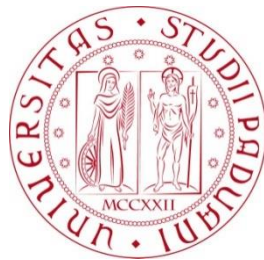


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze Statistiche



**I punti nel tennis sono indipendenti e identicamente distribuiti?**

Relatore Prof. Francesco Lisi

Dipartimento di Scienze Statistiche

Laureando: Francesco Matteazzi

Matricola N 1096158

Anno Accademico 2016/2017



# Indice

Introduzione.....	1
1. Il contesto generale.....	5
1.1. Sport Analytics.....	5
1.2. I punti sono indipendenti e identicamente distribuiti?.....	7
1.3. Letteratura.....	8
2. Gli stati del match.....	11
2.1. Introduzione.....	11
2.2. Individuazione degli Stati.....	12
3. I dati.....	15
3.1. Point by Point Data.....	15
3.2. La scelta dei dati.....	16
4. Analisi.....	20
4.1. Analisi Preliminari.....	20
4.1.1. Ranking.....	20
4.1.2. Runs Test.....	23
4.2. L'approccio parametrico.....	27
4.3. L'approccio non parametrico.....	33
4.3.1. Le probabilità stimate.....	34
4.3.2. Simulazioni Monte Carlo.....	43
Conclusioni.....	65
Appendice A.....	69
Bibliografia	



# Introduzione

“The fifth set is not about tennis, it’s about nerves”. Con queste parole, il famoso tennista Boris Becker, sottolinea un aspetto fondamentale proprio non solo del tennis ma di tutta la realtà sportiva: nei momenti chiave di un incontro, più che le qualità tecniche e le competenze tattiche servono qualità psicologiche e solidità mentale per arrivare ad ottenere il successo. Qualsiasi sportivo sa bene che, la capacità di non farsi influenzare da quello che è successo negli istanti precedenti e rimanere concentrato solamente sull’azione corrente, è sicuramente una delle chiavi per poter raggiungere una vittoria. Tuttavia, quanto più questa capacità è fondamentale, tanto più è difficile estraniarsi dal passato quando si è consapevoli che ogni sbaglio potrebbe condizionare negativamente il risultato finale. A partire dal ragionamento di qui sopra sorgono diverse domande. La prima riguarda l’identificazione di questi momenti chiave e, di conseguenza, l’identificazione di atteggiamenti differenti da parte degli atleti coinvolti nel giocarli. La seconda è relativa invece al modo in cui sia possibile cogliere tutto ciò per mezzo di strumenti statistici.

Da un punto di vista statistico la traduzione di quanto fino ad ora affermato è riassumibile nella domanda: i momenti di una partita sono tra loro indipendenti o sono legati da una qualche forma di dipendenza? Calandosi nella realtà tennistica, questo può essere riformulato come segue: i punti giocati durante un incontro sono tra loro indipendenti e identicamente distribuiti o sono affetti da una forma di dipendenza? Statisticamente parlando, ipotizzare una che i punti di un incontro siano tra loro indipendenti vuol dire porre un’assunzione estremamente forte circa la natura dei punti stessi. Si sta infatti supponendo che non esista alcun tipo di legame né tra due scambi successivi né tra due scambi qualunque giocati all’interno di una partita.

Da un punto di vista sportivo, questa assunzione potrebbe essere spiegata prendendo ad esempio un giocatore totalmente razionale, in grado di giocare ogni scambio esattamente come quello precedente senza farsi influenzare dal momento, indipendentemente dal fatto che sia il primo o l'ultimo punto della partita e indipendentemente dal fatto che si stia disputando un torneo minore o la finale di Wimbledon. Dovendo esprimere un'opinione a priori, sarebbe per certi versi sorprendente che un'assunzione così forte risultasse statisticamente valida per tutti i giocatori. Non sarebbe però errato credere che una tale ipotesi possa risultare valida solo per alcuni giocatori, in particolare per quelli più avvezzi a giocare con abitudine certe partite e certi punti rilevanti; in sintesi, per i migliori giocatori al mondo.

Questo lavoro prende le mosse dalla volontà di coniugare una profonda passione per lo Sport con il percorso di studi affrontato. Negli ultimi anni, la sempre maggiore attenzione alla performance sportiva ha trovato nuovo terreno fertile nella crescente capacità di raccogliere, immagazzinare e soprattutto interpretare i dati. Da questo punto di vista, la Sport Analytics, ovvero il nuovo connubio tra Sport e Statistica, rappresenta non solo una delle grandi novità degli ultimi decenni ma anche una grande prospettiva per il futuro. L'attenzione al dettaglio statistico sta interessando trasversalmente tutto il panorama sportivo, dalla pallacanestro al calcio fino appunto al tennis, e sta avendo fortissimo utilizzo in diversi rami della pratica sportiva: dalla preparazione tattica di squadra al miglioramento tecnico individuale, dal monitoraggio delle pratiche di recupero dagli infortuni, fino al puro interesse accademico senza dimenticare le grandi opportunità fornite dal mercato delle scommesse sportive.

Riassumendo, lo scopo di questa tesi è quello di verificare l'affermazione di indipendenza e identica distribuzione dei punti giocati, in una partita di tennis, da diverse coppie di giocatori. Per fare questo si è fissato come punto di partenza l'identificazione di una serie di stati del match, per ognuno dei quali

è stata calcolata la probabilità di vincere il punto al servizio. Sotto l'ipotesi di indipendenza, lo stato del match considera esclusivamente il servizio dei due giocatori. Per confrontare le varie probabilità calcolate nei vari stati con la probabilità di riferimento si è fatto ricorso a due approcci. Dapprima un approccio parametrico nella forma di un modello di regressione logistica, in secondo luogo un approccio non parametrico nella forma di diverse simulazioni Monte Carlo implementate nei vari stati del match.

L'elaborato è costituito da quattro capitoli. Il Capitolo 1 si prefigge di introdurre il problema dando una rapida descrizione del contesto e della letteratura precedente. Nel Capitolo 2 verrà illustrato il concetto di stato del match e saranno esposti i criteri che hanno portato alla definizione degli stessi. Il Capitolo 3 presenterà, invece, i dati dandone prima una descrizione generale per poi presentarli nel dettaglio. Infine, con il Capitolo 4 verranno esposte le analisi condotte: dapprima le semplici analisi preliminari sui dati e, a seguire, verranno presentati gli approcci parametrici e non parametrici adottati per affrontare il problema in esame. Chiuderanno l'elaborato le Conclusioni e un'Appendice in cui verranno esposte brevemente le regole e alcuni termini tecnici del gioco del tennis.





