

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Matematica

Gioco, Partita e Incontro:
Modellizzazione Markoviana del Tennis

Tesi di Laurea in Probabilità

Relatore:
Chiar.mo Prof.
Andrea Pascucci

Presentata da:
Antonio Di Feo

Anno Accademico 2024/2025

a Rossana

Introduzione

La teoria della probabilità e i processi stocastici offrono strumenti matematici molto potenti per descrivere e prevedere l'evoluzione di sistemi complessi nel tempo. Anche negli sport, infatti, questi strumenti ricoprono un ruolo centrale per gli addetti ai lavori: basti pensare in primo luogo alle scommesse sportive, in cui ogni quota consultabile sui portali di settore è data da un differente algoritmo basato su vari processi aleatori. Inoltre, anche gli stessi sportivi e le loro squadre di analisti utilizzano abitualmente questi processi. Ad esempio, nel calcio, negli ultimi anni vengono spesso consultati gli *expected goals* (xG) di una squadra in una determinata partita. Essi non sono altro che la somma delle probabilità che ha avuto ogni tiro effettuato di entrare in porta, ossia il valore atteso (grazie alla linearità di quest'ultimo) della variabile aleatoria indicante il numero di goal totali segnati in quel determinato incontro.

Il calcio però è uno sport caratterizzato da un flusso continuo, ovvero non esistono “stati” rigidi indipendenti. Il tennis invece (come il ping pong oppure la pallavolo) è uno sport composto da “punti”, ovvero situazioni teoricamente indipendenti dove solo il giocatore (o la squadra) 1 o il giocatore 2 conquista quella particolare fase dell'incontro. Questa struttura si presta molto bene al fine di vedere in modo ancor più approfondito tale attività come una catena di eventi, che è proprio ciò che abbiamo studiato in questo lavoro. Il punteggio particolare dei *game* di tennis (così come quello dei *set*), infatti, si presta bene per essere osservato come una successione di esiti indipendenti l'uno dall'altro. È chiaro tuttavia che, nella pratica reale, tale indipendenza non è rigorosamente verificata. Infatti se un giocatore (specialmente se di livello non professionistico) si trovasse ad affrontare un punto particolarmente lungo e faticoso per lui, è piuttosto normale che non solo nel punto successivo, ma anche in quelli a seguire, la sua forza e agilità ne risentano. Anche per questo motivo, in gran parte del nostro studio abbiamo separato il suddetto game in due fasi (alle quali abbiamo attribuito probabilità diverse): una prima fase dall'inizio del game fino al punteggio di *parità* (verrà spiegato nelle sezioni di seguito) e una seconda parte che porta fino alla fine del game. Questo è stato fatto soprattutto per tenere conto del fattore psicologico, che in uno sport come il tennis è

particolarmente importante. Infatti nell'eventuale fase del game che va dal punteggio di *parità* in avanti, la tensione del momento può farsi sentire anche in giocatori più esperti, specialmente negli incontri importanti.

Grazie a queste premesse, nel nostro studio un game di tennis è stato modellizzato tramite una *catena di Markov* a 17 stati, con p e \bar{p} costruite come illustrato in precedenza. Sono state quindi sfruttate le proprietà matematiche delle catene di Markov finite per calcolare le cosiddette probabilità di assorbimento, al fine di ricavare rigorosamente le probabilità di vittoria del gioco. Inoltre si è determinato il numero atteso di punti disputati per stimare la durata media di un singolo game. Lo stesso tipo di lavoro è stato svolto con un set di tennis. È stata quindi trovata un'estensione del modello ottenuto in modo che anche in questo macrosettore della partita (tramite combinazioni di probabilità e valori attesi trovati in precedenza per i game) si possa giungere a dei risultati validi di probabilità e di durata anche per il set. È stata inoltre dedicata una parte del nostro lavoro al *tiebreak*, un particolare game che si disputa per decidere le sorti del set in caso di situazione di parità. A causa della peculiare regola sull'alternanza frequente dei turni di servizio però, la modellizzazione di questa fase ha richiesto la costruzione di uno spazio degli stati che è risultato significativamente più esteso e complesso.

Infine, la validità del modello qui esposto è stata confrontata con dei dati reali ricavati da tutti gli incontri disputatisi tra tre giocatori fra i più importanti della storia di questo sport: Rafael Nadal, Roger Federer e Novak Djokovic. L'elevato numero di partite disputate tra questi campioni e il loro altissimo livello di gioco (che tende anche a mitigare la variabilità psicologica) offrono infatti una base statistica solida per effettuare un confronto piuttosto rigoroso. L'analisi ha evidenziato una significativa convergenza tra i risultati del nostro modello, ottimizzato tramite un'opportuna scelta dei parametri, e i dati reali degli incontri. Il lavoro ha preso anche le mosse dal recente studio di Carrari et al. [3].

L'elaborato è strutturato come di seguito. Nel *Capitolo 1* verrà introdotto il concetto di catena di Markov illustrando alcune proprietà utili al nostro lavoro. Inoltre verrà descritto anche il sistema di punteggio del tennis. Nel *Capitolo 2* sarà illustrato il fulcro del nostro modello stocastico calcolando le probabilità di vittoria e le durate attese del game, del set e anche del caso particolare del tiebreak. Infine nel *Capitolo 3* saranno utilizzati come base di validazione i dati reali dei campioni precedentemente nominati e verranno confrontati con le applicazioni della nostra teoria.

Indice

Introduzione	i
1 Preliminari e notazioni	1
1.1 Le catene di Markov in tempi discreti	1
1.2 Proprietà	2
1.2.1 “Hitting time” e proprietà di assorbimento	4
1.3 Il tennis e il suo punteggio	6
2 Studio delle probabilità di vittoria e della lunghezza di un match	9
2.1 Probabilità di vittoria di un game	9
2.2 Durata di un game	13
2.3 Probabilità di vittoria e durata di un set	17
2.3.1 Il tiebreak	17
2.3.2 Probabilità di vittoria di un set	19
2.3.3 Durata di un set	22
3 Confronto con esempi reali	27
3.1 Analisi degli scontri diretti tra Nadal, Djokovic e Federer	27
3.2 Valore atteso della durata degli incontri e confronti	29
3.3 Commenti e conclusioni	30
Bibliografia	33

Capitolo 1

Preliminari e notazioni

1.1 Le catene di Markov in tempi discreti

Sia $\{X_i\}_{0 \leq i \leq N}$ un insieme finito di variabili aleatorie¹ $X_i : \Omega \rightarrow I_n$. Definiamo un elemento particolarmente utile per la teoria successiva.

Definizione 1.1 (Matrice di transizione [1]). Sia $I := \{i_1, \dots, i_N\}$ un insieme finito di stati di probabilità. Sia $N := |I|$ e sia $M \in \mathcal{M}_{N,N}(\mathbb{R})$ una matrice $N \times N$ a coefficienti reali. Si dice che M è una matrice di transizione (o matrice stocastica) se ogni vettore riga di M è una distribuzione di probabilità, ovvero se $m_{ij} \in [0, 1]$ per ogni $i, j \in I$ e $\sum_{j=1}^N m_{ij} = 1$ per ogni $i = 1, \dots, N$ [1].

Le matrici di transizione sono alla base della teoria successiva. Infatti ogni loro riga può essere vista come una distribuzione di probabilità associata a una diversa variabile aleatoria. Di seguito, con le catene di Markov, si può apprezzare un caso particolare di queste matrici molto utile ai nostri fini.

Definizione 1.2 (Catena di Markov finita [1]). Un insieme $\{X_i\}_{0 \leq i \leq N}$ di variabili aleatorie si dice Catena di Markov finita con distribuzione iniziale λ e matrice di transizione P se valgono entrambe le condizioni seguenti:

(i) a X_0 è associata la distribuzione λ ;

(ii) per ogni $n \geq 0$, la distribuzione di X_n condizionata a tutte le $X_s = i_s$, per ogni $s = 0, \dots, n-1$, dipende solo dal valore $X_{n-1} = i_{n-1}$; ovvero ogni X_n dipende soltanto dal valore della variabile aleatoria precedente e ha quindi distribuzione² associatagli $\{p_{nj} \mid j \in I\}$ (questa è detta proprietà Markoviana o di perdita di memoria).

¹ $I_n := \{r \in \mathbb{N} \mid 1 \leq r \leq n\}$.

² p_{ab} è la componente corrispondente alla a-esima riga e alla b-esima colonna della matrice P .

In altre parole, data la distribuzione $\lambda = (\lambda_1, \dots, \lambda_N)$ associata a X_0 , deve valere:

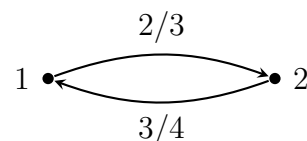
$$(1) \mathbb{P}(X_0 = i_k) = \lambda_k \text{ per ogni } k = 1, \dots, N;$$

$$(2) \forall n = 0, \dots, N-1 \quad \mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_k, \dots, X_n = i_r) = p_{i_r i_{n+1}} \\ \forall i_k, \dots, i_r \in I$$

Si dice che $(X_n)_{0 \leq n \leq N}$ è Markov(λ, P).

Per ogni catena di Markov finita, oltre ad una matrice di transizione, può essere associato un diagramma con delle frecce. Vediamo due esempi:

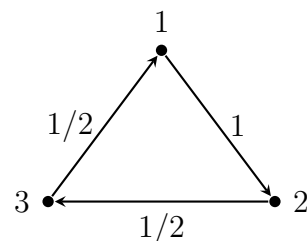
Esempio 1.
$$P = \begin{pmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{pmatrix}$$



In questo caso, leggendo la matrice P , la prima riga ci indica le probabilità degli eventi possibili sapendo di trovarci nel caso 1. La seconda riga, invece, indica le probabilità sapendo di provenire dall'evento 2.

Idem per il seguente esempio 2, con l'aggiunta di una terza riga e quindi di un terzo evento associatogli.

Esempio 2.
$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$



1.2 Proprietà

Vediamo ora alcune proprietà utili delle catene di Markov finite.

Teorema 1.3. *Un processo di variabili aleatorie $(X_n)_{0 \leq n \leq N}$ è un Markov(λ, T) se e solo se per ogni $i_0, \dots, i_N \in I$ vale che*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N) = \lambda_{i_0} t_{i_0 i_1} t_{i_1 i_2} \dots t_{i_{N-1} i_N}. \quad (1.1)$$

Di seguito si enuncerà un altro teorema che fa intuire ancora meglio il concetto di *perdita di memoria*. Prima, però, introduciamo il vettore $\delta_i := (\delta_{ij} : j \in I)$ per indicare la *massa unitaria* in i . Essa è la distribuzione di probabilità banale sullo stato i , ovvero:

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{altrimenti} \end{cases}$$

Teorema 1.4 (Proprietà di Markov). Sia $(X_n)_{0 \leq n \leq N}$ un Markov (λ, T) . Allora, sapendo che $X_m = i$, vale che $(X_{m+n})_{0 \leq n \leq N}$ è un Markov (δ_i, T) ed esso è indipendente dalle variabili aleatorie X_0, \dots, X_m .

