesi al continuo.

5.1 Definizioni

Come al solito $(\Omega, \mathcal{F}, P)$ è il nostro spazio di probabilità di riferimento.

Definizione 5.1.1 : *Una successione $\{\mathcal{F}_n\}$ di $\sigma-$algebre contenute in $\mathcal{F}$ tali che*

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$$

è detta una **filtrazione**. *Uno spazio di probabilità dotato di una filtrazione è detto spazio filtrato. Se un processo $\{X_n\}$ è tale che per ogni n X_n è $\mathcal{F}_n$ misurabile, allora si dice* **adattato** *alla filtrazione $\{\mathcal{F}_n\}_n$.*

Se un processo $\{X_n\}$ è tale che per ogni n X_n è $\mathcal{F}_{n-1}$ misurabile, allora si dice **prevedibile** *rispetto alla filtrazione $\{\mathcal{F}_n\}_n$.*

Notiamo che dato un processo $\{X_n\}$ su $(\Omega, \mathcal{F}, P)$, si può definire una filtrazione rispetto alla quale risulti automaticamente adattato, detta per l'appunto **filtrazione naturale** di $\{X_n\}$.

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n)$$

Essendo $\mathcal{F}_n$ per definizione la più piccola $\sigma-$algebra rispetto alla quale $X_1, \dots, X_n$ risultano tutte misurabili, la filtrazione naturale è la più piccola filtrazione rispetto alla quale il processo risulta adattato.

Definizione 5.1.2 : *Una successione di v.a.* $\{X_n\}$ *è detta una* **martingala** *rispetto a* $\{\mathcal{F}_n\}$ *se*

(a) ogni X_i *è integrabile (ovvero* $E(|X_i|) < +\infty$*) e* $\mathcal{F}_i-$*misurabile (adattato);*

(b) $E(X_{i+1}|\mathcal{F}_i) = X_i$ *o equivalentemente* $E(X_{i+1} - X_i|\mathcal{F}_i) = 0$ *q.o.*

Osservazione 5.1.1 : *Il concetto di martingala è intrinseco al processo e non alla filtrazione presa in considerazione. Infatti se* $\{X_n\}$ *è una martingala rispetto a* $\{\mathcal{F}_n\}$ *allora lo è anche rispetto alla sua filtrazione naturale* $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$*. Infatti per minimalità* $\mathcal{G}_n \subseteq \mathcal{F}_n$ *per ogni* n *e si ha*

$$E(X_{i+1}|\mathcal{G}_i) = E(E(X_{i+1}|\mathcal{F}_i)\mathcal{G}_i) = E(X_i|\mathcal{G}_i) = X_i.$$

Infine osserviamo che una martingala ha media costante, poichè

$$E(X_{i+1}) = E(E(X_{i+1}|\mathcal{F}_i)) = E(X_i), \quad \forall\, i.$$

La teoria delle martingale proviene dalla modellizzazione dei giochi di scommesse. Se la vincita ad un certo gioco può essere rapresentata tramite una martingala, allora essa è in media costante ovvero, se opportunamente riscalata, nulla. In altre parole, in media non si vince e non si perde ed il gioco è detto equo. Occorre però stare attenti ch e avere media costante per una successione di v.a. non è una condizione sufficiente per essere una martingala.

Diamo ora alcuni esempi di martingale. A meno di specificare altrimenti, d'ora in poi assumeremo la proprietà di martingala rispetto alla filtrazione naturale del processo.

Esempi 5.1.1 :

1. **Somme di v.a. indipendenti.** *Sia* $\{X_n\}$ *una successione di v.a. indipendenti, tutte integrabili ed a media nulla e consideriamo* $S_n = \sum_{i=1}^{n} X_i$*, allora* $\{S_n\}$ *è una martingala.*

   Infatti S_n *è* $\mathcal{G}_n-$*misurabile ed integrabile* $\Big(E(|S_n|) \leq \sum_{i=1}^{n} E(|X_i|) < +\infty \Big)$ *e*

   $$E(S_n|\mathcal{G}_{n-1}) = E\Big(\sum_{i=1}^{n} X_i|\mathcal{G}_{i-1} \Big) = \sum_{i=1}^{n-1} X_i + E(X_n|\mathcal{G}_{n-1}) = S_{n-1} + E(X_i) = S_{n-1}$$

   dove negli ultimi due passaggi abbiamo usato l'indipendenza e la media nulla delle v.a.

   Se le v.a. sono tali che $E(X_1) = \mu \neq 0$*, allora è il processo* $\{S_n - n\mu\}$ *ad essere una martingala (verificare).*

2. **Martingala lascia o raddoppia**.

Prendiamo lo spazio di probabilità $([0,1],\mathcal{B}([0,1]),\lambda)$*, dove con* λ *denotiamo la misura di Lebesgue. Per* $k \in \mathbb{N}$ *indichiamo con* $\mathcal{D}_k$ *la* $k-$*esima* $\sigma-$*algebra diadica, ovvero*

$$\mathcal{D}_0 = \{\emptyset,[0,1]\},\ldots,\mathcal{D}_k = \sigma(\{\,[0,\frac{1}{2^k}],[\frac{1}{2^k},\frac{2}{2^k}],\ldots,[\frac{2^k-1}{2^k},1]\,\})$$

e poniamo $f_n = 2^n\mathbf{1}_{[0,2^{-n}]}$ *(double or nothing martingale). Allora* $\{f_n\}$ *è una martingala . Essendo tutte le* f_n *limitate e quindi integrabili su* $[0,1]$*, è sufficiente dimostrare la proprietà di martingala sulle semialgebre degli intervalli diadici di ordine* n *che generano la* $n-$*esima* $\sigma-$*algebra. Infatti per* $\mathcal{D}_{n-1}$ *gli intervalli diadici di ordine* $n-1$*, che denotiamo con* $\{I_1,\ldots,I_{2^{n-1}}\}$*, costituiscono gli atomi della* $\sigma-$*algebra (ovvero non esistono insiemi contenuti in essi che siano* $\mathcal{D}_{n-1}-$*misurabili. Dunque se una funzione è misurabile rispetto a* $\mathcal{D}_{n-1}$ *essa deve essere costante sugli atomi della* $\sigma-$*algebra. Per definizione* $E(f_n|\mathcal{D}_{n-1})$ *è* $\mathcal{D}_{n-1}-$*misurabile, quindi costante su ogni* $\{I_1,\ldots,I_{2^{n-1}}\}$*, inoltre deve verificare*

$$\int_{I_j} E(f_n|\mathcal{D}_{n-1})d\lambda = \int_{I_j} f_n d\lambda, \qquad \forall\, j = 1,\ldots,2^{n-1}$$

perciò si ha per $I_1 = [0,\dfrac{1}{2^{n-1}}]$

$$\begin{aligned} E(f_n|\mathcal{D}_{n-1})\frac{1}{2^{n-1}} &= E(f_n|\mathcal{D}_{n-1})\lambda(I_1) = \int_{I_1} E(f_n|\mathcal{D}_{n-1})d\lambda \\ &= \int_{I_1} f_n d\lambda = \int_0^{2^{-(n-1)}} 2^n\mathbf{1}_{[0,2^{-n}]}d\lambda \int_0^{2^{-n}} 2^n d\lambda = 1 \end{aligned}$$

mentre per $j = 2,\ldots,2^{n-1}$

$$E(f_n|\mathcal{D}_{n-1})\lambda(I_j) = \int_{I_j} E(f_n|\mathcal{D}_{n-1})d\lambda = \int_{I_j} f_n d\lambda = 0$$

dunque $E(f_n|\mathcal{D}_{n-1}) = 2^{n-1}\mathbf{1}_{[0,2^{-(n-1)}]} = f_{n-1}$.