# Capitolo 1

## Il Contesto Generale

Lo scopo di questo capitolo è di fornire una rapida descrizione del concetto di Sport Analytics al fine di chiarire il contesto nel quale questo elaborato si inserisce. Il secondo paragrafo introdurrà il problema analizzandolo sia da un punto di vista sportivo sia con un'ottica statistica. Infine, il terzo paragrafo presenterà un excursus della precedente letteratura concernente il tema in esame allo scopo di definire le basi su cui poggia questo elaborato e le novità introdotte dallo stesso rispetto agli autori precedenti.

### **1.1. Sport Analytics**

Così come gli uomini che ne costituiscono l'essenza, anche lo Sport è in continuo mutamento. In particolare, ciò che si sta affermando negli ultimi decenni a livello professionistico è una spasmodica e ossessiva ricerca del risultato. La volontà di primeggiare su rivali e avversari sta portando ad una maniacale attenzione al dettaglio. L'espressione "I dettagli fanno la differenza", non è certo una frase fatta quando ci si riferisce all'ambito sportivo. In un contesto in cui la competitività è a livelli estremi non è concesso sbagliare, per quanto piccolo possa essere l'errore, se si vuole ottenere la vittoria. Questa crescente attenzione al particolare, allo studio degli avversari e di stessi ha trovato, e sta trovando, terreno fertile nel contesto della Sport Analytics. La possibilità di disporre di grandi quantità di dati, inerenti qualsiasi aspetto legato alla pratica sportiva, ha spinto ad uno uso sempre maggiore della statistica e degli strumenti di cui essa dispone.

Come tutte le grandi novità, anche l'avvento della Sport Analytics non è stato semplice ed immediato; quando Billy Beane, general manager degli Oakland Athletics, decise di ricorrere a dei modelli statistici per assemblare la sua squadra di baseball fu considerato un pazzo. A distanza di pochi mesi, dopo una cavalcata di venti vittorie consecutive la più lunga della storia della MLB, la sua scelta venne totalmente rivalutata. La sua storia, raccontata nel libro *Moneyball: The art of winning an unfair game* di Michael Lewis (2003), ha aperto la strada a quella che ora si sta affermando sempre di più come una consuetudine.

In ambito tennistico, l'espressione che viene utilizzata per indicare l'applicazione di strumenti statistici e analitici al gioco è Tennis Analytics. Così come per gli altri sport, anche le applicazioni relative al tennis sono molteplici. Valutare come si muove un avversario, quando preferisce giocare un dritto o un rovescio, quali sono le sue traiettorie preferite, sono informazioni che possono risultare fondamentali sia nella preparazione di un incontro sia durante una partita per analizzarne in tempo reale l'andamento. Allo stesso modo recuperare queste informazioni su se stessi è altrettanto fondamentale per correggere alcuni punti deboli del proprio gioco. Un'altra diretta ed immediata applicazione di strumenti statistici concerne la previsione del risultato finale di un incontro. Il mercato delle scommesse è infatti indissolubilmente legato alla pratica sportiva professionistica, e la capacità di analizzare al meglio i dati raccolti è sicuramente requisito fondamentale per i bookmakers. Infine, l'analisi statistica dei dati può aiutare ad indagare ancora più a fondo un incontro provando a cogliere nei numeri gli aspetti psicologici che caratterizzano una partita il che risulta essere vero soprattutto nel caso del tennis, sport in cui la componente mentale gioca un ruolo determinante. Si consideri, ad esempio, la finale dell'ultimo torneo di Wimbledon; dopo circa 1 ora e 40 minuti di gioco, Roger Federer si trova a giocare il match point sul suo servizio. Dall'altro lato del campo il suo avversario, Marin Cilic, non ha più alcun

marginale d'errore. Da un punto di vista emotivo una situazione di questo genere risulta essere diametralmente opposta rispetto a quella vissuta durante il primo scambio della stessa finale. Ma tutto questo si rispecchia nei dati? Federer avrà servito quell'ultimo punto come aveva servito durante tutto l'arco dell'incontro? All'opposto, il fatto di ritrovarsi psicologicamente e sportivamente con le spalle al muro avrà avuto un qualche effetto statistico su quella che è stata la risposta di Cilic?

## **1.2. I punti sono indipendenti e identicamente distribuiti?**

Riprendendo l'esempio precedente, affermare che Cilic, così come Federer, abbia giocato l'ultimo punto dell'incontro senza farsi influenzare dalla situazione corrente significa affermare che non esiste, da un punto di vista statistico, alcuna forma di dipendenza tra due punti giocati all'interno di una partita. Assumere che i punti di un incontro siano indipendenti e identicamente distribuiti significa porre una caratterizzazione molto forte circa la natura dei punti stessi, sia da un punto di vista statistico che sportivo. Può sembrare piuttosto ovvio che i punti giocati all'interno di una partita siano tra loro diversi ma assumendoli indipendenti e identicamente distribuiti, ciò che si vuole verificare è se, e quanto, queste loro differenze risultino rilevanti. L'assunzione i.i.d. è sicuramente una semplificazione di un processo più complesso ma può comunque essere considerata una buona approssimazione? Assumere, infine, una siffatta relazione tra i punti significa assumere che la probabilità di vincere un punto rimane costante durante l'intero incontro. Le implicazioni sportive di una tale assunzione seguono dalle considerazioni sopra riportate. In primo luogo, significa assumere che non vi sia alcuna differenza di approccio alla partita, da parte dei due giocatori, sia che si stia giocando la finale di Wimbledon sia che si stia giocando il primo turno di un torneo minore. Significa, assumere che non vi siano differenze significative nell'approccio alle

varie situazioni che occorrono durante una partita. Ovvero i giocatori non cambiano il loro approccio al gioco mantenendolo costante dal primo scambio fino all'ultimo match point. Significa, soprattutto, che ogni giocatore non si fa influenzare né da quello che è successo prima né da quello che potrebbe succedere dopo. Un'espressione come "Pensare punto per punto" può risultare un cliché ma, qualunque sportivo sa che essa rappresenta uno status mentale vincente. La capacità di vivere esclusivamente il momento presente è sintomo di grande solidità mentale e forza psicologica e rappresenta, di conseguenza, una delle chiavi del successo. Assumere che questo atteggiamento risulti proprio di tutti i giocatori potrebbe essere per certi versi sorprendente ma potrebbe aiutare anche a capire meglio il gioco e la sua stessa evoluzione.