Dimostrazione. Bisogna far vedere che, per ogni evento A determinato da X_0, \dots, X_m , si ha che

$$\begin{aligned} & \mathbb{P}(\{X_m = i_m, \dots, X_{m+n} = i_{m+n}\} \cap A \mid X_m = i) \\ &= \delta_{ii_m} t_{i_m i_{m+1}} \cdots t_{i_{m+n-1} i_{m+n}} \mathbb{P}(A \mid X_m = i), \end{aligned} \quad (1.2)$$

poi la tesi seguirà dal Teorema 1.3. Consideriamo prima il caso di un evento elementare

$$E = \{X_0 = i_1, \dots, X_m = i_m\}.$$

Qui, quindi, bisogna far vedere che

$$\begin{aligned} & \frac{\mathbb{P}(X_0 = i_0, \dots, X_{m+n} = i_{m+n}, i = i_m)}{\mathbb{P}(X_m = i)} \\ &= \delta_{ii_m} t_{i_m i_{m+1}} \cdots t_{i_{m+n-1} i_{m+n}} \frac{\mathbb{P}(X_0 = i_0, \dots, X_m = i_m, i = i_m)}{\mathbb{P}(X_m = i)} \end{aligned}$$

che è vero nuovamente grazie al Teorema 1.3. In generale, allora, ogni evento A determinato da X_0, \dots, X_m può essere scritto come unione disgiunta e numerabile di eventi elementari E_k come appena visti:

$$A = \bigcup_{k=1}^{\infty} E_k.$$

Quindi l'uguaglianza (1.2) segue sommando le corrispettive identità di ogni E_k . \square

Un'ultima proprietà interessante riguarda le matrici di transizione. Grazie al prossimo risultato infatti, si possono sfruttare le operazioni delle matrici (elevamento a potenza in particolare) per ottenere informazioni sulle catene di Markov associategli.

Teorema 1.5. Sia $(X_n)_{0 \leq n \leq N}$ un Markov (λ, T) . Allora, per ogni $0 \leq n, m \leq N$, vale che: ³

$$(i) \quad \mathbb{P}(X_n = j) = (\lambda^\top T^n)_j;$$

$$(ii) \quad \mathbb{P}_i(X_n = j) = \mathbb{P}(X_{n+m} = j \mid X_m = i) = t_{ij}^{(n)}.$$

Dimostrazione. (i) Per il Teorema 1.3 vale che:

$$\begin{aligned} \mathbb{P}(X_n = j) &= \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} \mathbb{P}(X_0 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} \lambda_{i_1} t_{i_1 i_2} \cdots t_{i_{n-1} j} = (\lambda^\top T^n)_j. \end{aligned}$$

³ $t_{ij}^{(n)}$ è la componente di posto i, j della matrice T^n .

- (ii) Grazie alla proprietà di Markov (vedasi Teorema 1.4), sapendo che $X_m = i$, si ha che $(X_{m+n})_{n \geq 0}$ è un $Markov(\delta_i, T)$. Quindi, ponendo $\lambda = \delta_i$ nel punto (i), vale lo stesso ragionamento fatto per X_n .

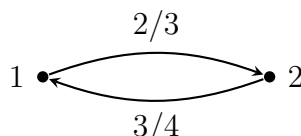
□

Vediamo ora ancora come poter applicare questo teorema al caso visto nell'Esempio 1.

Esempio 3. Torniamo quindi a

$$T = \begin{pmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{pmatrix}$$

associata al diagramma:



Ora, sfruttando l'uguaglianza $T^{n+1} = T^n T$, si può scrivere:

$$t_{11}^{(n+1)} = t_{11}^{(n)} t_{11} + t_{12}^{(n)} t_{21}.$$

Quindi, sostituendo con i valori di T dati:

$$t_{11}^{(n+1)} = t_{11}^{(n)} \frac{1}{3} + t_{12}^{(n)} \frac{3}{4}.$$

Inoltre, sappiamo che $t_{11}^{(n)} + t_{12}^{(n)} = \mathbb{P}_1(X_n = 1 \text{ oppure } 2) = 1$. Quindi, sostituendo $t_{12}^{(n)}$ nell'uguaglianza precedente, si ottiene una ricorrenza per $t_{11}^{(n)}$:

$$\begin{aligned} t_{11}^{(n+1)} &= t_{11}^{(n)} \frac{1}{3} + (1 - t_{11}^{(n)}) \frac{3}{4} \\ &= t_{11}^{(n)} \left(\frac{1}{3} - \frac{3}{4} \right) + \frac{3}{4} \\ &= -\frac{5}{12} t_{11}^{(n)} + \frac{3}{4}. \end{aligned}$$

1.2.1 “Hitting time” e proprietà di assorbimento

Definizione 1.6. Sia $(X_n)_{0 \leq n \leq N}$ una catena di Markov con matrice di transizione T . L'hitting time (o tempo di primo ingresso) di un sottoinsieme A di I è la variabile aleatoria $H^A : \Omega \rightarrow \mathbb{N} \cup \{0\} \cup \{\infty\}$ data da:

$$H^A(\omega) := \inf\{n \geq 0 \mid X_n(\omega) \in A\},$$

ponendo che l'estremo inferiore dell'insieme vuoto è ∞ . Si pone inoltre la probabilità a partire da i che $(X_n)_{0 \leq n \leq N}$ "colpisca" A come:

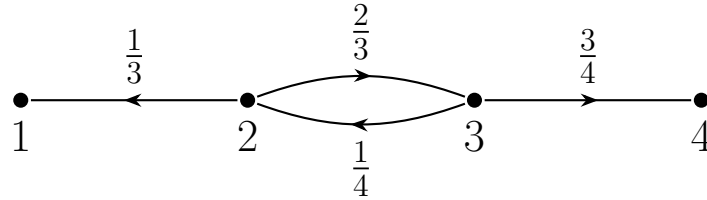
$$h_i^A := \mathbb{P}(H^A < \infty).$$

Infine, se A è una classe chiusa (ovvero: una volta verificato un suo evento, quelli successivi saranno tutti appartenenti ad esso), allora h_i^A è detta probabilità di assorbimento di A . Il tempo medio impiegato da $(X_n)_n$ per raggiungere A è dato da

$$k_i^A := \mathbb{E}_i(H^A) = \left(\sum_{n < \infty} n \mathbb{P}(H^A = n) \right) + \infty \mathbb{P}(H^A = \infty).$$

Guardiamo ora un esempio di queste definizioni appena date:

Esempio 4. Consideriamo la catena di Markov relativa al seguente diagramma.



Vogliamo sapere la probabilità di assorbimento in 4 partendo dall'evento 2 e il tempo medio di assorbimento in 1 oppure in 4. Notiamo che questi due stati sono chiusi, come si evince dal diagramma.

Introduciamo quindi:

$$h_i = \mathbb{P}_i(H^{\{4\}}), \quad k_i = \mathbb{E}_i(H^{\{1,4\}}).$$

Chiaramente si ha che $h_1 = 0$, $h_4 = 1$ e $k_1 = k_4 = 0$. Analizziamo ora stato per stato cosa può succedere. Se per esempio partiamo dallo stato 2, con probabilità $2/3$ andremo in 3 e con probabilità $1/3$ andremo in 1. Quindi⁴

$$h_2 = \frac{2}{3}h_3 + \frac{1}{3}h_1, \quad k_2 = 1 + \frac{2}{3}k_3 + \frac{1}{3}k_1.$$

Allo stesso modo si ottiene:

$$h_3 = \frac{1}{4}h_2 + \frac{3}{4}h_4, \quad k_3 = 1 + \frac{1}{4}k_2 + \frac{3}{4}k_4.$$

Quindi, sostituendo con quanto notato precedentemente:

$$h_2 = \frac{2}{3}h_3 = \frac{2}{3} \left(\frac{1}{4}h_2 + \frac{3}{4} \right),$$

$$k_2 = 1 + \frac{2}{3}k_3 = 1 + \frac{2}{3} \left(1 + \frac{1}{4}k_2 \right).$$

⁴ $k_2 = 1 + \dots$ perché 1 rappresenta il tempo di un passo completato dallo stato 2 a quello immediatamente successivo.

Quindi, risolvendo le due equazioni di primo grado, si ottiene che, partendo dallo stato 2, la probabilità di ottenere (e quindi poi di rimanere in) 4 è $3/5$, e il tempo di assorbimento è 2. A questo punto si trovano in modo semplice anche h_3 e k_3 .

1.3 Il tennis e il suo punteggio

Il tennis è uno sport individuale uno contro uno a punteggio. Si gioca in un campo rettangolare diviso a sua volta in due rettangoli (ognuno da $11,885\text{ m} \times 8,23\text{ m}$), separati da una rete alta $0,914\text{ m}$ al centro e $1,07\text{ m}$ alle estremità laterali del campo.

Analizziamo ora il punteggio di questo sport, perché ci servirà nello studio seguente (utilizziamo il regolamento ufficiale *ITF* [2]). Un *match* (utilizziamo questo termine inglese per evitare di confondere) di tennis si divide in “*game*” (anche detti “giochi”) e “*set*” (anche dette partite). Per vincere un *match*, un giocatore deve aggiudicarsi due o tre *set* (a seconda del tipo di torneo che si sta giocando). La vittoria di un *set* viene raggiunta quando il giocatore conquista sei *game* oppure, nel caso in cui si arrivi al punteggio di cinque *game* pari, di sette *game*. Nel caso in cui si giunga al punteggio di sei giochi pari però, si entra in una parte di gioco chiamata “*tie-break*”, che analizzeremo nella sezione 2.3.1. Per vincere un *game*, il giocatore deve conquistare per primo quattro punti, chiamati in ordine 15, 30, 40 e *gioco*. Se durante un *game* si dovesse arrivare al punteggio di 40 *pari*, per vincerlo occorre conquistare un vantaggio di due punti rispetto al proprio avversario. Quindi se il giocatore A fa punto, egli conquista il “*vantaggio*” (lo chiameremo in seguito Adv_A). Se A fa successivamente ancora punto, allora vince il *game* (Win_A), altrimenti si ritorna nella condizione di parità (*Deuce*). Il meccanismo di un *game* viene mostrato attraverso il diagramma in Figura 1.1.

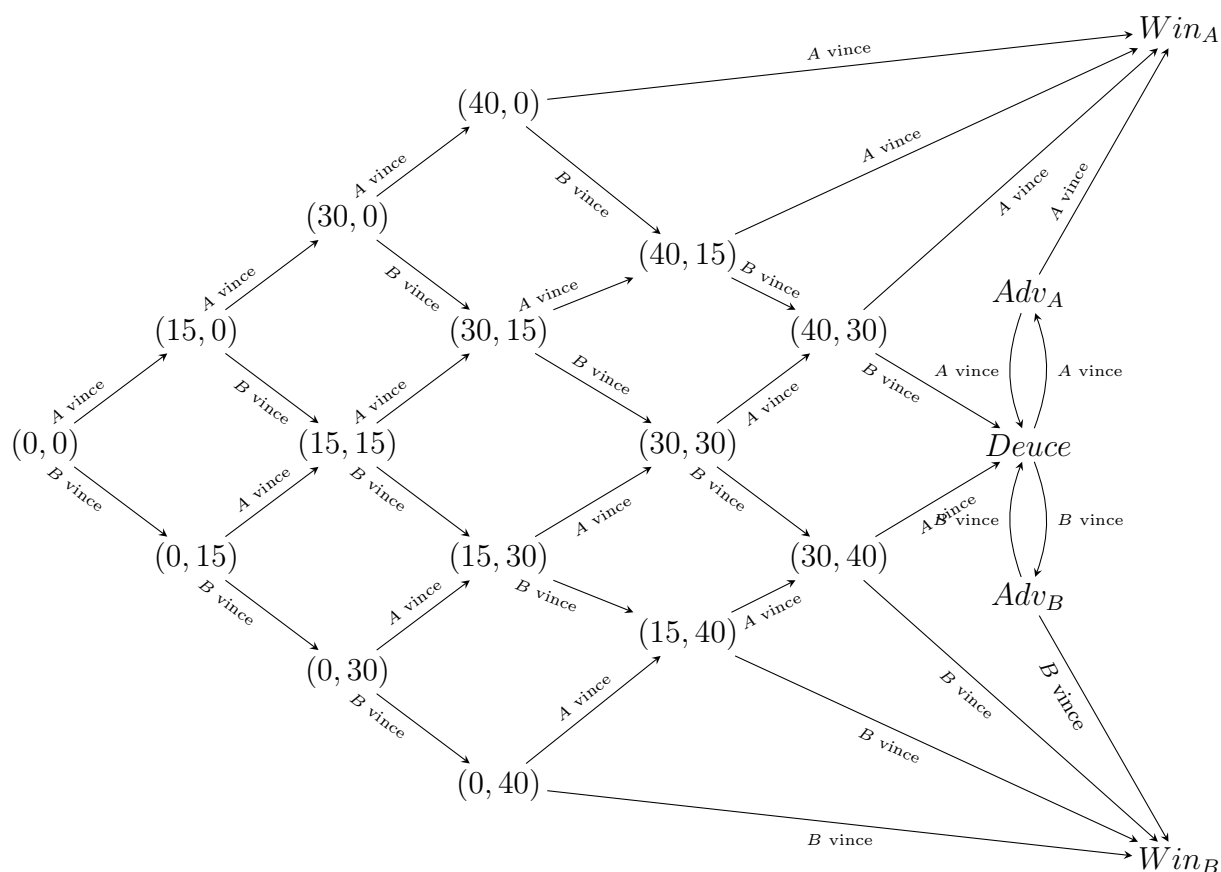


Figura 1.1: Diagramma del sistema di punteggio di un game di tennis. Nella figura i punti sono espressi attraverso una coppia ordinata di due numeri. Considerando l'incontro tra due giocatori A e B, il primo degli elementi della coppia rappresenta il punteggio del giocatore A, il secondo il punteggio del giocatore B.