Perchè si chiama martingala lascia o raddoppia? Pensiamo di star giocando a testa e croce con una moneta equa con un capitale iniziale $f_0 = 1$*, puntando sempre su testa tutto il capitale vinto precedentemente. Allora si avrà*

$$f_1 = \begin{cases} 2 & \frac{1}{2} \quad \textit{se } T \\ 0 & \frac{1}{2} \quad \textit{se } C \end{cases} \qquad \textit{in generale} \quad f_k = \begin{cases} 2^k & \frac{1}{2^n} \quad \textit{se } T\ldots T \ k \textit{ volte} \\ 0 & 1-\frac{1}{2^n} \quad \textit{altrimenti.} \end{cases}$$

3. *Sia $\{X_n\}$ una successione di v.a. i.i.d con $P(X_1 = 1) = P(X_1 = -1) = \dfrac{1}{2}$ (dunque $E(X_i) = 0$, $E(X_i^2) = 1$ per ogni i). Allora, con la stessa notazione di prima, $S_n^2 - n$ è una martingala; infatti l'integrabilità e la misurabilità sono chiare ($E(|S_n^2 - n|) \leq 2n$) e*

$$\begin{aligned} E(S_{n+1}^2 - (n+1)|\mathcal{G}_n) &= E(S_{n+1}^2|\mathcal{G}_n) - (n+1) \\ &= E((S_n + X_{n+1})^2|\mathcal{G}_n) - (n+1) \\ &= E((S_n^2 + X_{n+1}^2 + 2S_nX_{n+1}|\mathcal{G}_n) - (n+1) \\ &= S_n^2 + E(X_{n+1}^2|\mathcal{G}_n) + 2S_nE(X_{n+1}|\mathcal{G}_n) - (n+1) \\ &= S_n^2 + E(X_{n+1}^2) + 2S_nE(X_{n+1}) - (n+1) = S_n^2 - n \end{aligned}$$

dove abbiamo ottenuto l'ultima uguaglianza per indipendenza.

Dalla dimostrazione si osserva che l'ultimo esempio rimane valido anche se le v.a. non sono identicamente distribuite, qualora continuino a valere le condizioni $E(X_i) = 0, E(X_i^2) = 1$ per ogni i. Se, inoltre la successione di v.a. è indipendente con $E(X_i) = 0, E(X_i^2) = \sigma^2$ per ogni i, allora il processo $\{S_n^2 - n\sigma^2\}$ è una martingala (verificare).

Nel caso in cui si abbia una martingala finita, allora la si può caratterizzare con un'unica v.a.

Proposizione 5.1.1 : *Sia n fissato. Se $X_1, \ldots, X_n$ è una martingala finita rispetto alla filtrazione $\mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$, allora deve esistere una v.a. integrabile Y tale che*

$$E(Y|\mathcal{F}_k) = X_k, \qquad \forall\, k = 1, \ldots, n.$$

Viceversa, se Y è una v.a. integrabile, allora la sequenza $X_k = E(Y|\mathcal{F}_k)$ per $k = 1, \ldots, n$ è una martingala finita rispetto a $\mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$.

Dimostrazione: Nella prima direzione, la tesi viene direttamente dalla definizione di martingala, scegliendo $Y = X_n$.

Viceversa $E(E(Y|\mathcal{F}_{k+1})|\mathcal{F}_k) = E(Y|\mathcal{F}_k)$ per le proprietà di media condizionata e per la definizione di filtrazione. $\square$

La seconda parte della dimostrazione rimane valida anche nel caso di una successione infinita, mentre la prima potrebbe non essere verificata.

Controesempio: siano $\{X_i\}_{i\in\mathbb{N}}$ v.a. i.i.d. tali che $P(X_1 = 1) = P(X_1 = -1) = \dfrac{1}{2}$ e sia $S_n = \displaystyle\sum_{i=1}^{n} X_i$, allora non esiste alcuna v.a. integrabile Y tale che $S_n = E(Y|\mathcal{G}_n)$.

Dimostrazione: supponiamo per assurdo che esista una tale v.a. Y, allora per la disuguaglianza condizionale di Jensen si dovrebbe avere

$$E(|Y|\,|\mathcal{G}_n) \geq |E(Y|\mathcal{G}_n)| = |S_n|$$

da cui $E(|S_n|) \leq E(|Y|)$ per ogni n, ma $E(|S_n|) \longrightarrow +\infty$ per $n \to +\infty$ (verificarlo per esercizio), che implica una contraddizione . □

Definizione 5.1.3 : *Sia $\{\mathcal{F}_n\}$ una filtrazione, una successione di v.a. $\{X_n\}$ è detta una* **submartingala** *(rispettivamente una* **supermartingala***) rispetto a $\{\mathcal{F}_n\}$ se*

1. *ogni X_i è integrabile e la successione è adattata;*
2. *$E(X_{i+1}|\mathcal{F}_i) \geq X_i$ (rispettivamente $E(X_{i+1}|\mathcal{F}_i) \leq X_i$).*

Chiaramente una martingala è sia una submartingala sia una supermartingala.

Osservazione 5.1.2 : *per una subamrtingala si ha*

$$E(X_n) = E(E(X_n|\mathcal{F}_{n-1})) \geq E(X_{n-1})$$

ovvero la successione delle medie è crescente in n, mentre per una supermartingala

$$E(X_{n-1}) \geq E(X_n)$$

è descrescente in n.

Esempio 5.1.1 *Sia $\{X_n\}$ una successione di v.a. indipendenti, tutte con la stessa media μ e $S_n^X = \sum_{i=1}^{n} X_i$.*

Allora $S_n^X - n\mu$ è una martingala e se $\mu > 0$

$$E(S_n^X|\mathcal{G}_{n-1}) = S_{n-1}^X + \mu > S_{n-1}^X$$

ovvero S_n^X è una submartingala, mentre si ottiene una supermartingala se $\mu < 0$.

La seguente proposizione è immediata.

Proposizione 5.1.2 : *Se $\{X_n\}$ è una martingala a valori reali e $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ è una funzione convessa (rispettivamente concava) e $E(|\phi(X_i)|) < +\infty$ per ogni i, allora $\{\phi(X_n)\}$ è una submartingala (rispettivamente una supermartingala).*

Dimostrazione: grazie alla disuguaglianza condizionale di Jensen. □

Esempio 5.1.2 : *la precedente proposizione implica che se* $\{X_n\}$ *è una martingala, allora* $\{|X_n|^p\}$ *per* $p \geq 1$ *è una submartingala per la convessità delle potenze. Analogamente* $\{|X_n|^{\frac{1}{p}}\}$ *per* $p \geq 1$ *è una supermartingala.*

Introduciamo ora un teorema molto importante nella teoria delle martingale e che ha un suo corrispondente anche al continuo.

Teorema 5.1.1 : **Teorema di decomposizione di Doob per submartingale**
Sia $\{X_n\}$ *una submartingala e sia* $Y_i = X_i - X_{i-1}$ *la corrispondente successione delle differenze. Allora possiamo scrivere in maniera unica*

$$Y_i = Z_i + D_i,$$

dove Z_i *è la successione delle differenze di una martingala e per ogni* i*,* $D_i \geq 0$ *e* $\mathcal{G}_{i-1}-$*misurabile (ovvero è un processo prevedibile).*

Dimostrazione: Poniamo $D_i = E(Y_i|\mathcal{G}_{i-1})$, allora per le proprietà di submartingala, si ha $D_i \geq 0$ e per la media condizionata è $\mathcal{G}_{i-1}-$misurabile.

La decomposizione è unica, infatti se, per assurdo, ve ne fossero due

$$\tilde{Z}_i + \tilde{D}_i = Y_i = Z_i + D_i,$$

allora $\tilde{Z}_i - Z_i = D_i - \tilde{D}_i$, dove $\tilde{Z}_i - Z_i$ è una successione delle differenze di una martingala e quindi lo è anche $D_i - \tilde{D}_i$, che è $\mathcal{G}_{i-1}-$misurabile, da cui si conclude che $D_i = \tilde{D}_i$. □

Rivisitiamo quanto abbiamo detto in termini di teoria dei giochi d'azzardo. Supponiamo si giochino una serie di partite ad un qualche gioco dove ogni volta si vince o perde una posta unitaria. Tutti i possibili esiti delle partite fino al tempo n costituiscono l'insieme delle informazioni sul gioco accumulate fino a quel tempo, matematicamente l'$n-$esima $\sigma-$algebra della filtrazione. In altre parole, se indichiamo con X_n il capitale (fortuna) in mano al giocatore dopo l'$n-$esima partita, allora la filtrazione di riferimento è quella naturale $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$ generata dalla successione delle fortune, ovvero il giocatore conosce solo ciò che viene rivelato dall'andamento delle partite.

Se, rispetto alla filtrazione $\{\mathcal{G}_n\}$, è valida la condizione di martingala per la successione dei capitali, questo vuol dire che la fortuna attesa dopo la prossima partita è uguale alla fortuna attuale; in media non si vince nè si perde e dunque il gioco è equo.