### **1.3. La letteratura**

L'analisi matematica e statistica inerente il tennis si fonda dunque sull'assunzione che i punti nel siano tra loro indipendenti e identicamente distribuiti. Sul fatto che quest'assunzione sia accettabile o meno si è già discusso in letteratura nonostante non siano presenti molti studi a riguardo. Alcuni di questi studi non affrontano direttamente il problema della struttura i.i.d. dei punti ma considerano diversi aspetti ad essa collegati e che in essa possono riflettersi. Altri, invece, considerano direttamente l'assunzione di indipendenza provando a verificarla per mezzo di modelli o test.

Alla prima categoria di studi è ascrivibile la proposta di Knight e O' Donoghue (2012) relativa all'importanza dei break points. Knight e O' Donoghue, confrontando situazioni di break points e non, concludono che la probabilità di vincere il punto sulla risposta subisce un incremento in situazioni di break points. I due mettono così in luce come in situazioni diverse si assista a

differenze in termini di probabilità di vincere il punto indicando dunque possibili deviazioni rispetto all'assunzione di indipendenza. Sulla stessa scia si inseriscono gli studi di Klaassen e Magnus, relativi al torneo di Wimbledon, inerenti dapprima l'eventuale vantaggio ottenuto dal servire per primi nel set (1999) e l'effetto che giocare l'ultimo decisivo set ha sulle probabilità di vincere il match (1999). Ancora una volta la volontà è di quella di isolare particolari situazioni di gioco allo scopo di valutare se in questi particolari contesti si assiste o meno a deviazioni rispetto l'iniziale assunzione di indipendenza. Ancora, Klaassen e Magnus hanno provato a valutare l'effetto che può avere il cambio delle palle durante l'incontro (1999), basandosi sui dati relativi al torneo di Wimbledon. Altro aspetto estremamente rilevante nel tennis e possibile discriminante di deviazioni rispetto all'indipendenza è rappresentato dalla superficie su cui un particolare incontro è disputato. È risaputo, infatti, che le proprietà tipiche di ciascuna superficie possono o meno favorire un giocatore a seconda dello stile di gioco e delle caratteristiche fisiche e tecniche dello stesso. Si deve a O' Donoghue (2013) il tentativo di tener conto di questo particolare effetto esclusivamente legato al tennis. Infine, un'interessante proposta in letteratura si deve a Morris (1977) e alla volontà di distinguere i punti sulla base di una variabile definita, dallo stesso Morris, importance. Il concetto di importance si ripete in molti dei successivi studi relativi all'ipotesi di i.i.d. ed è un importante assunto di questo stesso elaborato.

Per quel che concerne gli studi che affrontano direttamente la questione dell'indipendenza dei punti, va innanzitutto segnalato l'elaborato di Klaassen e Magnus (2001) relativo alla stima di un modello per dati di panel allo scopo di verificare l'esistenza o meno di un effetto del punto precedente su quello corrente. Perseguendo lo stesso scopo ma utilizzando diversi approcci per testare l'ipotesi di indipendenza, Pollard e Pollard (2011) hanno affrontato il problema prendendo in considerazione undici incontri giocati nel 2011 da Nadal nei tornei del Grande Slam. Sia Klaassen e Magnus, sia Pollard e Pollard

sono giunti alla conclusione che sono presenti evidenze statistiche di deviazione rispetto all'ipotesi di indipendenza ma che queste deviazioni risultano essere per lo più trascurabili rendendo comunque l'ipotesi i.i.d. un'approssimazione accettabile.

Altri lavori che sono stati tenuti in considerazione nella produzione di questo elaborato sono sicuramente il libro *Analyzing Wimbledon* (2004) di Klaassen e Magnus che raccoglie e amplia i già citati lavori precedenti; l'articolo di McHale e Morton (2011) relativo alla stima di un modello di Bradley-Terry per la previsione del risultato di un incontro. Per quanto riguarda gli aspetti di natura matematica concernenti la derivazione delle formule analitiche per il calcolo dell'importance, sono stati punti di riferimento gli elaborati di Madurska (2012), Barnett (2006), Clowes, Cohen e Tomljanovic (2007).

Questo elaborato, inserendosi in un sentiero già battuto, fonda le sue basi sulla letteratura sopra citata. Partendo da una così solida base, l'obiettivo di questo testo è stato quello di provare a fornire nuove indicazioni relative all'assunzione di i.i.d. differenziandosi però dai precedenti studi. In questo senso si è deciso di condurre analisi diverse rispetto a quelle precedentemente svolte prendendo in considerazione diverse situazioni all'interno di un incontro. Infine si è cercato di considerare nuovi dati che permettessero di ampliare il più possibile lo spettro degli scenari disponibili.

# Capitolo 2

## Gli Stati del Match

In questo capitolo verrà presentato uno degli elementi di innovazione rispetto a quanto è già stato fatto in letteratura: la definizione di stato del match. Nel primo paragrafo verrà data una definizione di che cosa sia uno stato del match, mentre nel secondo paragrafo verranno presentati i vari stati presi in considerazione in questo elaborato.

### **2.1. Introduzione**

Ragionando da un punto di vista sportivo è indiscutibilmente chiaro come all'interno di un incontro alcuni momenti risultino essere decisamente più importanti rispetto ad altri. Chiaramente, per quel che concerne il tennis, match e set points rappresentano l'esempio più chiaro e concreto di questo ma di certo non sono gli unici momenti chiave. A maggior ragione in un sport come il tennis, in cui la sfida tra due avversari si gioca non solo su un piano tecnico tattico e fisico ma anche e, forse, soprattutto mentale sono molteplici le fasi che possono risultare decisive nello spostare l'inerzia di una partita. Solo per fare alcuni esempi, riuscire a strappare un game sul servizio dell'avversario o riuscire a vincere una serie consecutiva di punti sono azioni che possono rivelarsi decisive nell'arco di un incontro. Sulla base di queste considerazioni è dunque lecito chiedersi se in certe situazioni particolari di gioco l'atteggiamento dei due sfidanti rimanga lo stesso rispetto ad altre meno rilevanti. In termini statistici, la probabilità di vincere un punto al servizio risulta essere statisticamente diversa a seconda delle varie situazioni di gioco?

Con lo scopo di evidenziare deviazioni dall'ipotesi di indipendenza e identica distribuzione si è cercato, dunque, di individuare alcuni momenti topici all'interno di una partita nei quali queste stesse deviazioni si possono verificare. Nel prosieguo di questo elaborato, con l'espressione stato del match si intenderà dunque una particolare situazione occorsa all'interno del match stesso.