Capitolo 2

Studio delle probabilità di vittoria e della lunghezza di un match

2.1 Probabilità di vittoria di un game

Cominciamo prendendo due giocatori di tennis A e B che si scontrano in un match di tennis. Iniziamo l'analisi delle probabilità studiando un singolo game. Chiamiamo $p \in]0, 1[$ la probabilità che A vinca un punto durante il suddetto. C'è però da considerare il fatto che durante il corso di un game la probabilità di vincere un punto non rimane costante. Infatti in un'eventuale fase decisiva subentrano altri fattori nelle teste dei due giocatori, tra questi ci può essere soprattutto la tensione dovuta al momento importante.

Al fine di semplificare la trattazione [3], abbiamo scelto quindi di considerare un momento del game che fa cambiare la probabilità di vincere (e quindi automaticamente quella di perdere) il punto ai due giocatori, ovvero il punteggio di *30 pari*. Fino a questa situazione assumiamo quindi che la probabilità che A vinca il punto sia p , dal *30 pari* in poi (punteggio che noi identificheremo con il *40 pari* per semplicità e chiameremo quindi *Deuce*) la probabilità che A vinca il punto sarà $\bar{p} \neq p$ (in generale).

Nella Tabella 2.1 sono raffigurate quindi le situazioni di punteggio possibili durante un game (considerando quanto scritto sopra) e il numero corrispondente da noi associatogli, in modo da semplificare le notazioni che seguiranno. Il punteggio viene raffigurato tramite la coppia ordinata (a, b) , con $a, b \in \{0, 15, 30, 40\}$. Se $a = b = 30$ oppure $a = b = 40$ allora si accede al punteggio chiamato *Deuce* e quindi poi si va ai vantaggi e alla vittoria del game, come spiegato in 1.3.

Questo schema può essere visto come una *Catena di Markov* consistente in 17 eventi possibili (vedasi Figura 2.1). La matrice di transizione associata al processo la chiameremo $P = (p_{ij})_{i,j \in S}$, dove S è l'insieme dei 17 stati. Sempre dallo schema si può notare

Tabella 2.1: Punteggi e stati associatigli

Punteggi	(0, 0)	(15, 0)	(0, 15)	(30, 0)	(15, 15)	(0, 30)	(40, 0)	(30, 15)	(15, 30)
Stati	1	2	3	4	5	6	7	8	9
Punteggi	(0, 40)	(40, 15)	(15, 40)	<i>Deuce</i>	<i>Adv_A</i>	<i>Adv_B</i>	<i>Win_A</i>	<i>Win_B</i>	
Stati	10	11	12	13	14	15	16	17	

che gli stati 16 e 17 sono chiusi. Le corrispettive probabilità e tempi di assorbimento le analizzeremo in seguito.

Analizziamo ora l'evento in cui il giocatore al servizio A vince il game. La probabilità di questo può essere calcolata applicando un teorema importante sulle catene di Markov¹.

Teorema 2.1. *Il vettore h^A delle probabilità di assorbimento di un Markov(λ, T) ($h^A := (h_i^A : i \in I)$) ha come sue componenti le soluzioni minime non negative² del seguente sistema lineare associatogli:*

$$\begin{cases} h_i^A = \sum_{j \in I} t_{ij} h_j^A & \text{se } i \notin A \\ h_i^A = 1 & \text{se } i \in A. \end{cases} \quad (2.1)$$

Dimostrazione. Verifichiamo prima che h^A soddisfa (2.1). Se $X_0 = i \in A$, allora $H^A = 0$, quindi $h_i^A = 1$. Se invece $X_0 = i \notin A$, allora $H^A \geq 1$, quindi per la proprietà delle catene di Markov

$$\mathbb{P}_i(H^A < \infty \mid X_1 = j) = \mathbb{P}_j(H^A < \infty) = h_j^A,$$

e quindi

$$\begin{aligned} h_i^A &= \mathbb{P}_i(H^A < \infty) = \sum_{j \in I} \mathbb{P}_i(H^A < \infty, X_1 = j) \\ &= \sum_{j \in I}^3 \mathbb{P}_i(H^A < \infty \mid X_1 = j) \mathbb{P}_i(X_1 = j) \\ &= \sum_{j \in I} t_{ij} h_j^A. \end{aligned}$$

Per verificare la minimalità, ora supponiamo che $x = (x_i : i \in I)$ sia un qualunque vettore con $x_i \geq 0$ per ogni $i \in I$ che risolva (2.1). Allora $h_i^A = x_i = 1$ per ogni $i \in A$. Sia $i \notin A$, allora

$$x_i = \sum_{j \in I} t_{ij} x_j = \sum_{j \in A} t_{ij} + \sum_{j \notin A} t_{ij} x_j.$$

¹Vedere il paragrafo 1.2.1.

²Per soluzione minima non negativa si intende che se $x = (x_i : i \in I)$ è un altro vettore di soluzioni con $x_i \geq 0$ per ogni i , allora $x_i \geq h_i$ per ogni i .

³L'uguaglianza è valida per definizione di probabilità condizionata.

Sostituendo x_j con la formula appena trovata otteniamo

$$\begin{aligned} x_i &= \sum_{j \in A} t_{ij} + \sum_{j \notin A} t_{ij} \left(\sum_{k \in A} t_{jk} + \sum_{k \notin A} t_{jk} x_k \right) \\ &= \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \sum_{j \notin A} \sum_{k \notin A} t_{ij} t_{jk} x_k. \end{aligned}$$

Ripetendo ancora la sostituzione per x nell'ultimo membro dell'uguaglianza, dopo n step otteniamo

$$\begin{aligned} x_i &= \mathbb{P}_i(X_1 \in A) + \cdots + \mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) \\ &\quad + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} t_{ij_1} t_{j_1 j_2} \cdots t_{j_{n-1} j_n} x_{j_n}. \end{aligned}$$

Ora, siccome $x_i \geq 0$ per ogni $i = 1, \dots, n$, notiamo che tutti termini dell'ultima somma sono non negativi. Infatti essi sono probabilità oppure prodotti di probabilità con componenti di x . Quindi possiamo scrivere che $x_i \geq \mathbb{P}_i(H^A \leq n)$ per ogni n , e mandando n all'infinito si ottiene:

$$x_i \geq \lim_{n \rightarrow \infty} \mathbb{P}_i(H^A \leq n) = \mathbb{P}_i(H^A < \infty) = h_i,$$

che è quanto si voleva dimostrare. \square

Quindi, analizzando il nostro caso, la probabilità di vincere una game per il giocatore al servizio A coincide con la probabilità di assorbimento dello stato 16 a partire dallo stato 1 della catena di Markov appena vista. Applichiamo dunque il Teorema 2.1: occorre trovare la minima soluzione non negativa di

$$\begin{cases} h_i = \sum_{j \in S} p_{ij} h_j & \text{se } 1 \leq i \leq 15 \\ h_{16} = 1, h_{17} = 0, \end{cases}$$

dove p_{ij} sono le probabilità di transizione definite prima e le h_i sono le probabilità di assorbimento dato lo stato iniziale i ($h_{16} = 1$ visto che stiamo analizzando il caso "A vince il game").

La soluzione può essere calcolata piuttosto facilmente e si ottiene che

$$h_1(p, \bar{p}) = p^2 \left[5p^2 - 4p^3 + 4(p-1)^2 p \bar{p} - \frac{2(p-1)^2 \bar{p}^2 (p(4\bar{p}-2) - 2\bar{p}-3)}{2\bar{p}^2 - 2\bar{p} + 1} \right].$$

Denotando quindi con P_{xY}^G la probabilità che Y vinca il game mentre X è al servizio, si ottiene che:

$$\begin{aligned} P_{aA}^G &= h_1(p_A, \bar{p}_A), & P_{aB}^G &= h_1(1-p_A, 1-\bar{p}_A), \\ P_{bB}^G &= h_1(p_B, \bar{p}_B), & P_{bA}^G &= h_1(1-p_B, 1-\bar{p}_B). \end{aligned}$$

È da notare che, sapendo che $P_{xX}^G + P_{xY}^G = 1$ e che $h_1(1 - p_X, 1 - \bar{p}_X) = 1 - h_1(p_X, \bar{p}_X)$, se $p_X = \bar{p}_X$ le probabilità di assorbimento coincidono con quelle analizzate in esempi di partite reali di livello professionistico⁴. Nella Tabella 2.2 vengono riportati i valori di h_1 per le diverse combinazioni di p e \bar{p} .

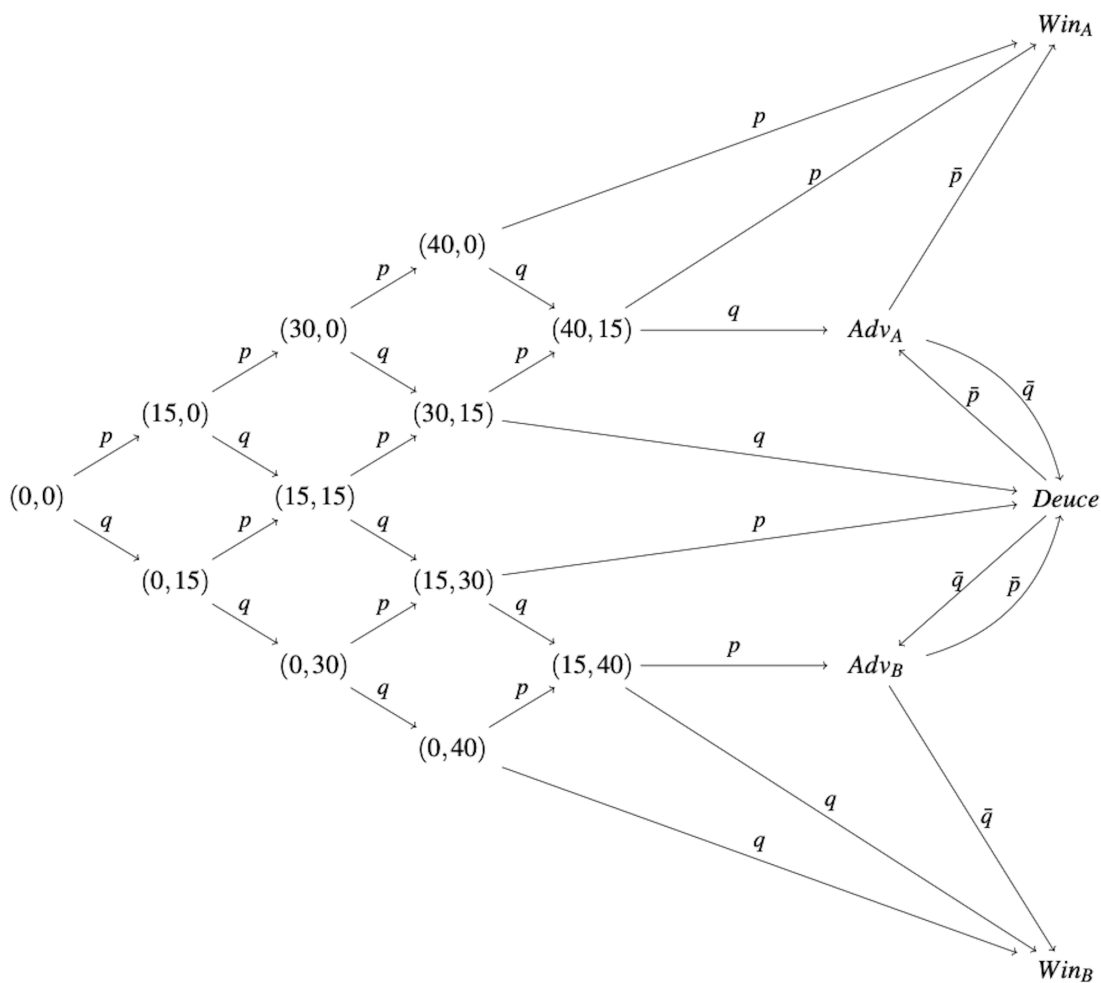


Figura 2.1: Catena di Markov dei punteggi di un game, nel modo in cui sono stati presi in esame nel nostro studio (ovvero il (30,30) viene equiparato al Deuce). p, q, \bar{p}, \bar{q} indicano le probabilità dell'evento successivo a quello della freccia di riferimento, dato l'evento di partenza antecedente la freccia. Ovviamente $q = 1 - p$ e $\bar{q} = 1 - \bar{p}$.