Se invece fosse la condizione di submartingala ad essere verificata, questo vorrebbe dire che il gioco è favorevole al giocatore, poichè in media la fortuna aumenta. Nel caso si avesse una supermartingala, il gioco sarebbe invece sfavorevole al giocatore.

Dato un gioco equo, è naturale pensare che il giocatore possa tentare di attuare una strategia di scommesse che porti il gioco in suo favore, ovvero che in media aumenti il capitale vinto/perso. Supponiamo perciò che il giocatore parta con un capitale iniziale X_0 ed indichiamo con $\Delta_n = X_n - X_{n-1}$ la vincita o perdita netta di capitale, subita dopo l'$n-$esima partita avendo scommesso una unità per partita. Al tempo n il giocatore decide di investire sulla sua scommessa una quantità aleatoria H_n di denaro, che deve risultare $\mathcal{G}_{n-1}-$misurabile, poichè può basare le sue decisioni solamente su quanto ha osservato fino al tempo $n-1$, non avendo nessuna conoscenza degli andamenti futuri (altrimenti il gioco non sarebbe equo). Matematicamente stiamo dicendo che il giocatore appronta una strategia di investimento rappresentata da un processo $\{H_n\}$ prevedibile.

Di conseguenza, la ricchezza accumulata al tempo n dal giocatore, seguendo la strategia $\{H_n\}$, è data da

$$Y_n = X_0 + H_1\Delta_1 + \cdots + H_n\Delta_n = X_0 + \sum_{i=1}^{n} H_i\Delta_i.$$

Allora per la prevedibilità del processo $\{H_n\}$,

$$\begin{aligned} E(Y_n|\mathcal{G}_{n-1}) &= E(X_0 + \sum_{i=1}^{n} H_i\Delta_i|\mathcal{G}_{n-1}) \\ &= X_0 + \sum_{i=1}^{n-1} H_i\Delta_i + H_nE((X_n - X_{n-1})|\mathcal{G}_{n-1}) \\ &= X_0 + \sum_{i=1}^{n-1} H_i\Delta_i = Y_{n-1}. \end{aligned}$$

Se $\{X_n\}$ è una martingala, tale risulta anche $\{Y_n\}$, cioè qualsiasi strategia di scommessa "onesta" (prevedibile) non altera l'equità del gioco.

Se invece si parte da un gioco rappresentato da una submartingala (favorevole al giocatore), chiaramente una strategia positiva, $H_n \geq 0$, $\forall n$, mantiene il gioco favorevole, poichè in media la fortuna tende ad aumentare. Viceversa se si approntasse una strategia negativa, $H_n \geq 0$, $\forall n$, questa renderebbe il gioco sfavorevole. La situazione è speculare, quando si parte da una supermartingala.

Si potrebbe pensare di approntare strategie di scommessa aleatorie anche nel tempo, vedremo che neanche queste alterano l''equità del gioco, ovvero che preservano la proprietà di martingala del processo originale.

Definizione 5.1.4 : *definiamo* **norma** L^p*, per* $p \geq 1$ *di una martingala* $\{X_n\}$ *la quantità*

$$(\sup_{n\geq 1} E(|X_n|^p))^{\frac{1}{p}} = (\lim_{n\to+\infty} E(|X_n|^p))^{\frac{1}{p}},$$

dove l'uguaglianza è resa possibile, poichè $\{|X_n|\}$ è una submartingala.

Diciamo inoltre che una martingala è **limitata in** L^p*, se la sua norma* L^p *è finita.*

Poichè siamo su spazi di misura finita, le solite inclusioni per gli spazi L^p valgono. Notiamo inoltre che qualsiasi martingala non negativa deve essere necessariamente limitata in L^1, infatti

$$E(|X_i|) = E(X_i) = E(X_0),, \qquad \forall\, i$$

per la costanza dellla media e dunque $\sup_{n\geq 0} E(X_n) = E(X_0) < +\infty$.

5.2 Tempi d'arresto e optional sampling theorem

I concetti esposti in questo paragrafo possono essere estesi dal tempo discreto a quello continuo.

Come al solito, sia dato uno spazio di probabilità filtrato $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_n, P)$.

Definizione 5.2.1 : *una v.a. τ è detta* **tempo d'arresto** *(t.a.) rispetto a $\{\mathcal{F}_n\}$ se è una v.a. a valori interi tale che l'insieme $\{\tau = n\}$ sia $\mathcal{F}_n-$misurabile.*

La condizione $\{\tau = n\} \in \mathcal{F}_n$, che caratterizza la precedente definizione, è da intendersi nel seguente senso: si decide di arrestare un qualche fenomeno aleatorio al tempo n solo sulla base dell'informazione disponibile fino a quel momento.

Esempio 5.2.1 : *siano $\{X_n\}$ una successione di v.a. i.i.d con $P(X_1 = -1) = P(X_1 = 1) = \dfrac{1}{2}$, $S_n = \sum_{i=1}^{n} X_i$ e $\tau = \min(4, \inf\{k : S_k = 1\})$, allora τ è un tempo d'arresto rispetto alla filtrazione naturale generata da $\{X_n\}$. Infatti τ può assumere soltanto i valori $1, 2, 3, 4$ e*

$$\begin{aligned}
&\{\tau = 1\} = \{X_1 = 1\} \in \sigma(X_1)\\
&\{\tau = 2\} = \emptyset \in \sigma(X_1, X_2)\\
&\{\tau = 3\} = \{X_1 = -1, X_2 = 1, X_3 = 1\} \in \sigma(X_1, X_2, X_3)\\
&\{\tau = 4\} = (\{\tau = 1\} \cup \{\tau = 2\} \cup \{\tau = 3\})^c \in \sigma(X_1, X_2, X_3, X_4).
\end{aligned}$$

Invece la v.a. $\gamma = \min(4, \sup\{k < 4 : S_k = 1\})$ non è un tempo d'arresto, poichè

$$\{\gamma = 1\} = \{X_1 = 1\} \cap \{X_1 + X_2 + X_3 \neq 1\} \in \sigma(X_1, X_2, X_3) \neq \sigma(X_1).$$

Proprietà dei tempi di arresto

1. La v.a. $\tau = n$ con $P = 1$ ` un tempo d'arresto, poichè è una funzione costante e quindi misurabile rispetto ad $\mathcal{F}_0 \subseteq \mathcal{F}_i$ per ogni i. Infatti

$$\{\tau = i\} = \begin{cases} \emptyset & \text{se} \quad i \neq n \\ \Omega & \text{se} \quad i = n \end{cases}$$

2. Se τ e Σ sono tempi di arresto allora anche $\tau + \Sigma$ lo è, infatti per ogni n

$$\{\tau + \Sigma = n\} = \{\tau = n - \Sigma\} = \bigcup_{k=0}^{n} \{\tau = n - k, \Sigma = k\} = \bigcup_{k=0}^{n} \{\tau = n - k\} \cap \{\Sigma = k\}$$

e $\{\Sigma = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$, $\{\tau = n - k\} \in \mathcal{F}_{n-k} \subseteq \mathcal{F}_n$, per cui anche le intersezioni $\{\tau = n - k\} \cap \{\Sigma = k\} \in \mathcal{F}_n$ per ogni $k = 0, \ldots, n$ e così anche la loro unione.

La v.a. $\tau - \Sigma$ invece può non essere un tempo di arresto poichè

$$\{\tau - \Sigma = n\} = \bigcup_{k \in \mathbb{N}} \{\Sigma = k\} \cap \{\tau = n + k\},$$

ma l'evento $\{\tau = n + k\}$ può non appartenere a $\mathcal{F}_n$.

3. Se τ e Σ sono tempi di arresto allora lo sono anche $\tau \vee \Sigma$ e $\tau \wedge \Sigma$, infatti

$$\begin{aligned} \{\tau \vee \Sigma = n\} &= \{\tau = n, \Sigma \leq n\} \cup \{\tau < n, \Sigma = n\} \\ &= \bigcup_{k=0}^{n} \{\tau = n\} \cap \{\Sigma = k\} \cup \bigcup_{k=0}^{n-1} \{\tau = k\} \cap \{\Sigma = n\} \in \mathcal{F}_n \end{aligned}$$

perchè $\{\tau = n\} \in \mathcal{F}_n$, $\{\Sigma = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$, $\{\tau = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$, $\{\Sigma = n\} \in \mathcal{F}_n$ e cosí anche la loro unione.