## 2.2. Individuazione degli Stati

Il punto di partenza nella definizione degli stati del match è stata la letteratura. Sono stati, infatti, ripresi quanti di questi particolari momenti erano già stati considerati e ad essi ne sono stati aggiunti di nuovi con l'obiettivo di ottenere una visione il più ampia possibile su tutte le possibili situazioni cardine di un incontro di tennis. Nella Tabella 1, sono elencati gli stati del match considerati.

STATO del MATCH	NOME
Game-Points (break-, set-, match-points)	GP
Tie-break Points	TB
Break- o Set- Points	IP1
Break o Set- o Tie-break Points	IP2
Tie-break o Game-Points	IP3
Set	Set
Previous Point	PP
Game-Points e Previous Points	GPP
Break- o Set- Points e Previous Points	BSPP
Tie-break o Game-Points e Previous Points	TBGPP
Break o Set- o Tie-break Points e Previous Points	TBBSPP
Importance	IMP

*Tabella 1. Stati del Match*

Partendo dalla classificazione riportate in Tabella 1, si può asserire come siano stati considerati tutti i momenti caratterizzanti un incontro di tennis. Una prima e fondamentale discriminante nella definizione degli stati riguarda il servizio.



Dato che lo scopo di questo elaborato è quello di valutare l'ipotesi di i.i.d. partendo dalla probabilità di vincere un punto al servizio, ciascuno degli stati sopra elencati considera i soli punti giocati al servizio, dal giocatore di riferimento, all'interno dello stato stesso. Partendo da questi assunti si è arrivati alla definizione di dodici stati che in alcuni casi non risultano tra loro totalmente disgiunti. Si consideri ad esempio il caso di IP1 e IP2: IP1 considera tutti i punti che sono Game-Points includendo, di conseguenza, tutti i Break-, i Set- e i Match-Points. Al contrario IP2 si limita alla sola identificazione di Break- e Set-Points. Questo aspetto risulta ancora più marcato se si prendono in considerazione gli stati GPP, BSPP, TBGPP, TBBSPP. La scelta di definire in questo modo gli stati è stata dettata dalla necessità di analizzare il più dettagliatamente possibile le varie situazioni di gioco, considerandole dapprima singolarmente ed in seguito in modo aggregato per valutare la permanenza o meno di eventuali deviazioni rispetto all'ipotesi di indipendenza. Operativamente, per definire gli stati, per ogni punto giocato è stata considerata una variabile dicotomica tale da assumere valore 1 se il punto corrente è stato giocato nello stato di riferimento. Per fare un esempio:

$$GP_i = \begin{cases} 1 & \text{se } i - \text{esimo punto è un game point} \\ 0 & \text{altrimenti} \end{cases}$$

Anche per lo stato denominato Imp è stata considerata una variabile dicotomica la cui costruzione ha richiesto però alcuni passaggi supplementari. Dapprima è stata calcolata l'importance, seguendo la definizione di Morris (1977). Sia  $g$  la probabilità di vincere un game condizionatamente al servizio e al punteggio corrente, siano inoltre  $s$  ed  $m$  le probabilità di vincere, rispettivamente, un set e un match condizionatamente al punteggio. Infine, siano  $x$ ,  $y$  le situazioni di punteggio dei giocatori A e B. Sulla base di ciò si definiscono le variabili: importance of point in game,  $imp_{ps}$ , importance of

game in set,  $imp_{gs}$ , importance of set in match,  $imp_{sm}$ , e importance of point in match  $imp$ . Nello specifico:

$$imp_{ps} = g(x + 1, y) - g(x, y + 1)$$

$$imp_{gs} = s(x + 1, y) - s(x, y + 1)$$

$$imp_{sm} = m(x + 1, y) - m(x, y + 1)$$

$$imp = imp_{ps} * imp_{gs} * imp_{sm}$$

Basandosi poi sulla distribuzione dei valori di  $imp$  ottenuti, per ogni giocatore di ciascuna coppia considerata, si è fissata una soglia pari al terzo quantile della distribuzione dell'importanza stessa. Infine è stata definita una variabile dicotomica tale da assumere valore 1 per punti caratterizzati da un valore di importance maggiore rispetto alla soglia, 0 altrimenti.

$$Imp_i = \begin{cases} 1 & \text{se importance } i - \text{esimo punto} > \text{soglia} \\ 0 & \text{altrimenti} \end{cases}$$

Per quanto concerne, infine, lo stato denominato Set, è stata considerata una variabile categoriale con valori da 1 a 5, nel caso di partite al meglio dei cinque set, o da 1 a 3, nel caso di partite al meglio dei tre set, in modo da considerare separatamente i punti giocati nei diversi set di una partita.

# Capitolo 3

## I Dati

In questo capitolo verranno presentati i dati utilizzati in questo elaborato. Dopo una breve presentazione generale riguardante la particolare natura dei dati cui si è fatto ricorso, seguiranno le motivazioni che hanno portato alla scelta dei dati utilizzati e una presentazione dettagliata delle coppie di giocatori sui cui sono state implementate le analisi.

### **3.1. Point by Point Data**

La volontà di analizzare e cogliere differenti strutture dei punti all'interno di un incontro, porta con sé la necessità di far ricorso ad una particolare forma di dati. A differenza di quanto accade, per esempio, nel mondo delle scommesse sportive dove per elaborare un modello è sufficiente conoscere il vincitore di una partita, l'individuazione di eventuali deviazioni dall'ipotesi di i.i.d. richiede un maggiore dettaglio nei dati. È necessario, infatti, far ricorso a dati point by point; detto altrimenti, all'intera sequenza dei punti giocati.

Ancora una volta, questo approccio si fonda su quanto è stato già fatto in letteratura. Gli studi di Magnus e Klaassen si basano infatti su un dataset di dati punto per punto relativi al torneo di Wimbledon, sia maschile che femminile, nel periodo 1992-1995. Allo stesso modo, Pollard e Pollard, hanno condotto le loro analisi basandosi sui dati punto per punto relativi ad undici incontri disputati da Nadal contro diversi giocatori, nel 2011, nei quattro tornei del Grande Slam. Dati punto per punto sono stati, infine, utilizzati sia da Knight e

O'Donoghue, riferiti ai quattro tornei del Grande Slam giocati tra il 2008 e il 2009, sia da O'Donoghue, relativi ai tornei del Grande Slam disputati nel 2011.

Per quel che concerne questo elaborato si è deciso di recuperare i dati dal software OnCourt. OnCourt, oltre a fornire una gran quantità di statistiche e di informazioni utili ad una prima analisi descrittiva, permette anche di recuperare l'intera sequenza dei punti giocati in una partita.