⁴Vedere per esempio [4].

p	\bar{p}									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.001	0.004	0.011	0.021	0.034	0.049	0.062	0.071	0.078	0.081
0.2	0.011	0.022	0.043	0.076	0.119	0.165	0.204	0.233	0.252	0.263
0.3	0.040	0.061	0.099	0.158	0.234	0.312	0.378	0.425	0.455	0.472
0.4	0.102	0.132	0.185	0.264	0.363	0.464	0.549	0.607	0.643	0.663
0.5	0.206	0.242	0.302	0.391	0.500	0.609	0.697	0.758	0.794	0.812
0.6	0.357	0.392	0.451	0.535	0.636	0.736	0.815	0.868	0.898	0.913
0.7	0.545	0.575	0.622	0.688	0.766	0.842	0.901	0.939	0.960	0.969
0.8	0.748	0.767	0.795	0.835	0.881	0.924	0.957	0.978	0.989	0.993
0.9	0.922	0.929	0.938	0.951	0.965	0.979	0.989	0.995	0.998	0.999
1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tabella 2.2: Tabella delle probabilità di vincere un game al variare di $p, \bar{p} \in]0, 1]$.

2.2 Durata di un game

Valutiamo ora la possibile durata di un game, intendendo per “durata” il numero di punti giocati.

Separiamo il caso in cui il giocatore al servizio vince il game (cioè in cui si verifica l'evento 16) da quello in cui il ricevente vinca il game (si verifica l'evento 17). Se consideriamo il caso del servitore vincente, la lunghezza media del game può essere trovata facilmente [5] attraverso il seguente enunciato.

Teorema 2.2. *Data una catena di Markov finita $(X_n)_{0 \leq n \leq N}$ con matrice di transizione associata T , la matrice di transizione associata alla catena di Markov condizionata all'ultimo evento dello spazio degli stati $\{1, \dots, N\}$ T' è data da:*

$$t'_{ij} = t_{ij} \frac{h_j}{h_i} \quad \text{con } i, j \in \{1, \dots, N\},$$

dove $(h_i)_{1 \leq i \leq N}$ sono le probabilità di assorbimento in N .

Ora vediamo un risultato citato in [1] che sarà utile per calcolare i valori attesi richiesti.

Teorema 2.3. *Sia $(X_n)_{0 \leq n \leq N}$ una catena di Markov finita, con P matrice di transizione associatagli. Sia $k^A = (k_i^A : i \in I)$ con $I = \{0, \dots, N\}$ il vettore costituito dai tempi*

medi di assorbimento in A (si veda la Definizione 1.6); allora il vettore k^A è dato dalla più piccola soluzione non negativa del sistema lineare composto dalle seguenti $N + 1$ equazioni:

$$\begin{cases} k_i^A = 0 & \text{se } i \in A \\ k_i^A = 1 + \sum_{j \notin A} p_{ij} k_j^A & \text{se } i \notin A. \end{cases} \quad (2.2)$$

Dimostrazione. Prima di tutto mostriamo che k^A soddisfa il sistema (2.2). Se $X_0 = i \in A$, allora $H^A = 0$, quindi $k_i^A = 0$. Se invece $X_0 = i \notin A$, allora $H^A \geq 0$, quindi, per la proprietà di Markov vale che:

$$\mathbb{E}_i[H^A \mid X_1 = j] = 1 + \mathbb{E}_j[H^A]$$

e

$$\begin{aligned} k_i^A &= \mathbb{E}_i[H^A] = \sum_{j \in I} \mathbb{E}_i[H^A \mathbf{1}_{X_1=j}] \\ &= \sum_{j \in I} \mathbb{E}_i[H^A \mid X_1 = j] \mathbb{P}_i(X_1 = j) \\ &= 1 + \sum_{j \notin A} p_{ij} k_j^A. \end{aligned}$$

Supponiamo ora che $y = (y_i : i \in I)$ sia una soluzione di (2.2). Allora se $i \in A$ vale che $k_i^A = y_i = 0$. Se $i \notin A$ invece,

$$\begin{aligned} y_i &= 1 + \sum_{j \notin A} p_{ij} y_j \\ &= 1 + \sum_{j \notin A} p_{ij} \left(1 + \sum_{k \notin A} p_{jk} y_k \right) \\ &= \mathbb{P}_i(H^A \geq 1) + \mathbb{P}_i(H^A \geq 2) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} y_k. \end{aligned}$$

Ripetendo la sostituzione fino al termine finale, quindi, si ottiene dopo n step che

$$y_i = \mathbb{P}_i(H^A \geq 1) + \cdots + \mathbb{P}_i(H^A \geq n) + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} y_{j_n}.$$

Dunque, se $y \geq 0$,

$$y_i \geq \mathbb{P}_i(H^A \geq 1) + \cdots + \mathbb{P}_i(H^A \geq n)$$

e, mandando $n \rightarrow \infty$, si ottiene

$$y_i \geq \sum_{r=1}^{\infty} \mathbb{P}_i(H^A \geq r) = \mathbb{E}_i[H^A] = k_i^A.$$

La dimostrazione è finita perché è stato mostrato che k_i^A è la soluzione più piccola non negativa dell'equazione. \square

Grazie a questo teorema, ora risolvendo il seguente sistema possiamo trovare i tempi di assorbimento attesi nei 16 stati dato lo stato iniziale uguale a i .

$$\begin{cases} k_i = 1 + \sum_{j \in I \setminus \{i\}} p'_{ij} k_j & \text{se } 1 \leq i \leq 15 \\ k_{16} = 0. \end{cases}$$

⁵Segue quindi che la soluzione per $i = 1$ ad esempio sarà:

$$\begin{aligned} k_1(p, \bar{p}) = & 4 \left[\bar{p}^2(9 - 4\bar{p} + 12\bar{p}^3) + p^3(-5 + 26\bar{p} - 56\bar{p}^2 + 60\bar{p}^3 - 32\bar{p}^4) \right. \\ & - 2p\bar{p}(-3 + 17\bar{p} - 14\bar{p}^2 + 6\bar{p}^3 + 12\bar{p}^4) \left. \right] \left[(1 - 2\bar{p} + 2\bar{p}^2) \left(2\bar{p}^2(3 + 2\bar{p}) \right. \right. \\ & \left. \left. - 4p\bar{p}(-1 + 4\bar{p} + 2\bar{p}^2) - 4\bar{p}^3(1 - 3\bar{p} + 3\bar{p}^2) + p^2(5 - 18\bar{p} + 24\bar{p}^2 + 4\bar{p}^3) \right) \right]^{-1} \\ & + \left[4 \left(p^2(6 - 36\bar{p} + 89\bar{p}^2 - 92\bar{p}^3 + 48\bar{p}^4 + 12\bar{p}^5) \right) \right] \\ & \left[(1 - 2\bar{p} + 2\bar{p}^2) \left(2\bar{p}^2(3 + 2\bar{p}) - 4p\bar{p}(-1 + 4\bar{p} + 2\bar{p}^2) - 4\bar{p}^3(1 - 3\bar{p} + 3\bar{p}^2) \right. \right. \\ & \left. \left. + p^2(5 - 18p + 24\bar{p}^2 + 4\bar{p}^3) \right) \right]^{-1}. \end{aligned}$$

Usando la stessa notazione usata della Sezione 2.1, otteniamo il valore atteso dei punti giocati nei 4 possibili stati di assorbimento.

$$\begin{aligned} k_{aA}^G &= k_1(p_A, \bar{p}_A), & k_{aB}^G &= k_1(1 - p_A, 1 - \bar{p}_A), \\ k_{bB}^G &= k_1(p_B, \bar{p}_B), & k_{bA}^G &= k_1(1 - p_B, 1 - \bar{p}_B). \end{aligned}$$

Nella Tabella 2.3 viene presentato il valore atteso della lunghezza dei game condizionato all'evento in cui il giocatore A al servizio vince il game, ovvero k_{aA}^G . Commentando i dati della tabella si può notare che le probabilità per cui si ottiene la massima durata del game sono $p \approx 0$ e $\bar{p} \approx 0.5$. Non c'è da meravigliarsi, perché stiamo analizzando lunghezze del game *condizionate*. Infatti, se $p \approx 0$ si vede che il percorso che arriva allo stato 16 non raggiunge quasi mai questo stato senza aver prima raggiunto il *Deuce* (la parità) e, da quest'ultimo stato in poi, il \bar{p} vicino a 0.5 rende questa parte del gioco la più lunga possibile.

Dalle equazioni precedenti, può essere facilmente derivata la durata attesa di un game, sommando i prodotti della durata condizionata alla vittoria del game di ciascun giocatore con la sua rispettiva probabilità di assorbimento, ovvero facendo:

$$k_1(p, \bar{p})h_1(p, \bar{p}) + k_1(1 - p, 1 - \bar{p})h_1(1 - p, 1 - \bar{p}).$$

⁵Ricordiamo che $p'_{ij} = p_{ij} \frac{h_j}{h_i}$.

p	\bar{p}									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	5.912	6.804	7.521	8.064	8.332	8.274	7.959	7.525	7.088	6.709
0.2	5.388	6.268	7.100	7.739	8.075	8.064	7.783	7.374	6.956	6.591
0.3	5.108	5.792	6.607	7.310	7.721	7.774	7.542	7.172	6.783	6.441
0.4	4.938	5.423	6.117	6.811	7.274	7.395	7.226	6.906	6.558	6.247
0.5	4.814	5.141	5.673	6.283	6.750	6.923	6.821	6.564	6.269	6.000
0.6	4.708	4.917	5.285	5.760	6.171	6.366	6.325	6.138	5.907	5.690
0.7	4.599	4.721	4.948	5.265	5.572	5.745	5.746	5.629	5.469	5.316
0.8	4.471	4.531	4.644	4.812	4.988	5.101	5.116	5.059	4.972	4.886
0.9	4.293	4.311	4.345	4.396	4.453	4.492	4.500	4.484	4.458	4.431
1.0	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000

Tabella 2.3: Tabella dei valori attesi delle durate di un game al variare di $p, \bar{p} \in]0, 1]$, dando per scontato che il giocatore al servizio vinca il game.

p	\bar{p}								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	4.455	4.489	4.533	4.567	4.580	4.586	4.569	4.539	4.511
0.2	4.977	5.086	5.201	5.301	5.355	5.349	5.280	5.193	5.097
0.3	5.455	5.639	5.831	5.992	6.075	6.048	5.929	5.763	5.593
0.4	5.808	6.044	6.287	6.483	6.565	6.513	6.351	6.119	5.898
0.5	5.969	6.220	6.467	6.673	6.750	6.673	6.467	6.220	5.969
0.6	5.898	6.119	6.351	6.513	6.565	6.483	6.287	6.044	5.808
0.7	5.593	5.763	5.929	6.048	6.075	5.992	5.831	5.639	5.455
0.8	5.097	5.193	5.280	5.349	5.355	5.301	5.201	5.086	4.977
0.9	4.511	4.539	4.569	4.586	4.580	4.567	4.533	4.489	4.455

Tabella 2.4: Tabella dei valori attesi delle durate di un game al variare di $p, \bar{p} \in]0, 1[$.