Analogamente

$$\begin{aligned} \{\tau \wedge \Sigma = n\} &= \{\tau = n, \Sigma > n\} \cup \{\tau \geq n, \Sigma = n\} \\ &= \{\tau = n\} \cap \{\Sigma > n\} \cup \{\tau \geq n\} \cap \{\Sigma = n\} \\ &= \{\tau = n\} \cap \{\Sigma \leq n\}^c \cup \{\tau \leq n - 1\} \cap \{\Sigma = n\} \in \mathcal{F}_n \end{aligned}$$

Come conseguenza, se τ è un tempo di arresto, per ogni $n \in \mathbb{N}$ fissato, anche la v.a. $\tau \wedge n$ lo è ed inoltre è limitata.

Se abbiamo un processo $\{X_n\}$ adattato ad una filtrazione $\{\mathcal{F}_n\}$ e un tempo di arresto τ rispetto alla stessa filtrazione tale che $P(\tau < +\infty) = 1$, possiamo definire la v.a.

$$X_\tau = \sum_{k=0} X_k \mathbf{1}_{\{\tau=k\}},$$

conseguentemente

$$E(X_\tau) = E(E(X_\tau|\tau)) = E(\sum_{k=0}^{+\infty} E(X_k \mathbf{1}_{\{\tau=k\}}|\tau)) = E(\sum_{k=0}^{+\infty} E(X_k|\tau)\mathbf{1}_{\{\tau=k\}})$$

e se $\tau \perp \{X_n\}$ allora

$$E(X_\tau) = \sum_{k=0}^{+\infty} E(X_k)P(\tau = k).$$

Inoltre possiamo definire il processo arrestato al tempo τ come

$$X_{n\wedge\tau} = X_n \mathbf{1}_{\{\tau>n\}} + X_\tau \mathbf{1}_{\{\tau\leq n\}}.$$

Proposizione 5.2.1 : *Se $\{X_n\}$ è una martingala e τ un tempo di arresto rispetto alla filtrazione naturale generata da $\{X_n\}$, allora anche $\{X_{n\wedge\tau}\}$ è una martingala.*

Dimostrazione: Basta considerare la successione delle differenze $D_i = X_i - X_{i-1}$. Allora la successione delle differenze del processo arrestato è data da

$$Z_i := X_{i\wedge\tau} - X_{(i-1)\wedge\tau} = \begin{cases} 0 & \text{se} \quad T \leq i-1 \\ D_i & \text{se} \quad T > i-1 \end{cases}$$

e quindi necessariamente $E(Z_i|\mathcal{G}_{i-1}) = 0$. $\square$

Osservazione 5.2.1 : *Se rielaboriamo la precedente proposizione alla luce della teoria dei giochi di azzardo, esattamente come prima, qualsiasi strategia di scommessa basata su un tempo di arresto preserva la proprietà di martingala, ovvero mantiene un gioco equo se tale è in partenza.*

Sia infatti $\{X_n\}$ una martingala rispetto ad una filtrazione $\{\mathcal{F}_n\}$ (che rappresenta le vincite fino al tempo n), τ un t.a. rispetto alla stessa filtrazione ed $\{H_n\}$ un processo prevedibile, cioè per ogni n $H_n \in \mathcal{F}_{n-1}$.

Allora la ricchezza totale ottenuta giocando il gioco equo rappresentato da $\{X_n\}$, seguendo la strategia $\{H_n\}$ arrestata al tempo τ è

$$\tilde{Y}_n = X_0 + H_{1\wedge\tau}(X_{1\wedge\tau} - X_0) + \cdots + H_{n\wedge\tau}(X_{n\wedge\tau} - X_{(n-1)\wedge\tau})$$

ed il processo $\{\tilde{Y}_n\}$ *è adattato alla filtrazione, poichè*

$$\tilde{Y}_n = X_0 + \sum_{i=0}^{n-1} Y_i \mathbf{1}_{\{\tau=i\}} + Y_n \mathbf{1}_{\{\tau\geq n\}}$$

dove il processo $\{Y_n\}$ *è stato definito a pag. 98 e le v.a.* $Y_i\mathbf{1}_{\{\tau=i\}}$ *sono* $\mathcal{F}_i-$*misurabili (e quindi anche* $\mathcal{F}_n-$*misurabili) e* $Y_n\mathbf{1}_{\{\tau\geq n\}}$ *è* $\mathcal{F}_n-$*misurabile. Analogamente il processo* $\{H_{n\wedge\tau}\}$ *rimane prevedibile, perchè*

$$H_{n\wedge\tau} = \sum_{i=0}^{n-1} H_i \mathbf{1}_{\{\tau=i\}} + H_n \mathbf{1}_{\{\tau\geq n\}}.$$

Dunque anche $\{\tilde{Y}_n\}$ *è una martingala, difatti*

$$\begin{aligned} E(\tilde{Y}_n|\mathcal{F}_{n-1}) &= E(\tilde{Y}_{n-1} + H_{n\wedge\tau}(X_{n\wedge\tau} - X_{(n-1)\wedge\tau})|\mathcal{F}_{n-1}) \\ &= \tilde{Y}_{n-1} + H_{n\wedge\tau}E((X_{n\wedge\tau} - X_{(n-1)\wedge\tau})|\mathcal{F}_{n-1}) = \tilde{Y}_{n-1}. \end{aligned}$$

Vediamo ora due applicazioni della proposizione 5.2.1.

1. Siano $\{X_n\}_n$ una successione i.i.d di v.a. con $P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}$, $S_n = \sum_{i=1}^{n} X_i$ ed $a, b \in \mathbb{Z}$ nonnegativi.

   Vogliamo calcolare $P(S_n = a$ prima che $S_n = -b)$. Per far ciò, definiamo

   $$\tau = \inf\{S_n = a \,\text{oppure}\;\; S_n = -b\}$$

   τ è un t.a. poichè $\{\tau > n\} = \{\max_{i\leq n} |S_i| \leq \min(a, b)\} \in \mathcal{G}_n$.Inoltre $P(\tau < +\infty) = 1$; infatti notiamo che $\{S_n\}$ è anche una CM a stati numerabili con $S_0 = 0$

   La catena è chiaramente irriducibile e quindi tutti gli stati hanno lo stesso carattere. Per determinare se la catena è persistente o meno applichiamo il criterio 2.3.2 con $j_0 = 0$. Impostando il sistema, si ottiene

$$\begin{cases} x_1 = \frac{1}{2}x_2 \\ \vdots \\ x_k = \frac{1}{2}x_{k-1} + \frac{1}{2}x_{k+1} \\ \vdots \\ x_{-1} = \frac{1}{2}x_{-2} \\ \vdots \\ x_{-k} = \frac{1}{2}x_{-(k-1)} + \frac{1}{2}x_{-(k+1)} \\ \vdots \\ 0 \leq x_i \leq 1 \qquad \forall\, i \in \mathbb{Z} \end{cases}$$

la cui soluzione verifica $x_k = kx_1$ e $x_{-k} = kx_{-1}$ per ogni $k \in \mathbb{N}$, che è limitata per ogni k se e soltanto se $x_k = 0$, $\forall\, k$. L'unica soluzione del sistema è la triviale e quindi la catena è persistente, che implica $P(\tau < +\infty) = 1$.

Per quanto detto precedentemente, $\{S_{n\wedge\tau}\}$ è una martingala e quindi deve avere media costante

$$E(S_{n\wedge\tau}) = E(S_0) = 0.$$

Inoltre, per $n \to +\infty$, $S_{n\wedge\tau} \longrightarrow S_\tau$ q.o., poichè $P(\tau < +\infty) = 1$. Per il teorema della convergenza dominata (ricordiamo che $|S_{n\wedge\tau}| \leq \max(a, b)$), abbiamo $E(S_\tau) = 0$.

Ció implica

$$0 = E(S_\tau) = aP(S_\tau = a) - bP(S_\tau = -b) = ap - b(1-p)$$

da cui $\dfrac{b}{a+b} = P(S_n = a \text{ prima che } S_n = -b)$.

2. Diamo un esempio di una martingala che è limitata in L^1 ma non in L^2.

   Con la stessa notazione dell'esempio precedente, definiamo $S_k = 1 + \sum_{i=1}^{k} X_i$ ed il tempo si arresto $\tau = \inf\{k : S_k = 0\}$. Il processo $\{S_k\}$ è chiaramente una martingala e, come sopra, sappiamo che anche il processo definito tramite $R_n = S_{n\wedge\tau}$ lo è. Ques'ultima martingala per costruzione è non negativa e quindi deve essere anche limitata in L^1. Vogliamo dimostrare che $\{R_n\}$ non è limitata in L^2.