### **3.2. La scelta dei dati**

Sulla base delle considerazioni fatte, si può notare come la letteratura precedente si riferisca, nella scelta dei dati, a contesti omogenei e ben definiti. Le analisi precedentemente citate si riferiscono, infatti o ad un singolo torneo o ai soli tornei del Grande Slam senza considerare partite disputate nei tornei minori. Un altro aspetto comune è la mancanza d'interesse nei confronti di chi gioca gli incontri inclusi nei dataset; l'attenzione non cade tanto sui giocatori in se quanto piuttosto sulla qualità che li caratterizza (Magnus e Klaassen). Pur fondandosi sulla letteratura, questo elaborato cerca di svincolarsi dai precedenti approcci e uno dei primi punti di differenza riguarda appunto la scelta dei dati.

Il primo, e fondamentale, elemento di differenza rispetto ai precedenti elaborati riguarda la scelta di concentrarsi non tanto su un singolo, o su pochi tornei, bensì di focalizzarsi su specifiche coppie di giocatori. Dati, dunque, due giocatori è stato considerato lo storico dei loro scontri diretti. Questa scelta, oltre a rappresentare un nuovo approccio al problema, porta con se diverse implicazioni: in primo luogo, considerare due specifici giocatori non rende più necessaria l'introduzione di variabili atte a cogliere la qualità dei giocatori stessi. Considerare, poi le sequenze dei testa a testa, permette di riferirsi ad un vasto paniere di situazioni; per ogni sequenza infatti si contano partite tra loro

molto differenti in termini di contesto di gioco. I dati fanno, infatti, riferimento a decine di tornei differenti, che siano tornei del Grande Slam, ATP500 o ATP1000, includendo partite al meglio dei tre e dei cinque set e prendendo in considerazione, soprattutto, diverse superfici di gioco. Ancora, un'ulteriore tentativo di innovazione rispetto alla precedente letteratura consiste nel dar maggior enfasi al tennis femminile al fine di cogliere eventuali differenze rispetto alla realtà maschile.

Sulla base delle considerazioni fatte, si è dunque cercato di prendere in considerazione tutte quelle coppie di giocatori tali da coniugare al meglio questa ricerca di variabilità nei dati con le esigenze statistiche necessarie in un'ottica di analisi. Se, infatti, da un punto di vista statistico questa variabilità nei dati rappresenta di certo un vantaggio, è altrettanto vero che può indurre a diverse complicazioni. La prima, fondamentale, riguarda il numero di dati a disposizione. Non sempre infatti, le sequenze punto per punto considerate, racchiudono un numero elevato di incontri disputati e di conseguenza, di punti giocati. Questo risulta per lo più vero quando almeno uno dei due giocatori della coppia analizzata, è di livello medio basso. Al contrario, i giocatori all'apice delle classifiche di Ranking, vantano una storia di incontri ben più nutrita. Nonostante questo aspetto possa, senza dubbio, rappresentare una forte limitazione si è optato comunque per considerare, a dispetto della dimensione campionaria, anche queste particolari coppie di giocatori. Questa decisione è stata dettata dalla necessità di non limitarsi ad osservare i soli giocatori di alto livello. Come, infatti è già stato osservato, è indubbio che i migliori giocatori, e le migliori giocatrici, al mondo siano in grado di giocare con maggiore regolarità di altri; ma questo è altrettanto vero per gli altri giocatori? Considerare, infine, giocatori differenti in termini di qualità è necessario se si vuol rispondere a questo quesito.

Basandosi su quanto fino ad ora affermato si è arrivati a considerare trentatré coppie, di cui venti maschili e tredici femminili, riassuntive di 280 partite e 45.867 punti giocati da quarantotto giocatori, ventotto uomini e venti donne.

	<b>Uomini</b>	<b>Donne</b>
<b>Sequenze head-to-head</b>	20	13
<b>Numero di Giocatori</b>	28	20
<b>Numero di Partite</b>	190	90
<b>Numero di Punti</b>	33243	12624

*Tabella 2. Head-to-head*

Di seguito, sono riportate le tabelle con i dettagli degli scontri diretti considerati.

<b>Coppia</b>	<b>Numero di Partite</b>	<b>Numero di Punti</b>
Kerber – Pliskova	7	1030
Halep – Kuznetsova	7	1013
Radwanska – Kerber	11	1573
S. Williams – Sharapova	12	1475
Wozniacki – Cibulkova	10	1429
Errani – Cornet	6	952
Cibulkova – Kvitova	7	836
Giorgi – Pliskova	4	594
V. Williams – Stosur	5	804
Keys – Radwanska	4	518
V. Williams – S. Williams	8	1170
Gavrilova – Garcia	4	535
S. Zhang - Konta	5	695

*Tabella 3. Dettaglio head-to-head donne*

<b>Coppia</b>	<b>Numero di Partite</b>	<b>Numero di Punti</b>
Djokovic – Federer	35	6553
Federer – Nadal	18	2918
Berdych – Ferrer	9	1551
Del Potro – Federer	17	3325
Federer – Ferrer	8	1273
Nadal – Fognini	12	2115
Goffin – Tsonga	6	1001
Tipsarevic – Dimitrov	4	582
Verdasco – Lopez	4	660
Seppi – Haase	6	1011
Seppi – Muller	5	857
Struff – Kohlschreiber	4	672
Herbert – Struff	4	597
Isner – Lopez	7	1523
Fognini – Vinolas	11	1558
Murray – Querrey	5	952
Murray – Dimitrov	11	1928
Wawrinka – Paire	10	1584
Raonic – Agut	5	865
Thiem – Goffin	9	1718

*Tabella 4. Dettaglio head-to-head uomini*

# Capitolo 4

## Analisi

Questo capitolo rappresenta il cuore dell'elaborato. Nel primo paragrafo, ad una prima analisi descrittiva dei dati seguirà l'illustrazione dei primi test diagnostici svolti. Nel secondo paragrafo verranno presentate le analisi svolte secondo l'approccio parametrico. Chiuderà il capitolo, la divulgazione delle stime delle probabilità seguita dalle analisi condotte in un contesto non parametrico.

### 4.1. Analisi Preliminari

#### 4.1.1. Ranking

Una volta raccolti e organizzati i dati, il primo passo è stato condurre una serie di analisi a carattere esplorativo al fine di ottenere una visione d'insieme del contesto. In particolar modo, considerata l'eterogeneità dei giocatori inseriti nel database, una prima via per differenziarli tra loro è stata considerare i Ranking. Nello specifico si è deciso di considerare il miglior piazzamento raggiunto nelle classifiche, *Max Ranking*, e il peggiore, *Min Ranking*, in riferimento al periodo 2011-2017. La scelta di limitare il periodo agli ultimi sei anni è dovuta al fatto che i dati raccolti riguardano scontri diretti inclusi in questo lasso di tempo. È bene infine precisare che, a differenza di quanto fatto precedentemente in letteratura, l'analisi dei Ranking svolge in questo elaborato un mero ruolo descrittivo senza aver alcun impatto sulle analisi compiute.