I valori possibili che si calcolano applicando questa formula al variare di p e \bar{p} si trovano nella Tabella 2.4.

2.3 Probabilità di vittoria e durata di un set

Vogliamo ora valutare la probabilità di vittoria e la durata attesa di un set. Per farlo vogliamo usare i risultati relativi al game ottenuti nella sezione precedente; quindi considereremo ogni game indipendentemente dall'altro, a prescindere dalla situazione di punteggio. Approcceremo i set con una catena di Markov, come già fatto con i game, e da questa trarremo le probabilità e i valori attesi. Questo metodo ci permette di trovare la durata del set come il numero di game attesi, ma ciascuno di essi potrà avere a sua volta una lunghezza differente, come visto in 2.1.

Valuteremo in particolare la probabilità che un dato set finisca con uno particolare dei 14 possibili risultati (cioè $(6, 0), (6, 1), \dots, (7, 6), (6, 7)$), prendendo anche in considerazione il numero di *break* (ovvero game vinti dal giocatore che riceve il servizio) conseguiti da ciascuno dei due giocatori.

Cominceremo considerando il *tiebreak*, il game speciale giocato in caso di situazione di punteggio di 6 game pari; poi analizzeremo separatamente le probabilità di vittoria di un set e la sua durata attesa.

2.3.1 Il tiebreak

Il *tiebreak* è uno speciale tipo di game. Viene giocato quando il punteggio dei game nel corso di un set arriva nello stato di 6 game pari. In questo caso per decidere chi si aggiudicherà il set si gioca il suddetto tiebreak a 7 punti, contati in modo tradizionale $\{1, 2, 3, \dots\}$. In caso di 6 punti pari però, per vincere è necessario avere 2 punti di vantaggio sul proprio avversario. Durante il tiebreak, il giocatore che serve la pallina in campo viene cambiato ogni due punti, tranne dopo il primo, quando il giocatore che inizia a servire (lo stesso che aveva iniziato a battere nel primo game del set) lascia subito il servizio al suo avversario.

Ora, se assumiamo che le probabilità di vincere un punto del tiebreak da parte del giocatore al servizio siano tutte uguali per qualunque situazione di punteggio, questa parte speciale della partita può essere modellizzata tramite una catena di Markov. A causa però della particolare regola sullo scambio dei turni di servizio, lo spazio degli stati sarà composto in questo caso da 53 elementi, che ora verranno elencati:

$$\tilde{S} = \left(\{(i, j) : i, j \in \{0, \dots, 6\}, i + j \leq 10\} \setminus \{(5, 5)\} \right) \cup \\ \{\text{Deuce}_A, \text{Deuce}_B, \text{Adv}_{AA}, \text{Adv}_{BB}, \text{Adv}_{AB}, \text{Win}_A, \text{Win}_B\},$$

dove Deuce_A indica lo stato in cui i due giocatori hanno vinto pari punti da 5 in su, e la A sta a indicare che è lui che serve la pallina in campo (nel tiebreak infatti, come abbiamo

già visto per i game, ai nostri fini probabilistici lo stato di 5 punti pari è equivalente a quello di 6 punti pari, infatti in entrambe le situazioni un giocatore per vincere ha bisogno di conquistare due punti consecutivi). Lo stato Adv_{AA} indica invece il punteggio in cui il giocatore A è in vantaggio e servirà per conquistare il set, e allo stesso modo di conseguenza si interpretano gli altri simboli.

Le probabilità di transizione sono definite come nella sezione precedente, facendo attenzione però a quale giocatore, nella specifica situazione di punteggio, è al servizio. Per esempio, se definiamo $p_{a,b}$ come la probabilità di passare dal punteggio a al punteggio b , e definendo $p_{(0,0),(1,0)} = p_A$ con A al servizio nel primo punto, allora $p_{(1,0),(2,0)} = 1 - p_B$, $p_{(2,0),(3,0)} = 1 - p_B$, $p_{(3,0),(4,0)} = p_A$, e così via. In [4] (Newton and Keller, 2005, pagine 266-268) è riportato il calcolo delle probabilità di vittoria del tiebreak. Esse sono date dalla risoluzione di un sistema ricorsivo. Scriviamo qui di seguito l'equazione di partenza, ma non il resto del sistema e le sue soluzioni⁶.

$$p_A^T = \sum_{j=0}^5 p_A(7, j) + p_A(6, 6) \sum_{n=0}^{\infty} p_A(n+2, n). \quad (2.3)$$

Per quanto riguarda il numero atteso di punti giocati invece, bisogna risolvere ancora una volta un sistema di equazioni lineari:

$$\begin{cases} k_i = 1 + \sum_{j \in \bar{5}} p_{ij} k_j & \text{se } 1 \leq i \leq 51, \\ k_{52} = 0, \\ k_{53} = 0. \end{cases} \quad (2.4)$$

Risolvendo questo sistema, si ottiene che il numero atteso di punti giocati per completare un tiebreak è il seguente:

$$\begin{aligned} k_1(p_A, p_B) = & [-2 - 30p_B^2 + 71p_B^3 - 94p_B^4 + 73p_B^5 - 30p_B^6 + 5p_B^7 + \\ & + p_A p_B (-115 + 541p_B - 1166p_B^2 + 1483p_B^3 - 1124p_B^4 + 465p_B^5 - 80p_B^6) + \\ & + p_A^2 (-25 + 500p_B - 2650p_B^2 + 6514p_B^3 - 8716p_B^4 + 6557p_B^5 - 2600p_B^6 + 420p_B^7) + \\ & + p_A^3 (40 - 885p_B + 5730p_B^2 - 16276p_B^3 + 23769p_B^4 - 18398p_B^5 + 7000p_B^6 - 980p_B^7) + \\ & + p_A^4 (-25 + 792p_B - 6224p_B^2 + 20289p_B^3 - 32532p_B^4 + 26320p_B^5 - 9660p_B^6 + 1050p_B^7) + \\ & + p_A^5 (4 - 347p_B + 3553p_B^2 - 12428p_B^3 + 21784p_B^4 - 18466p_B^5 + 6510p_B^6 - 420p_B^7) \\ & + p_A^6 (1 + 58p_B - 716p_B^2 + 2996p_B^3 - 5698p_B^4 + 5040p_B^5 - 1680p_B^6)] \cdot \\ & \cdot [-p_B + p_A(-1 + 2p_B)]^{-1}. \end{aligned}$$

⁶Con $p_A(i, j)$ si intende la probabilità di arrivare al punteggio (i, j) , sempre considerando che A sia il giocatore che inizi a servire. p_A^T è invece la probabilità che A vinca il tiebreak partendo da un suo servizio.

Da notare che non abbiamo separato il caso di un tiebreak vinto dal giocatore A da quello vinto dal giocatore B . Infatti dividere le due computazioni sarebbe troppo oneroso e poco utile ai nostri fini. Tra l'altro, il numero di tiebreak giocati mediamente durante un match è piuttosto basso.

p_B	p_A										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	∞	22.12	13.59	11.21	10.09	9.34	8.74	8.20	7.73	7.33	7.00
0.1	22.38	16.18	13.55	12.00	10.91	10.06	9.36	8.75	8.23	7.78	7.40
0.2	13.80	13.61	13.02	12.28	11.50	10.73	10.01	9.35	8.76	8.25	7.81
0.3	11.29	12.04	12.29	12.18	11.81	11.27	10.65	10.00	9.35	8.76	8.24
0.4	10.07	10.92	11.51	11.81	11.84	11.62	11.21	10.65	10.01	9.35	8.72
0.5	9.30	10.05	10.73	11.28	11.62	11.74	11.62	11.28	10.73	10.05	9.30
0.6	8.72	9.35	10.01	10.65	11.21	11.62	11.84	11.81	11.51	10.92	10.07
0.7	8.24	8.76	9.35	10.00	10.65	11.27	11.81	12.18	12.29	12.04	11.29
0.8	7.81	8.25	8.76	9.35	10.01	10.73	11.50	12.28	13.02	13.61	13.80
0.9	7.40	7.78	8.23	8.75	9.36	10.06	10.91	12.00	13.55	16.18	22.38
1.0	7.00	7.33	7.73	8.20	8.74	9.34	10.09	11.21	13.59	22.12	∞

Tabella 2.5: *Tabella dei valori attesi delle durate di un tiebreak al variare di $p_A, p_B \in [0, 1]$.*

2.3.2 Probabilità di vittoria di un set

Consideriamo un set di tennis qualunque, comprendendo anche il caso di uno con il tiebreak finale. In questa parte calcoleremo esplicitamente la probabilità che il punteggio finale del set sia uno dei sette casi possibili seguenti: $(6, 0)$, $(6, 1)$, $(6, 2)$, $(6, 3)$, $(6, 4)$, $(7, 5)$, $(7, 6)$. Deriveremo poi i risultati per i rimanenti casi. Per semplificare la notazione, se il giocatore A (o rispettivamente B) comincia a servire nel primo game del set, denoteremo la probabilità di vittoria condizionata a questo evento con \mathbb{P}_a (rispettivamente \mathbb{P}_b).

Con qualche di abuso di notazione, definiamo innanzi tutto il vettore di variabili aleatorie (A, B) , il quale indica il punteggio finale di un set. Poi costruiamo un altro vettore di v.a. composto da 5 componenti $X := (x_0, x_1, x_2, x_3, x_4)$. x_0 indica chi ha iniziato a servire nel primo game (0 se ha iniziato A , 1 se ha iniziato B). Le altre variabili x_1, \dots, x_4 indicano quanti game sono terminati con l'esito rispettivo di aA, aB, bB, bA (vedasi 2.1). Quindi, per esempio, se un set con il primo servizio di A dovesse concludersi

con il punteggio di sei giochi a zero per il giocatore A , allora l'evento ad esso associatogli sarà ovviamente $\{(A, B) = (6, 0)\}$, il quale però coinciderà esattamente con l'evento $\{X = (0, 3, 0, 3, 0)\}$. Analogamente allora l'evento $\{(A, B) = (6, 1)\}$, sempre dato il fatto che A inizi a servire, coinciderà con $\{X = (0, 4, 0, 1, 2)\} \cup \{X = (0, 3, 1, 0, 3)\}$.