   Siano $M > 0$, $\Gamma^M = \inf\{k : R_k = M\}$ e $Q_k^M = R_{k\wedge\Gamma^M} = S_{k\wedge\tau\wedge\Gamma^M} = S_{k\wedge\tau_{0,M}}$, dove $\tau_{0,M} = \inf\{k : S_k = 0 \text{ oppure } S_k = M\}$.

Come abbiamo visto negli esempi precedenti, anche $S_n^2 - n$ è una martingala, perciò, per la proposizione 5.2.1, per ogni tempo di arresto ψ anche il processo $\{S^2_{n\wedge\psi} - n\wedge\psi\}$ è una martingala, per cui deve avere media costante in n

$$E(S^2_{n\wedge\psi} - n\wedge\psi) = E(S_0^2) = 1, \qquad \text{ovvero} \quad E(S^2_{n\wedge\psi}) = E(n\wedge\psi) + 1.$$

Scegliendo $\psi = \tau \wedge \Gamma^M$ si ottiene

$$E((Q_k^M)^2) = 1 + E(k\wedge\tau\wedge\Gamma^M) \leq 1 + E(k\wedge\tau) = E(R_k^2).$$

Applicando il teorema della convergenza dominata (ricordiamo che per ogni k $(Q_k^M)^2 \leq M^2$), si ha per $k \to +\infty$

$$E((Q_k^M)^2) \longrightarrow E(S^2_{\tau\wedge\Gamma^M}) = M^2 P(S_n = M \text{ prima che } S_n = 0) = M^2\frac{1}{M} = M.$$

Concludendo, per ogni $M > 0$ fissato, $E(R_k^2) \geq E((Q_k^M)^2) \longrightarrow M$. Per l'arbitrarietà di M necessariamente $E(R_k^2) \longrightarrow +\infty$ per $k \to +\infty$.

La proposizione 5.2.1 è una prima versione del noto teorema di Doob(optional sampling theorem), che afferma che una martingala ricampionata secondo una successione di tempi di arresto rimane una martingala. Lo presentiamo ora in generale.

Teorema 5.2.1 : *(Optional Sampling Theorem)*
Sia $\{\tau_n\}$ una successione crescente di tempi di arresto tali che $P(\tau_n < +\infty) = 1$ per ogni n. Siano inoltre $\{X_n\}$ una (sub)martingala e $\{\tilde{X}_n\} = \{X_{\tau_n}\}$. Se

(a) per ogni i, $E(|\tilde{X}_i|) < +\infty$;

(b) $\liminf_{N\to+\infty} \int_{\{\tau_n > N\}} |X_N| dP = 0$ per tutti gli n;

allora $\{X_{\tau_n}\}$ è una martingala.

In sostanza il teorema opzionale afferma che se la successione ha code uniformemente sufficientemente leggere allora la proprietà di martingala è preservata anche su tempi aleatori.

Dimostrazione: Indichiamo con $\tilde{\mathcal{G}}_n = \sigma(\tilde{X}_1, \ldots, \tilde{X}_n)$. Vogliamo dimostrare che per ogni $A \in \tilde{\mathcal{G}}_n$ si ha

$$\int_A \tilde{X}_{n+1} dP (\geq) = \int_A \tilde{X}_n dP. \tag{5.1}$$

Se denotiamo $D_j = A \cap \{\tau_n = j\}$, per disgiunzione $A = \bigcup_j D_j$ ed è quindi sufficiente dimostrare la (5.1) per ogni D_j. Prima mostriamo che $D_j \in \sigma(X_1, \dots, X_j) = \mathcal{G}_j$. Poichè $A \in \tilde{\mathcal{G}}_n$, per qualche $B \in \mathcal{B}(\mathbb{R}^n)$ deve essere $A = \{(\tilde{X}_1, \dots, \tilde{X}_n) \in B\}$, dunque

$$
\begin{aligned}
D_j &= \{(\tilde{X}_1, \dots, \tilde{X}_n) \in B, \tau_n = j\} \\
&= \bigcup_{j_1 \leq \dots \leq j_{n-1} \leq j} \{(\tilde{X}_1, \dots, \tilde{X}_n) \in B, \tau_1 = j_1, \dots, \tau_n = j\} \\
&= \bigcup_{j_1 \leq \dots \leq j_{n-1} \leq j} \{(X_{j_1}, \dots, X_{j_{n-1}}) \in B, \tau_1 = j_1, \dots, \tau_n = j\} \in \mathcal{G}_j.
\end{aligned}
$$

D'altra parte $\int_{D_j} \tilde{X}_n dP = \int_{\{\tau_n = j\} \cap A} X_{\tau_n} dP = \int_{D_j} X_j dP$, quindi per $N > j$ arbitrario, essendo la successione dei t.a. crescente, è possibile scrivere

$$
\begin{aligned}
\int_{D_j} \tilde{X}_{n+1} dP &= \sum_{i=j}^{N} \int_{D_j \cap \{\tau_{n+1} = i\}} \tilde{X}_{n+1} dP + \int_{D_j \cap \{\tau_{n+1} > N\}} \tilde{X}_{n+1} dP \\
&= \sum_{i=j}^{N} \int_{D_j \cap \{\tau_{n+1} = i\}} X_i dP + \int_{D_j \cap \{\tau_{n+1} > N\}} X_N dP \\
&- \int_{D_j \cap \{\tau_{n+1} > N\}} (X_N - \tilde{X}_{n+1}) dP
\end{aligned}
$$

sommando l'ultimo termine della sommatoria ed il secondo addendo, notiamo che

$$
\begin{aligned}
& \int_{D_j \cap \{\tau_{n+1} = N\}} X_N dP + \int_{D_j \cap \{\tau_{n+1} > N\}} X_N dP = \int_{D_j \cap \{\tau_{n+1} geN\}} X_N dP \\
(\geq) \quad = & \int_{D_j \cap \{\tau_{n+1} \geq N\}} X_{N-1} dP = \int_{D_j \cap \{\tau_{n+1} > N-1\}} X_{N-1} dP
\end{aligned}
$$

grazie alla proprietà di (sub)martingala di $\{X_n\}$. Iterando la precedente (dis)uguaglianza si giunge a

$$
\int_{D_j} \tilde{X}_{n+1} dP (\geq) = \int_{D_j \cap \{\tau_{n+1} > N\}} (X_N - \tilde{X}_{n+1}) dP.
$$

D'altronde $D_j \subseteq \{\tau_{n+1} \geq j\}$, quindi possiamo concludere

$$
\int_{D_j} \tilde{X}_{n+1} dP (\geq) = \int_{D_j} \tilde{X}_n dP - \int_{D_j \cap \{\tau_{n+1} > N\}} (X_N - \tilde{X}_{n+1}) dP.
$$

Per $N \to +\infty$, grazie all 'ipotesi (b), sappiamo che $\int_{D_j \cap \{\tau_{n+1} > N\}} X_N dP \longrightarrow 0$. Inoltre, poichè $\{\tau_{n+1} > N\} \downarrow 0$, grazie alla integrabilità della successione, possiamo applicare il teorema della convergenza dominata ed ottenere

$$\int_{D_j \cap \{\tau_{n+1} > N\}} (X_N - \tilde{X}_{n+1}) dP \longrightarrow 0,$$

che conclude la dimostrazione. $\square$

Corollario 5.2.1 : *sotto le stesse ipotesi del teorema 5.2.1 e sotto l'ulteriore condizione*

$$\limsup_n E(|X_n|) < +\infty$$

abbiamo

1. $E(X_1) \leq E(\tilde{X}_n) \leq \limsup_n E(X_n)$ *per ogni* n*;*
2. $E(|\tilde{X}_n|) \leq 2 \limsup_n E(|X_n|) - E(|X_1|)$ *per ogni* n*.*

Dimostrazione: Senza perdere di generalità, si può assumere che $\tau_1 \equiv 1$ (se si avesse $\tau_1 > 1$ sarebbe sufficiente rinumerare la successione imponendo $\tilde{\tau}_1 = 1$ e $\tilde{\tau}_2 = \tau_1 - 1$ etc.