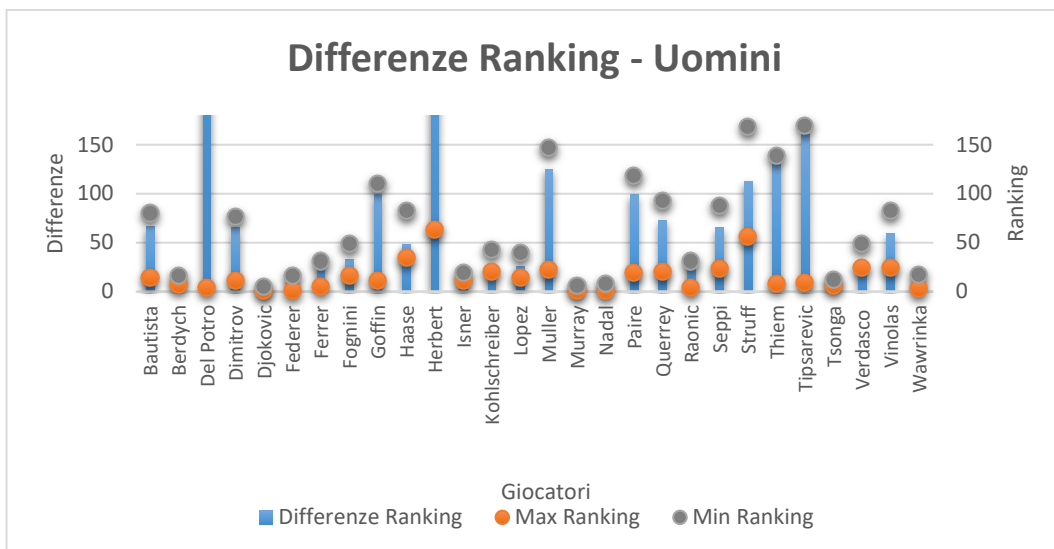


Figura 1. Differenze Ranking – Uomini

Per permettere una migliore interpretazione del grafico sopra riportato, si è deciso di limitare il valore dell'asse verticale ad un valore arbitrario di 180. In questo modo non risultano visibili i valori di *Min Ranking* per Del Potro e Herbert, rispettivamente pari a 1045 e 498.

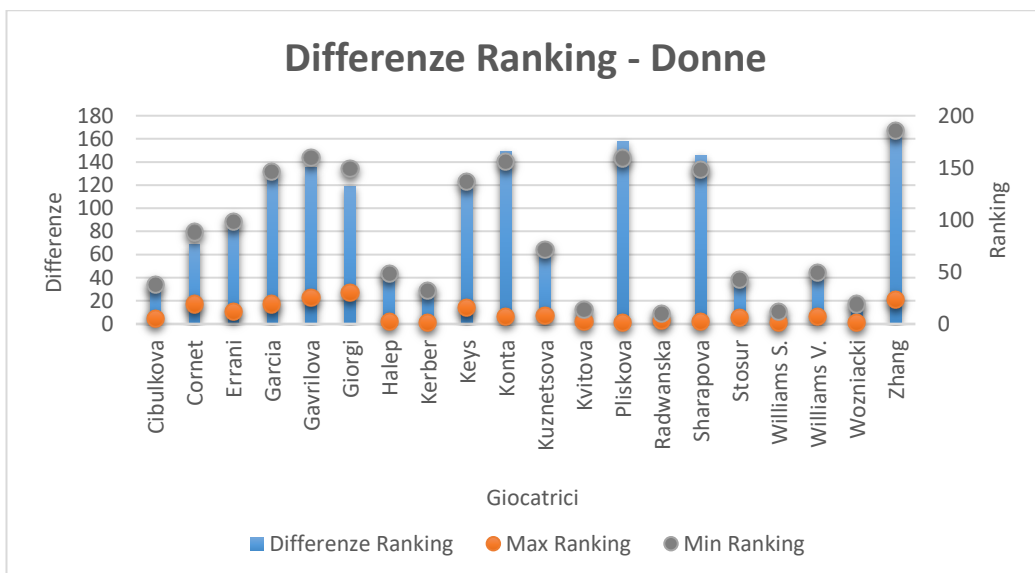


Figura 2. Differenze Ranking – Donne

In entrambi i grafici sopra riportati, l'asse delle ascisse riporta il nome di giocatori e giocatrici considerati. Sulle ordinate sono invece riportati i piazzamenti raggiunti nelle classifiche di Ranking e le differenze tra gli stessi. Risulta subito chiaro come sia possibile distinguere, sia per gli uomini che per le donne, comportamenti differenti tra i giocatori. Osservando, infatti, quei giocatori per cui le differenze tra miglior e peggior piazzamento risultano minori si può notare come questi siano coloro i quali stazionano tra le prime dieci posizioni del ranking. Per chi invece occupa posizioni lontane dal vertice della classifica, soprattutto per piazzamenti superiori al ventesimo, si può notare un maggior range di oscillazione. Per quel che concerne le donne, la situazione risulta ben più diversificata. È indubbio che le migliori giocatrici abbiano un ridotto raggio di oscillazione, si vedano ad esempio Serena Williams e Agnieszka Radwańska, ma è altrettanto vero che le variazioni in termini di Ranking risultano molto più marcate, rispetto agli uomini, anche in riferimento al vertice della classifica. Emblematico è, in questo senso, il caso dell'attuale numero uno al mondo Karolina Pliskova che nel 2011 occupava la posizione 159 della classifica WTA.

Meritano, infine, una nota i casi particolari di Del Potro e Sharapova. Entrambi rappresentano esempi di giocatori vincenti, noti per essere sempre stati ai vertici della disciplina. In questo senso, e sulla base delle precedenti osservazioni, variazioni così ampie come appaiono nei loro casi possono risultare contraddittorie. La spiegazione risiede, nella forzata lontananza dai campi da tennis, cui i due giocatori sono stati costretti. Per quel che concerne il tennista argentino, una serie infinita di infortuni lo hanno obbligato a rinunciare a diversi appuntamenti. Per quel che riguarda, invece, la tennista russa lo stop agonistico è stato dovuto ad una squalifica seguita ad un controllo anti doping cui la Sharapova è risultata essere positiva.