Per calcolare la probabilità che un set finisca con un determinato punteggio determineremo prima tutti i possibili stati di X che portano a quel finale di set e poi, grazie alle assunzioni dovute alle catene di Markov, sostituiranno le probabilità nelle loro formule. Vediamo quindi le equazioni per ogni singolo punteggio che renda il set vincente per il giocatore A , seguendo la notazione scritta nel precedente paragrafo.

$$\begin{aligned} \mathbb{P}_a[(6, 0)] &= \mathbb{P}[X = (0, 3, 0, 0, 3)] = (P_{aA}^G)^3 (P_{bA}^G)^3, \\ \mathbb{P}_b[(6, 0)] &= \mathbb{P}[X = (1, 3, 0, 0, 3)] = (P_{aA}^G)^3 (P_{bA}^G)^3, \\ \mathbb{P}_a[(6, 1)] &= 3 \cdot \mathbb{P}[(0, 4, 0, 1, 2)] + 3 \cdot \mathbb{P}[(0, 3, 1, 0, 3)] \\ &= 3[(P_{aA}^G)^4 P_{bB}^G (P_{bA}^G)^2] + 3[(P_{aA}^G)^3 P_{aB}^G (P_{bA}^G)^3], \\ \mathbb{P}_b[(6, 1)] &= 3 \cdot \mathbb{P}[(1, 3, 0, 1, 3)] + 3 \cdot \mathbb{P}[(1, 2, 1, 0, 4)] \\ &= 3[(P_{aA}^G)^3 P_{bB}^G (P_{bA}^G)^3] + 3[(P_{aA}^G)^2 P_{aB}^G (P_{bA}^G)^4]. \end{aligned}$$

Da notare che nel caso del punteggio finale $(6, 1)$ esistono 3 casi diversi in cui si può verificare il vettore aleatorio $X = (1, 3, 0, 1, 3)$ a seconda dell'ordine in cui vengono conseguiti i game idem per $(X = (1, 3, 0, 1, 3))$; per questo motivo si conterà la sua probabilità per 3 volte. Nei prossimi casi quindi troveremo dei coefficienti davanti alle rispettive probabilità; la loro somma sarà sempre uguale al numero di combinazioni di $A+B-1$ elementi presi per B volte (con (A, B) punteggio finale del set, il -1 è giustificato dal fatto che l'ultimo game deve necessariamente vincerlo A , visto il risultato):

$$\binom{A+B-1}{B}.$$

Quindi, andando avanti:

$$\begin{aligned}
\mathbb{P}_a[(6, 2)] &= 3 \cdot \mathbb{P}[(0, 4, 0, 2, 2)] + 12 \cdot \mathbb{P}[(0, 3, 1, 1, 3)] + 6 \cdot \mathbb{P}[(0, 2, 2, 0, 4)] \\
&= 3[(P_{aA}^G)^4(P_{bB}^G)^2(P_{bA}^G)^2] + 12[(P_{aA}^G)^3P_{aB}^G P_{bB}^G(P_{bA}^G)^3] + 6[(P_{aA}^G)^2(P_{aB}^G)^2(P_{bA}^G)^4], \\
\mathbb{P}_b[(6, 2)] &= 6 \cdot \mathbb{P}[(1, 4, 0, 2, 2)] + 12 \cdot \mathbb{P}[(1, 3, 1, 1, 3)] + 3 \cdot \mathbb{P}[(1, 2, 2, 0, 4)] \\
&= 6[(P_{aA}^G)^4(P_{bB}^G)^2(P_{bA}^G)^2] + 12[(P_{aA}^G)^3P_{aB}^G P_{bB}^G(P_{bA}^G)^3] + 3[(P_{aA}^G)^2(P_{aB}^G)^2(P_{bA}^G)^4], \\
\mathbb{P}_a[(6, 3)] &= 4 \cdot \mathbb{P}[(0, 5, 0, 3, 1)] + 24 \cdot \mathbb{P}[(0, 4, 1, 2, 2)] + 24 \cdot \mathbb{P}[(0, 3, 2, 1, 3)] + 4 \cdot \mathbb{P}[(0, 2, 3, 0, 4)] \\
&= 4[(P_{aA}^G)^5(P_{bB}^G)^3(P_{bA}^G)] + 24[(P_{aA}^G)^4P_{aB}^G(P_{bB}^G)^2(P_{bA}^G)^2] + \\
&\quad + 24[(P_{aA}^G)^3(P_{aB}^G)^2P_{bB}^G(P_{bA}^G)^3] + 4[(P_{aA}^G)^2(P_{aB}^G)^3(P_{bA}^G)^4], \\
\mathbb{P}_b[(6, 3)] &= 4 \cdot \mathbb{P}[(1, 4, 0, 3, 2)] + 24 \cdot \mathbb{P}[(1, 3, 1, 2, 3)] + 24 \cdot \mathbb{P}[(1, 2, 2, 1, 4)] + 4 \cdot \mathbb{P}[(1, 1, 3, 0, 5)] \\
&= 4[(P_{aA}^G)^4(P_{bB}^G)^3(P_{bA}^G)^2] + 24[(P_{aA}^G)^3P_{aB}^G(P_{bB}^G)^2(P_{bA}^G)^3] + \\
&\quad + 24[(P_{aA}^G)^2(P_{aB}^G)^2P_{bB}^G(P_{bA}^G)^4] + 4[P_{aA}^G(P_{aB}^G)^3(P_{bA}^G)^5], \\
\mathbb{P}_a[(6, 4)] &= \mathbb{P}[(0, 5, 0, 4, 1)] + 20 \cdot \mathbb{P}[(0, 4, 1, 3, 2)] + 60 \cdot \mathbb{P}[(0, 3, 2, 2, 3)] + \\
&\quad + 40 \cdot \mathbb{P}[(0, 2, 3, 1, 4)] + 5 \cdot \mathbb{P}[(0, 1, 4, 0, 5)] = \\
&= [(P_{aA}^G)^5(P_{bB}^G)^4P_{aA}^G] + 20[(P_{aA}^G)^4P_{aB}^G] + 20[(P_{aA}^G)^4P_{aB}^G(P_{bB}^G)^3(P_{bA}^G)^2] + \\
&\quad + 60[(P_{aA}^G)^3(P_{aB}^G)^2(P_{bB}^G)^2(P_{bA}^G)^3] + 40[(P_{aA}^G)^2(P_{aB}^G)^3P_{bB}^G(P_{bA}^G)^4] + 5[P_{aA}^G(P_{aB}^G)^4(P_{bA}^G)^5], \\
\mathbb{P}_b[(6, 4)] &= 5 \cdot \mathbb{P}[(1, 5, 0, 4, 1)] + 40 \cdot \mathbb{P}[(1, 4, 1, 3, 2)] + 60 \cdot \mathbb{P}[(1, 3, 2, 2, 3)] + \\
&\quad + 20 \cdot \mathbb{P}[(1, 2, 3, 1, 4)] + \mathbb{P}[(1, 1, 4, 0, 5)] = \\
&= 5[(P_{aA}^G)^5(P_{bB}^G)^4P_{bA}^G] + 40[(P_{aA}^G)^4P_{aB}^G(P_{bB}^G)^3(P_{bA}^G)^2] + 60[(P_{aA}^G)^3(P_{bB}^G)^2(P_{bB}^G)^2(P_{bA}^G)^3] + \\
&\quad + 20[(P_{aA}^G)^2(P_{aB}^G)^3P_{bB}^G(P_{bA}^G)^4] + [P_{aA}^G(P_{aB}^G)^4(P_{bA}^G)^5], \\
\mathbb{P}_a[(7, 5)] &= \mathbb{P}[(x_0, 6, 0, 5, 1)] + 25 \cdot \mathbb{P}[(x_0, 5, 1, 4, 2)] + 100 \cdot \mathbb{P}[(x_0, 4, 2, 3, 3)] + \\
&\quad + 100 \cdot \mathbb{P}[(x_0, 3, 3, 2, 4)] + 25 \cdot \mathbb{P}[(x_0, 2, 4, 1, 5)] + \mathbb{P}[(x_0, 1, 5, 0, 6)] = \\
&= [(P_{aA}^G)^6(P_{bB}^G)^5P_{bA}^G] + 25[(P_{aA}^G)^5P_{aB}^G(P_{bB}^G)^4(P_{bA}^G)^2] + 100[(P_{aA}^G)^4(P_{aB}^G)^2(P_{bB}^G)^3(P_{bA}^G)^3] + \\
&\quad + 100[(P_{aA}^G)^3(P_{aB}^G)^3(P_{bB}^G)^2(P_{bA}^G)^4] + 25[(P_{aA}^G)^2(P_{aB}^G)^4P_{bB}^G(P_{bA}^G)^5] + [P_{aA}^G(P_{aB}^G)^5(P_{bA}^G)^6] \\
&= \mathbb{P}_b[(7, 5)].
\end{aligned}$$

Come possiamo notare, in quest'ultima uguaglianza x_0 può essere uguale allo stesso modo a 0 oppure 1, il risultato delle probabilità rimane indifferente. Questo avviene grazie al fatto che per trovarsi nel punteggio di (7, 5) si è vincolati a trovarsi necessariamente nel punteggio di (5, 5) nei game precedenti.

Quando invece il risultato di un set è (7, 6), si hanno due possibili casi:

$$\mathbb{P}_a[(7, 6)] = \mathbb{P}_a[A \text{ vince il tiebreak} \mid (6, 6)] \cdot \mathbb{P}_a[(6, 6)]$$

e

$$\mathbb{P}_b[(7, 6)] = \mathbb{P}_b[A \text{ vince il tiebreak} \mid (6, 6)] \cdot \mathbb{P}_b[(6, 6)].$$

Analizziamo allora il secondo termine delle due somme:

$$\begin{aligned}
\mathbb{P}_a[(6, 6)] &= \mathbb{P}[(x_0, 6, 0, 6, 0)] + 26 \cdot \mathbb{P}[(x_0, 5, 1, 5, 1)] + 125 \cdot \mathbb{P}[(x_0, 4, 2, 4, 2)] + 200 \cdot \mathbb{P}[(x_0, 3, 3, 3, 3)] + \\
&\quad + 125 \cdot \mathbb{P}[(x_0, 2, 4, 2, 4)] + 26 \cdot \mathbb{P}[(x_0, 1, 5, 1, 5)] + \mathbb{P}[(x_0, 0, 6, 0, 6)] = \\
&= (P_{aA}^G)^6 (P_{bB}^G)^6 + 26[(P_{aA}^G)^5 P_{aB}^G (P_{bB}^G)^5 P_{bA}^G] + 125[(P_{aA}^G)^4 (P_{aB}^G)^2 (P_{bB}^G)^4 (P_{bA}^G)^2] + \\
&\quad + 200[(P_{aA}^G)^3 (P_{aB}^G)^3 (P_{bB}^G)^3 (P_{bA}^G)^3] + 125[(P_{aA}^G)^2 (P_{aB}^G)^4 (P_{bB}^G)^2 (P_{bA}^G)^4] + \\
&\quad + 26[P_{aA}^G (P_{aB}^G)^5 P_{bB}^G (P_{bA}^G)^5] + (P_{aB}^G)^6 (P_{bA}^G)^6 = \\
&= \mathbb{P}_b[(6, 6)],
\end{aligned}$$

mentre $\mathbb{P}_a[A \text{ vince il tiebreak} \mid (6, 6)]$ e $\mathbb{P}_b[A \text{ vince il tiebreak} \mid (6, 6)]$ possono essere trovate su [4], come già detto nel paragrafo 2.3.1. Per analizzare i casi rimanenti (ovvero quelli in cui $B > A$, cioè quando B vince il set), denotiamo

$$\begin{aligned}
\mathbb{P}_a[(A, B)] &= S_{A,B}(P_{aA}^G, P_{aB}^G, P_{bB}^G, P_{bA}^G), \\
\mathbb{P}_b[(A, B)] &= T_{A,B}(P_{aA}^G, P_{aB}^G, P_{bB}^G, P_{bA}^G),
\end{aligned}$$

con S e T funzioni che dipendono anche dai due parametri del risultato finale A e B .