Dall'optional sampling theorem consegue

$$E(\tilde{X}_n) \geq (=) E(\tilde{X}_1), \text{ e } E(\tilde{X}_n) = \lim_{N \to +\infty} \sum_{j=1}^{N} \int_{\{\tau_n = j\}} X_j dP,$$

ma

$$\sum_{j=1}^{N} \int_{\{\tau_n = j\}} X_j dP \leq (=) \sum_{j=1}^{N} \int_{\{\tau_n = j\}} X_N dP = E(X_N) - \int_{\{\tau_n > N\}} X_N dP.$$

Per l'ipotesi (b) del teorema, possiamo scegliere una sottosuccessione dei naturali lungo la quale $\int_{\{\tau_n > N\}} X_N dP \longrightarrow 0$, per $n \to +\infty$, il che implica che

$$E(\tilde{X}_n) \leq \limsup_n E(X_n).$$

Per la seconda parte notiamo che $\{X_n^+\}$ è una submartingala e dunque per il teorema, anche $\{\tilde{X}_n^+\}$ lo deve essere e possiamo analogamente concludere che

$$E(\tilde{X}_n^+) \leq \limsup_n E(X_n^+).$$

Per la successione originale si ottiene

$$\begin{aligned} E(\tilde{X}_n^-) &\leq E(\tilde{X}_n^+) - E(X_1) \\ E(|\tilde{X}_n|) &= E(\tilde{X}_n^+) + E(\tilde{X}_n^-) \leq 2 \limsup_n E(X_n^+) - E(X_1) \\ &\leq 2 \limsup_n E(|X_n|) - E(X_1). \qquad \square \end{aligned}$$

Corollario 5.2.2 : *sotto l'ipotesi (b) del teorema 5.2.1, se* $\text{limsup}_n E(|X_n|) < +\infty$*, allora* $\sup E(|\tilde{X}_n|) < +\infty$*, dunque vale anche l'ipotesi (a) del teorema e le relative conclusioni.*

Dimostrazione: Per $M > 0$ introduciamo i t.a. troncati $\zeta_{n,M} = \tau_n \wedge M$ e la successione $\tilde{X}_{n,M} = X_{\zeta_{n,M}}$. Poichè

$$E(|\tilde{X}_{n,M}|) = \sum_{j=1}^{M} \int_{\{\zeta_{n,M}=j\}} |X_j| dP \leq \sum_{j=1}^{M} E(|X_j|)$$

le condizioni dell'optional sampling theorem sono verificate da τ_n, $\tilde{X}_{n,M}$ e per il precedente Corollario esiste α indipendente da n e da M tale che $E(|\tilde{X}_{n,M}|) \leq \alpha$.

D'altra parte$\zeta_{n,M} \longrightarrow \tau_n$ per $M \to +\infty$, perciò $\tilde{X}_{n,M} \longrightarrow \tilde{X}_n$ e per il lemma di Fatou $E(|\tilde{X}_n|) \leq \alpha$. □

5.3 Teoremi di convergenza per le martingale

In questa sezione presentiamo alcuni risultati di convergenza delle martingale, utili anche per trasferire al continuo tante delle proprietà studiate nel contesto discreto.

Teorema 5.3.1 : *Versione debole della disuguaglianza massimale di Doob.*

Sia $\{X_n\}$ *una (sub)martingala e* $\lambda \in \mathbb{R}^+$ *e denotiamo con* $X^* = \sup_{k\geq 0} |X_k|$. *Allora vale*

$$P(X^* > \lambda) \leq \frac{||X||_1}{\lambda} = \frac{\lim_{n\to+\infty} E(|X_n|)}{\lambda} \tag{5.2}$$

Dimostrazione: Senza perdere di generalità, possiamo restringerci al caso di martingale non negative, applicando altrimenti la dimostrazione a $\{|X_n|\}$. Sia τ il t.a. così definito

$$\tau = \inf\{k \, : \, X_k > \lambda\}.$$

Vogliamo dimostrare che $E(X_{\tau\wedge n}) \leq E(X_n)$ per ogni n.

Fissato n, notiamo che per $k < n$ si ha

$$\begin{aligned} E(X_n \mathbf{1}_{\{\tau=k\}}) &= E(E(X_n \mathbf{1}_{\{\tau=k\}}|\mathcal{G}_k)) = E(\mathbf{1}_{\{\tau=k\}} E(X_n|\mathcal{G}_k)) \\ (\geq) &= E(\mathbf{1}_{\{\tau=k\}} X_k) = E(\mathbf{1}_{\{\tau=k\}} X_\tau). \end{aligned}$$

Sommando k da 1 ad $n-1$, otteniamo $E(X_n \mathbf{1}_{\{\tau<n\}})(\geq) = E(X_\tau \mathbf{1}_{\{\tau<n\}})$, da cui

$$\begin{aligned} E(X_n) &= E(X_n \mathbf{1}_{\{\tau\geq n\}}) + E(X_n \mathbf{1}_{\{\tau<n\}})(\geq) = E(X_{\tau\wedge n} \mathbf{1}_{\{\tau\geq n\}}) + E(X_\tau \mathbf{1}_{\{\tau<n\}}) \\ &= E(X_{\tau\wedge n} \mathbf{1}_{\{\tau\geq n\}}) + E(X_{\tau\wedge n} \mathbf{1}_{\{\tau<n\}}) = E(X_{\tau\wedge n}). \end{aligned}$$

Per definizione di τ abbiamo

$$E(X_{\tau\wedge n}) \geq \lambda P(X_{\tau\wedge n} > \lambda) = \lambda P(X_n^* > \lambda),$$

dove con X_n^* abbiamo indicato $\sup\limits_{k\leq n} X_k$. Dunque per $\to +\infty$ possiamo concludere

$$\frac{\lim_{n\to+\infty} E(X_n)}{\lambda} \geq \lim_{n\to+\infty} P(X_n^* > \lambda) = P(X^* > \lambda). \qquad \square$$

Per dimostrare il teorema di convergenza per le martingale ci serviremo dell'approccio di Doob che va a verificare che il numero di attraversamenti di ogni striscia si riduce a zero se il tempo è spinto ad infinito.

Consideriamo un intervallo fissato della retta reale $[\alpha, \beta]$ e si a$\{\zeta_n\}$ una successione di numeri reali. Definiamo, con la convenzione che $\inf \emptyset = \infty$, le quantità

$$\begin{aligned}
\tau_1 &= \inf\{k : \zeta_k \leq \alpha\} \\
\tau_2 &= \inf\{k > \tau_1 : \zeta_k \geq \beta\} \\
\tau_3 &= \inf\{k > \tau_2 : \zeta_k \leq \alpha\} \\
\tau_2 &= \inf\{k > \tau_3 : \zeta_k \geq \beta\} \\
&\vdots
\end{aligned}$$

ed il numero di attraversamenti (dal basso verso l'alto) dell'intervallo $[\alpha, \beta]$ come $U(\alpha, \beta) = \max\{k : \tau_{2k} < +\infty\}$. Allora la successione $\{\zeta_n\}$ converge se e solo se

1. $\limsup\limits_n |\zeta_n| < +\infty$;

2. $U(\alpha, \beta) < +\infty$ per ogni $\alpha, \beta \in \mathbb{Q}$.

Proposizione 5.3.1 : *Una successione di v.a.* $\{X_n\}$ *converge q.o. se*

1. $P(\limsup\limits_n |X_n| < +\infty) = 1$ *oppure* $\sup_i |X_i| < +\infty$ *q.o.;*

2. $P(U_X(\alpha, \beta) < +\infty) = 1$ *per ogni* $\alpha, \beta \in \mathbb{Q}$.

Dimostrazione: Sul complementare di $\bigcup\limits_{\alpha,\beta\in\mathbb{Q}} \{U_X(\alpha, \beta) = +\infty\} cup \{\limsup |X_n| = +\infty\}$, per quanto detto prima, la successione numerica $\{X_n(\omega)\}$ converge per quasi ogni $\omega \in \Omega$ fissato. D'altra parte, per le ipotesi, questo insieme risulta essere unione numerabile di insiemi d probabilità nulla e quindi di probabilità nulla anch'esso.