Questa prima analisi, pur nella sua semplicità, permette di avere un'idea più chiara e precisa in riferimento ai dati utilizzati e chiarisce ulteriormente le

motivazioni riguardanti la scelta delle coppie già precedentemente illustrate. Una siffatta analisi non fornisce tuttavia, particolari indicazioni inerenti il problema in esame. A questo scopo rispondono, invece, alcuni test statistici preliminari condotti sulle serie dei dati.

#### 4.1.2. Runs Test

Basandosi sul quesito dal quale questo elaborato prende le mosse, ovvero se i punti siano i.i.d., si è ritenuto opportuno analizzare se le sequenze dei punti vinti, per ognuna delle coppie considerate, fossero casuali o meno. Per far ciò sono stati implementati dei particolari test statistici, noti con il nome di runs test. Il Runs Test, o test delle sequenze, è un test non parametrico atto a verificare che gli elementi generati da un processo di Bernoulli siano tra loro mutualmente indipendenti. Il test si fonda dunque sul concetto di sequenze, runs, definendole come una serie di valori uguali adiacenti, rappresentati tramite notazione +/- o 1/0. Il sistema d'ipotesi cui il testo fa riferimento è:

$H_0$ : la sequenza è casuale;

$H_1$ : la sequenza non è casuale.

La statistica test è, infine, la differenza, standardizzata, tra il numero atteso, sotto  $H_0$ , e il numero osservato di runs. Per campioni numerosi questa statistica si distribuisce come una Normale Standard  $N(0,1)$ .

Il test sopra presentato è facilmente applicabile ai dati in esame; per ogni coppia di giocatori ed in riferimento ad ogni giocatore, è stata definita una variabile dicotomica del tipo:

$$p_{i\_fatto} = \begin{cases} 1 & \text{se l'i-esimo giocatore vince il punto corrente} \\ 0 & \text{altrimenti} \end{cases} \quad \text{per } i = A, B$$

Una volta definita questa variabile, si è deciso di testare l'ipotesi di indipendenza delle sequenze, sia sull'intera sequenza dei punti vinti e persi, sia

sulla sequenza dei soli punti vinti e persi sul proprio servizio per entrambi i giocatori. Le tabelle che seguono riportano, i p-value dei test e il numero di osservazioni e di runs rispetto ai quali i test sono stati condotti.

Coppia	Situazione	pval	runs	n1	n2	n
Kerber	No Serv	0,419	508	525	506	1031
vs	Serv Kerber	0,124	216	284	192	476
Pliskova	Serv Pliskova	0,538	277	314	241	555
Halep	No Serv	0,387	516	527	486	1013
vs	Serv Halep	0,035	10	12	9	21
Kuznetsova	Serv Kuznetsova	0,216	247	279	239	518
Radwanska	No Serv	0,003	32	32	33	66
vs	Serv Radwanska	0,057	366	444	350	794
Kerber	Serv Kerber	0,099	402	447	332	779
S. Williams	No Serv	0,113	699	836	639	1475
vs	Serv S.Williams	0,626	337	472	259	731
Sharapova	Serv Sharapova	0,058	347	380	364	744
Wozniacki	No Serv	0,095	741	757	672	1429
vs	Serv Wozniacki	0,496	341	418	280	698
Cibulkova	Serv Cibulkova	0,477	370	392	339	731
Errani	No Serv	0,089	500	491	461	952
vs	Serv Errani	0,089	500	491	461	952
Cornet	Serv Cornet	0,012	290	260	261	521
Cibulkova	No Serv	0,175	434	393	443	836
vs	Serv Cibulkova	0,312	228	238	203	441
Kvitova	Serv Kvitova	0,657	190	240	155	395
Giorgi	No Serv	0,145	282	282	312	594
vs	Serv Giorgi	0,123	131	164	124	288
Pliskova	Serv Pliskova	0,694	146	188	118	306
V. Williams	No Serv	0,087	555	552	618	1170
vs	Serv V. Williams	0,562	285	360	244	604
S. Williams	Serv S. Williams	0,026	231	374	192	566
V. Williams	No Serv	0,724	407	422	382	804
vs	Serv V.Williams	0,698	185	227	160	387
Stosur	Serv Stosur	0,530	215	222	195	417
Radwanska	No Serv	0,517	252	272	246	518
vs	Serv Radwanska	0,570	114	153	95	248
Keys	Serv Keys	0,528	129	151	119	270
Gavrilova	No Serv	0,005	233	299	236	535
vs	Serv Gavrilova	0,024	82	175	65	240
Garcia	Serv Garcia	0,159	133	171	124	295
Zhang	No Serv	0,515	355	321	374	695
vs	Serv Zhang	0,129	199	186	181	367
Konta	Serv Konta	0,578	155	193	135	328

Tabella 5. Runs Test – Donne

Coppia	Situazione	pval	runs	n1	n2	n
Djokovic vs Federer	No Serv	0,142	3229	3292	3267	6559
	Serv Djokovic	0,144	1591	2145	1221	3366
	Serv Federer	0,567	1477	2046	1147	3193
Federer vs Nadal	No Serv	<b>0,023</b>	1401	1442	1481	2923
	Serv Federer	0,253	687	931	564	1495
	Serv Nadal	0,333	645	917	511	1428
Berdych vs Ferrer	No Serv	0,279	760	778	773	1551
	Serv Berdych	<b>0,010</b>	382	494	269	763
	Serv Ferrer	0,363	356	504	284	788
Del Potro vs Federer	No Serv	<b>0,020</b>	1595	1581	1748	3329
	Serv Del Potro	0,605	786	1100	614	1714
	Serv Federer	0,581	673	1134	481	1615
Federer vs Ferrer	No Serv	0,669	633	685	588	1273
	Serv Federer	0,385	254	408	190	598
	Serv Ferrer	<b>0,044</b>	353	398	277	675
Nadal vs Fognini	No Serv	0,377	1041	1119	996	2115
	Serv Nadal	0,187	449	607	376	983
	Serv Fognini	0,102	586	620	512	1132
Goffin vs Tsonga	No Serv	0,365	490	473	528	1001
	Serv Goffin	0,335	237	317	179	496
	Serv Tsonga	0,503	220	349	156	505
Tipsarevic vs Dimitrov	No Serv	0,317	283	293	289	582
	Serv Tipsarevic	0,379	141	202	113	315
	Serv Dimitrov	0,692	121	176	91	267
Verdasco vs Lopez	No Serv	0,310	321	339	321	660
	Serv Verdasco	0,553	129	201	96	297
	Serv Lopez	0,405	177	225	138	363
Seppi vs Haase	No Serv	0,240	521	521	490	1011
	Serv Seppi	0,416	226	295	175	470
	Serv Haase	<b>0,006</b>	284	315	226	541