È facile notare quindi che

$$\begin{aligned}
\mathbb{P}_a[(B, A)] &= S_{A,B}(P_{aB}^G, P_{aA}^G, P_{bA}^G, P_{bB}^G), \\
\mathbb{P}_b[(B, A)] &= T_{A,B}(P_{aB}^G, P_{aA}^G, P_{bA}^G, P_{bB}^G).
\end{aligned}$$

2.3.3 Durata di un set

In questa sezione valuteremo il numero atteso di punti giocati durante un set. Assumiamo che A cominci a servire per primo, presenteremo qui solo i risultati validi nel caso di A vincente del set. Gli altri casi saranno poi facilmente ottenibili. Sia D la variabile aleatoria indicante la durata di un set. Valutiamo il suo valore atteso $\mathbb{E}_a[D] = \mathbb{E}_a[\mathbb{E}_a[D \mid (A, B)]]$, dove $\mathbb{E}_a[D \mid (A, B)]$ indica il valore atteso condizionato al risultato finale del set (A, B) e al fatto che A inizi a servire. Denotando quindi con $\mathbb{E}_a[D \mid (i, j)] = \mathbb{E}_a[D \mid (A, B) = (i, j)]$, avremo:

$$\begin{aligned}
\mathbb{E}_a[D] &= \sum_{i=0}^4 \mathbb{E}_a[D \mid (6, i)] \cdot \mathbb{P}_a[(A, B) = (6, i)] + \\
&\quad + \mathbb{E}_a[D \mid (7, 5)] \cdot \mathbb{P}_a[(A, B) = (7, 5)] + \\
&\quad + (\mathbb{E}_a[D \mid (7, 6)] + \mathbb{E}_a[D \mid (6, 7)]) \cdot \mathbb{P}_a[(A, B) = (6, 6)] + \\
&\quad + \sum_{i=0}^4 \mathbb{E}_a[D \mid (i, 6)] \cdot \mathbb{P}_a[(A, B) = (i, 6)] + \\
&\quad + \mathbb{E}_a[D \mid (5, 7)] \cdot \mathbb{P}_a[(A, B) = (5, 7)].
\end{aligned} \tag{2.5}$$

Vediamo ora i vari valori attesi con $A > B$ (eccetto il caso di (7, 6) e (6, 7) che valuteremo successivamente).⁷

$$\begin{aligned}
\mathbb{E}_a[D \mid (6, 0)] \cdot \mathbb{P}_a[(6, 0)] &= \mathbb{P}[(0, 3, 0, 0, 3)](3k_{aA}^G + 3k_{bA}^G), \\
\mathbb{E}_a[D \mid (6, 1)] \cdot \mathbb{P}_a[(6, 1)] &= 3\mathbb{P}[(0, 4, 1, 2)](4k_{aA}^G + k_{bB}^G + 2k_{bA}^G) + \\
&\quad + 3\mathbb{P}[(0, 3, 1, 0, 3)](3k_{aA}^G + k_{bA}^G + 3k_{bA}^G), \\
\mathbb{E}_a[D \mid (6, 2)] \cdot \mathbb{P}_a[(6, 2)] &= 3\mathbb{P}[(0, 4, 0, 2, 2)](4k_{aA}^G + 2k_{bB}^G + 2k_{bA}^G) + \\
&\quad + 12\mathbb{P}[(0, 3, 1, 1, 3)](3k_{aA}^G + k_{aB}^G + k_{bB}^G + 3k_{bA}^G) + \\
&\quad + 6\mathbb{P}[(0, 2, 2, 0, 4)](2k_{aA}^G + 2k_{aB}^G + 4k_{bA}^G), \\
\mathbb{E}_a[D \mid (6, 3)] \cdot \mathbb{P}_a[(6, 3)] &= 4\mathbb{P}[(0, 5, 0, 3, 1)](5k_{aA}^G + 3k_{bB}^G + k_{bA}^G) + \\
&\quad + 24\mathbb{P}[(0, 4, 1, 2, 2)](4k_{aA}^G + k_{aB}^G + 2k_{bB}^G + 2k_{bA}^G) + \\
&\quad + 24\mathbb{P}[(0, 3, 2, 1, 3)](3k_{aA}^G + 2k_{aB}^G + k_{bB}^G + 3k_{bA}^G) + \\
&\quad + 4\mathbb{P}[(0, 2, 3, 0, 4)](2k_{aA}^G + 3k_{aB}^G + 4k_{bA}^G), \\
\mathbb{E}_a[D \mid (6, 4)] \cdot \mathbb{P}_a[(6, 4)] &= \mathbb{P}[(0, 5, 0, 4, 1)](5k_{aA}^G + 4k_{bB}^G + k_{bA}^G) + \\
&\quad + 20\mathbb{P}[(0, 4, 1, 3, 2)](4k_{aA}^G + k_{aB}^G + 3k_{bB}^G + 2k_{bA}^G) + \\
&\quad + 60\mathbb{P}[(0, 3, 2, 2, 3)](3k_{aA}^G + 2k_{aB}^G + 2k_{bB}^G + 3k_{bA}^G) + \\
&\quad + 40\mathbb{P}[(0, 2, 3, 1, 4)](2k_{aA}^G + 3k_{aB}^G + k_{bB}^G + 4k_{bA}^G) + \\
&\quad + 5\mathbb{P}[(0, 1, 4, 0, 5)](k_{aA}^G + 4k_{aB}^G + 5k_{bA}^G), \\
\mathbb{E}_a[D \mid (7, 5)] \cdot \mathbb{P}_a[(7, 5)] &= \mathbb{P}[(0, 6, 0, 5, 1)](6k_{aA}^G + 5k_{bB}^G + k_{bA}^G) + \\
&\quad + 25\mathbb{P}[(0, 5, 1, 4, 2)](5k_{aA}^G + k_{aB}^G + 4k_{bB}^G + k_{bA}^2) + \\
&\quad + 100\mathbb{P}[(0, 4, 2, 3, 3)](4k_{aA}^G + 2k_{aB}^G + 3k_{bB}^G + 3k_{bA}^G) + \\
&\quad + 100\mathbb{P}[(0, 3, 3, 2, 4)](3k_{aA}^G + 3k_{aB}^G + 2k_{bB}^G + 4k_{bA}^G) + \\
&\quad + 25\mathbb{P}[(0, 2, 4, 1, 5)](2k_{aA}^G + 4k_{aB}^G + k_{bB}^G + 5k_{bA}^G) + \\
&\quad + \mathbb{P}[(0, 1, 5, 0, 6)](k_{aA}^G + 5k_{aB}^G + 6k_{bA}^G).
\end{aligned}$$

Osserviamo che, chiaramente, $\mathbb{E}_a[D \mid (i, j)]$ può essere ottenuto dividendo la precedente formula per $\mathbb{P}_a[(i, j)]$. Il caso del tiebreak (cioè $(A, B) = (7, 6)$ oppure $(6, 7)$) è alquanto delicato poiché, usando questa formula, non riusciamo a distinguere la durata di un

⁷Ricordiamo che k_{aB}^G è il valore atteso della durata di un game in cui A serve e B conquista il game, vedasi 2.1.

tiebreak vinto da A rispetto a quella di un tiebreak vinto da B . Infatti abbiamo:⁸

$$\begin{aligned}
& (\mathbb{E}_a[D \mid (7, 6)] + \mathbb{E}_a[D \mid (6, 7)]) \cdot \mathbb{P}_a[(6, 6)] = \\
& = \mathbb{P}_a[(0, 6, 0, 6, 0)](6k_{aA}^G + 6k_{bB}^G + k_a^T) + \\
& \quad + 26\mathbb{P}_a[(0, 5, 1, 5, 1)](5k_{aA}^G + k_{aB}^G + 5k_{bB}^G + k_{bA}^G + k_a^T) + \\
& \quad + 125\mathbb{P}_a[(0, 4, 2, 4, 2)](4k_{aA}^G + 2k_{aB}^G + 4k_{bB}^G + 2k_{bA}^G + k_a^T) + \\
& \quad + 200\mathbb{P}_a[(0, 3, 3, 3, 3)](3k_{aA}^G + 3k_{aB}^G + 3k_{bB}^G + 3k_{bA}^G + k_a^T) + \\
& \quad + 125\mathbb{P}_a[(0, 2, 4, 2, 4)](2k_{aA}^G + 4k_{aB}^G + 2k_{bB}^G + 4k_{bA}^G + k_a^T) + \\
& \quad + 26\mathbb{P}_a[(1, 5, 1, 5)](k_{aA}^G + 5k_{aB}^G + k_{bB}^G + 5k_{bA}^G + k_a^T) + \\
& \quad + \mathbb{P}_a[(0, 6, 0, 6)](6k_{aB}^G + 6k_{bA}^G + k_a^T).
\end{aligned}$$

Notiamo infine che, come visto per la probabilità di vittoria nella sezione 2.3.2, che se

$$\mathbb{E}_a[A, B] \cdot \mathbb{P}_a[A, B] = H(P_{aA}^G, P_{aB}^G, P_{bB}^G, P_{bA}^G, k_{aA}^G, k_{aB}^G, k_{bB}^G, k_{bA}^G),$$

allora

$$\mathbb{E}_a[B, A] \cdot \mathbb{P}_a[B, A] = H(P_{aB}^G, P_{aA}^G, P_{bA}^G, P_{bB}^G, k_{aB}^G, k_{aA}^G, k_{bA}^G, k_{bB}^G).$$

Nella Tabella 2.6 presentiamo i possibili valori attesi del numero di punti giocati in un set al variare di p_A e p_B , intendendo con p_A la probabilità che ha il giocatore A di fare punto durante un suo turno di battuta (per non complicare la tabella, consideriamo la probabilità per ogni punto costante nel corso del game, senza considerare il caso post-*Deuce*).

⁸Per k_a^T intendiamo il valore atteso del numero dei punti giocati in un tiebreak quando A inizia a servire.

p_B	p_A								
	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7
0.3	73.76	73.11	71.09	67.80	63.70	59.41	55.39	51.87	48.88
0.35	73.97	74.59	73.82	71.47	67.81	63.53	59.28	55.47	52.26
0.4	72.65	74.59	75.39	74.52	71.90	68.08	63.83	59.81	56.36
0.45	69.69	72.69	75.05	75.99	75.04	72.35	68.61	64.65	61.06
0.5	65.49	69.07	72.58	75.24	76.27	75.35	72.81	69.44	66.01
0.55	60.78	64.47	68.53	72.38	75.19	76.27	75.19	73.33	70.59
0.6	56.23	59.76	63.87	68.24	72.21	75.04	76.13	75.64	74.11
0.65	52.24	55.52	59.43	63.83	68.31	72.24	74.96	76.22	76.24
0.7	48.93	52.00	55.66	59.87	64.44	68.94	72.79	75.57	77.21

Tabella 2.6: *Tabella dei valori attesi delle durate (punti totali giocati) di un set al variare di $p_A, p_B \in]0, 1[$.*

Capitolo 3

Confronto con esempi reali

3.1 Analisi degli scontri diretti tra Nadal, Djokovic e Federer

In questa sezione useremo i dati degli incontri realmente disputati affinché possiamo verificare se il modello proposto sia correttamente capace di stimare le probabilità di vittoria di un game e la sua durata attesa, date le stime probabilistiche trovate. I dati reali sono stati ottenuti da www.atptour.com, il quale si occupa punto per punto di aggiornare i risultati e le statistiche delle partite ufficiali del circuito ATP.

Consideriamo ad esempio gli incontri che ci sono stati tra Rafael Nadal, Roger Federer e Novak Djokovic negli anni che vanno dal 2009 al 2014. Questi giocatori hanno avuto una quantità da record di incontri disputati l'uno contro l'altro. Nella Tabella 3.1 analizziamo appunto la dimensione dei dati messi a disposizione.

Giocatori	Incontri	Set	Game
Nadal vs. Djokovic	22	63	610
Nadal vs. Federer	17	46	448
Federer vs. Djokovic	18	51	486

Tabella 3.1: *Numero di scontri diretti, game e set totali disputati tra Nadal, Djokovic e Federer.*

In seguito valutiamo per ogni giocatore al servizio sia il caso di una costante probabilità di vittoria durante l'intero game, sia quello in cui ci sono due diverse probabilità (una prima del *Deuce* e una dopo). Per ognuno di questi tre giocatori calcoleremo le relative

frequenze di vittoria di un punto guardando tutte le partite giocate, e considereremo anche i punti vinti prima del *Deuce* e dopo.