$\square$

Lemma 5.3.1 : *Upcrossing lemma*

Sia $\{X_n\}$ una martingala e siano $\alpha < \beta$ due numeri reali. Allora

$$E(U_X(\alpha,\beta)) \leq \frac{\sup_i E((X_i-\alpha)^+)}{\beta-\alpha}$$

Dimostrazione: senza perdere di generalità possiamo dimostrare il teorema per $\alpha = 0$ e $\beta = 1$, potendo considerare altrimenti la successione $\dfrac{X_n-\alpha}{\beta-\alpha}$.

Prima di tutto notiamo che se τ è un t.a. e $\{Y_n\}$ una submartingala, allora

$$E(Y_\tau \mathbf{1}_{\{\tau\leq n\}}) \leq E(Y_n \mathbf{1}_{\{\tau\leq n\}}). \tag{5.3}$$

Infatti per n fissato e $k \leq n$, abbiamo

$$\begin{aligned} E(Y_n\mathbf{1}_{\{\tau=k\}}) &= E(E(Y_n\mathbf{1}_{\{\tau=k\}}|\mathcal{G}_k)) = E(\mathbf{1}_{\{\tau=k\}}E(|\mathcal{G}_k)) \\ &\geq E(\mathbf{1}_{\{\tau=k\}}Y_k) = E(\mathbf{1}_{\{\tau=k\}}Y_\tau) \end{aligned}$$

grazie alla proprietà di submartingala. Se sommiamo l'ultima disuguaglianza per $k = 1,\ldots,n$ otteniamo la (5.3), avendo l'uguaglianza nel caso di martingala.

Come secondo passo, ricordiamo che sia $\{X_n^+\}$ sia $\{|X_n|\}$ sono submartingale grazie alla disuguaglianza di Jensen. Definiamo dunque, con la convenzione che $\inf\emptyset = +\infty$, i seguenti tempi d'arresto

$$\begin{aligned} \tau_1 &= \inf\{k>0 \,:\, X_k \leq 0\} \\ \tau_2 &= \inf\{k>\tau_1 \,:\, X_k \geq 1\} \\ \tau_3 &= \inf\{k>tau_2 \,:\, X_k \leq 0\} \\ \tau_4 &= \inf\{k>\tau_3 \,:\, X_k \geq 1\} \\ &\vdots \end{aligned}$$

Allor aper ogni n fissato possiamo applicare l'optional sampling theorem alla successione crescente di t.a. <u>limitati</u> da n definiti come $\sigma_i = \tau_{2i+1}\wedge n$, per $i \in \mathbb{N}$, infatti

$$\begin{aligned} E(|X_{\sigma_i}|) &= E(|X_{\tau_{2i+1}}|) = \sum_{j=1}^{n-1} E(|X_j|\mathbf{1}_{\{\tau_{2i+1}=j\}}) + E(|X_n|\mathbf{1}_{\{\tau_{2i+1}\geq n\}} \\ &\leq \sum_{j=1}^{n} E(|X_j|) < +\infty. \end{aligned}$$

Inoltre per ogni $N > n$ si ha $\{\sigma_i > N\} = \emptyset$ e quindi anche la seconda condizione del teorema è automaticamente verificata. Dunque $\{X_i\}$ è unamartingala rispetto alla sua filtrazione naturale $\tilde{\mathcal{F}}_i = \sigma(X_{\sigma_1}, \ldots, X_{\sigma_i}) \subseteq \mathcal{F} - n = \sigma(X_1, \ldots, X_n)$ per ogni k, che implica

$$E(X_n|\tilde{\mathcal{F}}_i) = X_{\sigma_i}, \qquad \forall\, i.$$

Per definizione di t.a., l'evento $\{\sigma_i \leq n\} = \{\tau_{2i+1} \leq n\} \in \tilde{\mathcal{F}}_i$, che implica

$$\int_{\{\tau_{2i+1} \leq n\}} X_n dP = \int_{\{\tau_{2i+1} \leq n\}} X_{\tau_{2i+1}} dP \leq 0. \tag{5.4}$$

Analogamente vale

$$\int_{\{\tau_{2i} \leq n\}} X_n dP \geq 1 P(\tau_{2i} \leq n), \tag{5.5}$$

che insieme implicano

$$\int_{\{\tau_{2i} \leq n\} - \{\tau_{2i+1} \leq n\}} X_n dP \geq P(\tau_{2i} \leq n).$$

Gli insiemi $A_i = \{\tau_{2i} \leq n\} - \{\tau_{2i+1} \leq n\}$ sono disgiunti, dunque

$$\int_{\bigcup_i A_i} X_n dP \geq \sum_{i=1}^{+\infty} P(\tau_{2i} \leq n).$$

D'altra parte

$$\{\tau_{2i} \leq n\} = \{\text{almeno } i \text{ attraversamenti fino al tempo } n\} \to \{U_X(0,1) \geq i\}.$$

Sull'insieme $\bigcup\limits_i A_i$ si ha $X_n \equiv X_n^+$, quindi spingendo $n \to +\infty$ possiamo concludere

$$\int_\Omega X_n dP \geq \int_{\bigcup_i A_i} X_n^+ dP \leq \sum_{i=1}^{+\infty} P(\tau_{2i} \leq n) \to \sum_{i=1}^{+\infty} P(U_X(0,1) \geq i) = E(U_X(0,1)).$$

Prendendo l'estremo superiore di entrambi i membri della precedente relazione e notando che, per monotonia $\limsup E(X_n^+) = \sup E(X_n^+)$, otteniamo la tesi. $\square$

Lemma 5.3.2 : *Sia X una v.a. discreta non negativa ed integrabile, allora*

$$E(X) = \sum_{i=1}^{+\infty} P(X \geq i)$$

Dimostrazione:

$$E(X)=\sum_{i=1}^{+\infty} iP(X=i)=\lim_{k\to\infty}\sum_{i=1}^{k} iP(X=i)=\lim_{k\to\infty}\sum_{i=1}^{k} i[P(X\geq i)-P(X\geq i+1)]$$
$$=\lim_{k\to\infty}[P(X\geq 1)+\sum_{i=2}^{k} iP(X\geq i)-\sum_{i=1}^{k} iP(X\geq i+1)]$$
$$=\lim_{k\to\infty}[P(X\geq 1)+\sum_{i=2}^{k} iP(X\geq i)-\sum_{i=1}^{k}(i-1)P(X\geq i)]$$
$$=\lim_{k\to\infty}[P(X\geq 1)+\sum_{i=2}^{k} P(X\geq i)-kP(X\geq k+1)]$$
$$=\lim_{k\to\infty}[P(X\geq 1)+\sum_{i=2}^{k} P(X\geq i)]-\lim_{k\to\infty}P(X\geq k+1)] \quad =\sum_{i=1}^{+\infty}P(x\geq i) \quad \square$$

Teorema 5.3.2 : *Teorema di convergenza per le martingale.*
Sia $\{X_n\}$ una martingala limitata in L^1, allora X_n converge q.o. per $n\to+\infty$.

Dimostrazione: Grazie alla versione debole della disuguaglianza di Doob, abbiamo

$$P(\sup_i|X_i|>\lambda)\leq\frac{\sup_i E(|X_i|)}{\lambda}\geq\frac{M}{\lambda}\longrightarrow 0,\qquad \text{per }\ \lambda\to+\infty,$$

poichè la martingala è limitata, dunque la prima condizione della proposizione 5.3.1 è verificata. Dall'upcrossing lemma otteniamo che per qualsiasi coppia di razionali $\alpha<\beta$

$$P(U_X(\alpha,\beta)>\lambda)\leq\frac{E(U_X(\alpha,\beta))}{\lambda}\leq\frac{\sup_i E((X_i-\alpha)^+)}{\lambda(\beta-\alpha)}\leq\frac{M}{\lambda}$$

che tende a 0 per $\lambda\to+\infty$ e quindi anche la seconda condizione della proposizione è soddisfatta. $\square$

Definizione 5.3.1 : *Una successione di v.a. $\{Y_n\}$ si dice* **uniformemente integrable** *(u.i.) se per ogni $\epsilon>0$ esiste una costante $M>0$ tale che*

$$\int_{\{|Y_i|\geq M\}}|Y_i|dP<\epsilon,\qquad \forall\, i\in\mathbb{N}.$$

Osservazione 5.3.1 : *Siano $\{\mathcal{A}_i\}$ una filtrazione, $\mathcal{A} = \sigma(\bigcup_i \mathcal{A}_i)$ ed X una v.a. integrabile ed $\mathcal{A}-$misurabile. La successione definita da $X_n = E(X|\mathcal{A}_n)$ è una martingala uniformemente integrabile.*