Nelle Tabelle 3.2, 3.3 e 3.4 confrontiamo le percentuali di vittoria di un game basate sui dati reali (denotate con \hat{h}_1) con le probabilità secondo il nostro modello (h_1) calcolate usando la coppia (p, \bar{p}) che più gli si avvicina (valutiamo nella prima riga il caso $p = \bar{p}$ e nella seconda quello in cui $p \neq \bar{p}$). Ricordiamo comunque che $P_{aB}^G = 1 - P_{aA}^G$.

p	\bar{p}	h_1	\hat{h}_1
$p_A = 0.59459$	$\bar{p}_A = 0.59459$	$P_{aA}^G = 0.72434$	0.71237
$p_B = 0.62811$	$\bar{p}_B = 0.62811$	$P_{bB}^G = 0.79119$	0.75563
$p_A = 0.60500$	$\bar{p}_A = 0.57190$	$P_{aA}^G = 0.71529$	0.71237
$p_B = 0.63881$	$\bar{p}_B = 0.59705$	$P_{bB}^G = 0.77704$	0.75563

Tabella 3.2: *Probabilità di vincere un game: Nadal (denotato con A) vs. Djokovic (denotato con B).*

p	\bar{p}	h_1	\hat{h}_1
$p_A = 0.6484$	$\bar{p}_A = 0.6484$	$P_{aA}^G = 0.8271$	0.8288
$p_B = 0.6073$	$\bar{p}_B = 0.6073$	$P_{bB}^G = 0.7507$	0.7257
$p_A = 0.6414$	$\bar{p}_A = 0.6676$	$P_{aA}^G = 0.83373$	0.8288
$p_B = 0.6186$	$\bar{p}_B = 0.5858$	$P_{bB}^G = 0.73789$	0.7275

Tabella 3.3: *Probabilità di vincere un game: Nadal (denotato con A) vs. Federer (denotato con B).*

p	\bar{p}	h_1	\hat{h}_1
$p_A = 0.6287$	$\bar{p}_A = 0.6287$	$P_{aA}^G = 0.7923$	0.7597
$p_B = 0.6449$	$\bar{p}_B = 0.6449$	$P_{bB}^G = 0.8211$	0.8142
$p_A = 0.6357$	$\bar{p}_A = 0.6079$	$P_{aA}^G = 0.7828$	0.7597
$p_B = 0.6186$	$\bar{p}_B = 0.6321$	$P_{bB}^G = 0.8163$	0.8142

Tabella 3.4: *Probabilità di vincere un game: Federer (denotato con A) vs. Djokovic (denotato con B).*

3.2 Valore atteso della durata degli incontri e confronti

In modo simile a quanto fatto in 3.1, stimiamo ora la lunghezza dei game secondo il modello Markoviano visto in 2.2 (denotata con k_1) e la confrontiamo con la durata media negli scontri tra i tre giocatori calcolata usando i dati reali (denotata con \hat{k}_1). I risultati sono visibili nelle Tabelle 3.5, 3.6 e 3.7; i parametri reali sono piuttosto vicini a quelli trovati dal modello, salvo qualche piccolo errore.

Caso	p	\bar{p}	k_1	\hat{k}_1
$p = \bar{p}$	$p_A = 0.59459$	$\bar{p}_A = 0.59459$	$K_{aA}^G = 6.39459$	6.28910
	$p_B = 0.62811$	$\bar{p}_B = 0.62811$	$K_{bB}^G = 6.20935$	5.71368
	$1 - p_A = 0.40541$	$1 - \bar{p}_A = 0.40541$	$K_{aB}^G = 6.81638$	7.01176
	$1 - p_B = 0.37189$	$1 - \bar{p}_B = 0.37189$	$K_{bA}^G = 6.77603$	6.75000
$p \neq \bar{p}$	$p_A = 0.60500$	$\bar{p}_A = 0.57190$	$K_{aA}^G = 6.30753$	6.28910
	$p_B = 0.63881$	$\bar{p}_B = 0.59705$	$K_{bB}^G = 6.12896$	5.71368
	$1 - p_A = 0.39500$	$1 - \bar{p}_A = 0.42810$	$K_{aB}^G = 6.99899$	7.01176
	$1 - p_B = 0.36119$	$1 - \bar{p}_B = 0.40295$	$K_{bA}^G = 7.02878$	6.75000

Tabella 3.5: *Confronto del numero di punti giocati in un game per Nadal(A) vs Djokovic(B).*

Caso	p	\bar{p}	k_1	\hat{k}_1
$p = \bar{p}$	$p_A = 0.6484$	$\bar{p}_A = 0.6484$	$K_{aA}^G = 6.0857$	5.9185
	$p_B = 0.6073$	$\bar{p}_B = 0.6073$	$K_{bB}^G = 6.3274$	6.1220
	$1 - p_A = 0.3516$	$1 - \bar{p}_A = 0.3516$	$K_{aB}^G = 6.7383$	6.9737
	$1 - p_B = 0.3927$	$1 - \bar{p}_B = 0.3927$	$K_{bA}^G = 6.8045$	6.9180
$p \neq \bar{p}$	$p_A = 0.6414$	$\bar{p}_A = 0.6676$	$K_{aA}^G = 6.1207$	5.9185
	$p_B = 0.6186$	$\bar{p}_B = 0.5858$	$K_{bB}^G = 6.2353$	6.1220
	$1 - p_A = 0.3586$	$1 - \bar{p}_A = 0.3324$	$K_{aB}^G = 6.5629$	6.9737
	$1 - p_B = 0.3814$	$1 - \bar{p}_B = 0.4142$	$K_{bA}^G = 7.0260$	6.9180

Tabella 3.6: *Confronto del numero di punti giocati in un game per Nadal(A) vs Federer(B).*

Caso	p	\bar{p}	k_1	\hat{k}_1
$p = \bar{p}$	$p_A = 0.6287$	$\bar{p}_A = 0.6287$	$K_{aA}^G = 6.2059$	5.8000
	$p_B = 0.6449$	$\bar{p}_B = 0.6449$	$K_{bB}^G = 6.1078$	6.0197
	$1 - p_A = 0.3713$	$1 - \bar{p}_A = 0.3713$	$K_{aB}^G = 6.7751$	6.7636
	$1 - p_B = 0.3551$	$1 - \bar{p}_B = 0.3551$	$K_{bA}^G = 6.7455$	7.0435
$p \neq \bar{p}$	$p_A = 0.6357$	$\bar{p}_A = 0.6079$	$K_{aA}^G = 6.1556$	5.8000
	$p_B = 0.6186$	$\bar{p}_B = 0.6321$	$K_{bB}^G = 6.0784$	6.0197
	$1 - p_A = 0.3643$	$1 - \bar{p}_A = 0.3921$	$K_{aB}^G = 6.9464$	6.7636
	$1 - p_B = 0.3814$	$1 - \bar{p}_B = 0.3679$	$K_{bA}^G = 6.8593$	7.0435

Tabella 3.7: *Confronto del numero di punti giocati in un game per Federer(A) vs Djokovic(B).*

Infine confrontiamo il valore della durata del set dato il giocatore che inizia a servire per primo. Usando il nostro modello, però, non possiamo confrontare i set terminati con un tiebreak, quindi useremo come dati reali tutti quelli che non sono finiti con il punteggio di (7,6) oppure (6,7). Nelle tabelle 3.8, 3.9 e 3.10 valutiamo il valore atteso dei punti giocati in un set ($k_{(p,\bar{p})}$ e $k_{(p,p)}$) e la media del numero di punti ottenute dai dati reali (\hat{k}). È da notare che, in questa valutazione, quando $p \neq \bar{p}$ il nostro modello si avvicina molto ai dati reali.

	\hat{k}	$k_{(p,p)}$	$k_{(p,\bar{p})}$
aA	49.50	59.75	59.95
aB	58.89	60.55	60.69
bB	56.02	58.13	58.51
bA	64.67	62.19	62.10

Tabella 3.8: *Confronto del numero di punti giocati in un set tra Nadal(A) e Djokovic(B), dato il giocatore che inizia a servire (indicato con la lettera minuscola).*

3.3 Commenti e conclusioni

In conclusione, il nostro studio ha evidenziato una correlazione piuttosto significativa tra dati reali e dati ricavati dalla modellizzazione analizzata. Ovviamente ci sono delle discrepanze in alcuni confronti singoli, come nel caso di Nadal-Federer per la durata at-

	\hat{k}	$k_{(p,p)}$	$k_{(p,\bar{p})}$
aA	58.37	57.32	57.51
aB	58.02	61.86	61.84
bB	50.96	59.15	59.36
bA	59.80	60.00	59.98

Tabella 3.9: *Confronto del numero di punti giocati in un set tra Nadal(A) e Federer(B), dato il giocatore che inizia a servire (indicato con la lettera minuscola).*

	\hat{k}	$k_{(p,p)}$	$k_{(p,\bar{p})}$
aA	58.37	58.32	58.39
aB	58.02	60.45	60.29
bB	50.96	57.63	57.62
bA	59.80	61.15	61.12

Tabella 3.10: *Confronto del numero di punti giocati in un set tra Federer(A) e Djokovic(B), dato il giocatore che inizia a servire (indicato con la lettera minuscola).*

tesa del game nel caso aA (Tabella 3.6) e di Nadal-Djokovic per la durata attesa del set nello stesso caso (Tabella 3.8). Queste alterazioni sono dovute in parte a un numero di dati reali che in quelle analisi non risulta elevatissimo e in parte anche ad alcune variabili esterne riguardo i singoli match che non abbiamo analizzato (condizioni atmosferiche degli incontri, quanti match si sono disputati su una determinata superficie piuttosto che su un'altra). Il lavoro viene ciononostante valorizzato dai riscontri soprattutto quando confrontato con incontri tra giocatori di tennis professionisti. A questi livelli, infatti, l'eccezionale tenuta fisica e mentale degli sportivi concede poca rilevanza al fattore psicologico e alla determinata situazione di punteggio (ad esempio il fatto di essere nel terzo set di gioco invece che nel primo).

La nostra analisi quindi risulta realistica, nonostante siano state fatte delle generalizzazioni sui parametri delle probabilità. Queste sono state assunte per ragioni di brevità di calcoli e semplicità computazionale. È stato comunque utile, però, separare i punteggi antecedenti la *Parità* e quelli successivi. Infatti, come si evince anche dalle Tabelle precedenti, per esempio, si nota una differenza che risulta notevole tra le durate attese di un set considerando le probabilità prima costanti e poi differenziate pre e post *Deuce*.

Bibliografia

- [1] J. R. Norris, *Markov chains*. Cambridge university press, 1998.
- [2] International Tennis Federation, *2026 ITF Rules of Tennis*, ITF Ltd, London, 2026.
- [3] A. Carrari, M. Ferrante, G. Fonseca et al., «A new markovian model for tennis matches,» *Electronic Journal of Applied Statistical Analysis*, vol. 10, n. 3, pp. 693–711, 2017.
- [4] P. K. Newton e J. B. Keller, «Probability of winning at tennis I. Theory and data,» *Studies in applied Mathematics*, vol. 114, n. 3, pp. 241–269, 2005.
- [5] J. G. Kemeny e J. L. Snell, «Finite markov chains, undergraduate texts in mathematics,» 1976.

Ringraziamenti

Desidero innanzitutto ringraziare il mio relatore, il Professor Andrea Pascucci, per la disponibilità, i preziosi consigli e la guida attenta durante la stesura di questo elaborato.

Un ringraziamento speciale va alla mia famiglia e a mio padre, per gli sforzi fatti e per avermi permesso di raggiungere questo traguardo.

Grazie ai miei amici di sempre, ai compagni di università incontrati qui all'Alma Mater e ai coinquilini di questi anni: Francesco, Mirko e Giovanni. Insieme abbiamo condiviso eventi belli e brutti, la fatica dello studio, le ansie degli esami, ma soprattutto momenti indimenticabili che porterò sempre con me. Grazie anche per avermi sopportato.

Infine, grazie a tutte le persone che, in un modo o nell'altro, hanno fatto parte di questo mio percorso.