Infatti notiamo che essendo X integrabile, si ha

$$\begin{aligned} &P(|X| > M) \leq \frac{E(|X|)}{M} \longrightarrow 0, \qquad \textit{per } M \to +\infty \\ &|X|\mathbf{1}_{\{|X|>M\}} \longrightarrow 0 \quad P \textit{ q.o. per } M \to +\infty. \end{aligned}$$

Per ogni M, d'altra parte $|X|\mathbf{1}_{\{|X|>M\}} \leq |X| \in L^1$ e quindi per il teorema della convegenza dominata

$$E(|X|\mathbf{1}_{\{|X|>M\}}) \longrightarrow 0 \qquad \textit{per } M \to +\infty$$

ovvero scelto $\epsilon > 0$ si può selezionare M così grande che

$$\int_{\{|X|>M\}} |X| dP < \epsilon.$$

Allora, fissato M e ponendo

$$X_n = E(X|\mathcal{A}_n) = E(X\mathbf{1}_{\{|X|>M\}}|\mathcal{A}_n) + E(X\mathbf{1}_{\{|X|\leq M\}}|\mathcal{A}_n) := Y_n + Z_n,$$

abbiamo $Z_n \leq M$ e

$$E(|Y_n|) = E(|E(X\mathbf{1}_{\{|X|>M\}}|\mathcal{A}_n)|) \leq E(E(|X|\mathbf{1}_{\{|X|>M\}}|\mathcal{A}_n)) = E(|X|\mathbf{1}_{\{|X|>M\}}) < \epsilon,$$

mentre utilizzando la disuguaglianza di Jensen per qualsiasi $\alpha \in \mathbb{R}$ si ha la seguente inclusione di insiemi

$$\{|X_n| > \alpha\} \subseteq \{E(|X|\,|\mathcal{A}) > \alpha\},$$

perciò scelto $\Gamma > M$ si ottiene

$$\begin{aligned} E(|Z_n|\mathbf{1}_{\{|X_n|>\Gamma\}} &= E(|E(X\mathbf{1}_{\{|X|\leq M\}}|\mathcal{A}_n)|\mathbf{1}_{\{|X_n|>\Gamma\}}) \\ &\leq E(E(|X|\mathbf{1}_{\{|X|\leq M\}}|\mathcal{A}_n)\mathbf{1}_{\{|X_n|>\Gamma\}}) \\ &\leq E(E(|X|\mathbf{1}_{\{E(|X|\,|\mathcal{A}_n)\leq M\}}|\mathcal{A}_n)|\mathbf{1}_{\{|X_n|>\Gamma\}}) \\ &\leq E(E(|X|\,|\mathcal{A}_n)\mathbf{1}_{\{E(|X|\,|\mathcal{A}_n)\leq M\}}\mathbf{1}_{\{E(|X|\,|\mathcal{A}_n)>\Gamma\}}) = 0 \end{aligned}$$

e concludendo

$$\int_{\{|X_n|>\Gamma\}} |X_n| dP \leq \int_{\{|X_n|>\Gamma\}} |Y_n| dP + \int_{\{|X_n|>\Gamma\}} |Z_n| dP \leq \int_\Omega |Y_n| dP + 0 < \epsilon$$

ovvero l'uniforme integrabilità.

Diamo ora una condizione sufficiente per l'u.i.

Proposizione 5.3.2 : *Sia $\{X_n\}_n$ una successione di v.a. tali che per qualche $\epsilon > 0$ vale*

$$\sup_n E(|X_n|^{1+\epsilon}) < +\infty,$$

allora$\{X_n\}_n$ è uniformemente integrabile.

Dimostrazione:

$$\int_{\{|X_n|>M\}} |X_n| dP = \int_{\{|X_n|>M\}} \frac{|X_n|^{1+\epsilon}}{|X_n|^\epsilon} dP \leq \frac{1}{M^\epsilon} \int_{\{|X_n|>M\}} |X_n|^{1+\epsilon} dP \leq \frac{E(|X_n|^{1+\epsilon})}{M^\epsilon} \to 0$$

per $M \to +\infty$. □

Infine vale un teorema di passaggio al limite sotto il segno di integrale.

Teorema 5.3.3 : *Sia $\{X_n\}_n$ una successione di v.a. uniformemente integrabili convergente $P-$q.o. ad una v.a. X.*

Allora X è integrabile e $E(X_n) \longrightarrow E(X)$.

Dimostrazione: Grazie all'uniforme integrabilità, notiamo che scelto α sufficientemente grande, sfruttando l'u.i. si ha

$$E(|X_n|) = \int_{\{|X_n|>\alpha\}} |X_n| dP + \int_{\{|X_n| \leq \alpha\}} |X_n| dP \leq \epsilon + \alpha,$$

dunque applicando il Lemma di Fatou, otteniamo

$$\int |X| dP = \int \liminf |X_n| dP \leq \liminf \int |X_n dP \leq \alpha + \epsilon < +\infty.$$

Denotiamo $Y_n = |X_n - X|$. Per ipotesi

$$Y_n \longrightarrow 0 \quad P - \text{q.o.}$$

ed è una successione u.i., poichè

$$\int_{\{|Y_n|>\alpha\}} |Y_n| dP \leq \int_{\{|X_n|+|X|>\alpha\}} (|X_n| + |X|) dP \leq \int_{\{|X_n|>\frac{\alpha}{2}\}} |X_n| dP + \int_{\{|X|>\frac{\alpha}{2}\}} |X| dP \leq \epsilon.$$

Inoltre per il teorema della convergenza dominata, $\lim_{n\to+\infty} Y_n \mathbf{1}_{\{Y_n \leq \alpha\}} = 0$ implica $E(Y_n \mathbf{1}_{\{Y_n \leq \alpha\}}) \longrightarrow 0$.

Dunque possiamo concludere che per n sufficientemente grande la seguente stima è valida

$$\int Y_n dP = \int_{\{Y_n \le \alpha\}} Y_n dP + \int_{\{Y_n > \alpha\}} Y_n dP < 2\epsilon,$$

da cui si ottiene la tesi, grazie all'arbitrarietà di ϵ. $\square$

In conclusione possiamo presentare il seguente

Teorema 5.3.4 : *$\{X_n\}_n$ è una martingala u.i. se e solo se esiste una v.a. X integrabile e misurabile rispetto alla $\sigma-$algebra $\mathcal{G} = \sigma(X_1, X_2, \dots)$, tale che*

$$E(X|\mathcal{G}_n) = X_n. \tag{5.6}$$

Dimostrazione: Un verso della dimostrazione è ovvio. Per il viceversa, ricordiamo che se $\{X_n\}_n$ è u.i., allora è anche limitata in L^1. Poichè la successione costituisce una martingala, dal teorema di convergenza per le martingale possiamo concludere che $X_n \longrightarrow X$ $P-$q.o., per qualche v.a. X integrabile (grazie al Lemma di Fatou) e $\mathcal{G}-$misurabile. Vogliamo dimostrare la (5.6).

Per ogni insiemem $A \in \mathcal{G}_n = \sigma(X_1, \dots, X_n)$, per definizione di media condizionata, si ha

$$\int_A E(X|\mathcal{G}_n) dP = \int_A X dP = \int_\Omega \mathbf{1}_A X dP.$$

Inoltre per ogni n fissato, come conseguenza della convergenza anche $X_{n+k}\mathbf{1}_A \longrightarrow X\mathbf{1}_A$ $P-$ q.o. per $k \to +\infty$. Grazie all'uniforme integrabilità possiamo passare al limite sotto il segni di media ed affermare

$$E(X_{n+k}\mathbf{1}_A) \longrightarrow E(X\mathbf{1}_A), \qquad k \longrightarrow +\infty.$$

Per la proprietà di martingala, $E(X_{n+k}\mathbf{1}_A) = E(X_n\mathbf{1}_A)$, dunque per ogni $A \in \mathcal{G}_n$,

$$E(X\mathbf{1}_A) = \lim_{k\to+\infty} E(X_{n+k}\mathbf{1}_A) = E(X_n\mathbf{1}_A)$$

ovvero la (5.6). $\square$

Bibliografia

[Ba] P. Baldi “Calcolo delle Probabilità e Statistica ed. McGraw-Hill.

[B] P. Billingsley Probability and measure ed. John Wiley 1984.