Tabella 6. Runs Test - Uomini

Coppia	Situazione	pval	runs	n1	n2	n
Seppi vs Muller	No Serv	0,467	423	440	417	857
	Serv Seppi	0,078	195	296	128	424
	Serv Muller	0,321	200	289	144	433
Struff vs Kohlschreiber	No Serv	0,028	308	322	350	672
	Serv Struff	0,467	150	219	117	336
	Serv Kohlschreiber	0,298	150	233	103	336
Herbert vs Struff	No Serv	0,156	281	279	318	597
	Serv Herbert	0,736	144	175	117	292
	Serv Struff	0,517	133	201	104	305
Isner vs Lopez	No Serv	0,108	281	279	318	597
	Serv Isner	0,511	144	175	117	292
	Serv Lopez	0,359	133	201	104	305
Fognini vs Vinolas	No Serv	0,468	769	825	733	1558
	Serv Fognini	0,649	351	455	283	738
	Serv Vinolas	0,203	422	450	370	820
Murray vs Querrey	No Serv	0,357	460	513	439	952
	Serv Murray	0,596	193	311	134	445
	Serv Querrey	0,456	236	305	202	507
Murray vs Dimitrov	No Serv	0,568	951	1002	926	1928
	Serv Murray	0,643	441	614	335	949
	Serv Dimitrov	0,918	471	591	388	979
Wawrinka vs Paire	No Serv	0,780	786	826	758	1584
	Serv Wawrinka	0,283	336	505	266	771
	Serv Paire	0,226	406	492	321	813
Raonic vs Bautista	No Serv	0,296	417	455	410	865
	Serv Raonic	0,793	189	302	139	441
	Serv Bautista	0,424	189	271	153	424
Thiem vs Goffin	No Serv	0,126	827	825	893	1718
	Serv Thiem	0,572	405	509	346	855
	Serv Goffin	0,629	395	547	316	863

Tabella 7. Runs Test - Uomini

Nelle tabelle sopra riportate, sono contrassegnati in rosso i p-value inferiori al 5%, sintomatici dei casi in cui l'ipotesi nulla del test viene rifiutata. Ciò che si può intuire dai risultati ottenuti è una generale conferma, eccettuato qualche sporadico caso, dell'ipotesi di casualità delle sequenze considerate. I casi in cui il test porta ad assumere non casualità non sembrano, dunque, rappresentare la norma connaturandosi invece come isolate deviazioni da una ben più

marcata tendenza generale. Ad avvalorare, infine, i risultati qui sopra esposti è la stessa letteratura e, nel caso specifico, le analisi condotte da Pollard e Pollard.

## 4.2. L'Approccio Parametrico

Una volta concluse le analisi e i test preliminari, si è proceduto con l'implementare le analisi vere e proprie. Così come precedentemente illustrato, le analisi svolte sono riconducibili a due ben distinti approcci statistici: parametrico e non parametrico. La differenza tra questi due criteri di analisi risiede nella definizione aprioristica, o meno, di ipotesi circa le caratteristiche dei dati in esame. La scelta di stimare un modello logistico, ipotizzando dunque che i dati siano originati da una distribuzione binomiale, caratterizza queste prime analisi come parametriche.

La decisione di procedere con un modello logistico è stata dettata da una serie di considerazioni. In primo luogo la struttura dei dati è del tutto coerente con le ipotesi sottostanti il modello. Ogni stato del match è, infatti, identificato da una variabile dicotomica i cui valori sono 0 o 1 a seconda della situazione di partita in cui ci si trova. Partendo da questa semplice considerazione si è cercato dunque di cogliere se ciascuno stato avesse o meno un qualche effetto statistico sulla sequenza dei punti vinti e persi al servizio. Sulla base di questa intuizione, per ogni sequenza head-to-head, per entrambi i giocatori coinvolti e per tutti gli stati del match considerati è stato stimato un modello logistico  $M_j$  del tipo:

$$\text{logit}(p_{i_{\text{fatto}}}) = \beta_0 + \beta_{1,j} + D_{i,j}$$

dove:

$i$  =  $i$ -esimo punto giocato, con  $i=1, \dots, n$ ;

$j$  =  $j$ -esimo stato, con  $j = 1, \dots, 12$ ;

$$D_{i,j} = \begin{cases} 1 & \text{se l}'i\text{-esimo punto è giocato nel } j\text{-esimo stato} \\ 0 & \text{altrimenti} \end{cases} .$$

Il parametro  $\beta_{1,j}$  descrive invece l'impatto che il j-esimo stato ha sulla trasformata logit di  $p_{i_{\text{fatto}}}$ .

Infine, si indicherà con  $M_0$  il modello stimato sotto l'ipotesi nulla di i.i.d.:

$$\text{logit}(p_{i_{\text{fatto}}}) = \beta_0.$$

Sulla base delle assunzioni costituenti il modello, l'interesse di queste analisi è rivolto verso la stima del parametro  $\beta_{1,j}$ . Rappresentando, infatti, l'impatto dello stato di riferimento è lecito attendersi che deviazioni dall'ipotesi di indipendenza e identica distribuzione corrispondano ad una sua stima significativamente diversa da zero. Al contrario, nel caso di situazioni i.i.d., la stima del parametro  $\beta_{1,j}$  risulterà statisticamente non significativa, ovvero i modelli stimati  $M_0$  e  $M_j$  risulteranno statisticamente equivalenti tra loro secondo il sistema d'ipotesi:

$$H_0: M_0 = M_j;$$

$$H_1: M_0 \neq M_j.$$

Al fine di verificare il sistema d'ipotesi esposto si è reso necessario ricorrere ad un qualche test statistico. Il test designato a questo scopo è il test basato sul log-rapporto di verosimiglianza, definito come:

$$\Lambda(x) = \frac{L(B_0|x)}{L(B_j|x)}.$$

Le tabelle di seguito riportate mostrano i risultati dei test condotti sia sulle coppie maschili, Tabelle 8 e 9, sia sulle coppie di giocatrici, Tabelle 10 e 11. In entrambi i casi, per una migliore interpretazione dei risultati, si è deciso di evidenziare tutti i casi in cui il modello ha portato a rifiutare l'ipotesi nulla. In particolare, in rosso sono indicati i p-value  $< 0.05$ , mentre sono caratterizzati dal colore blu i p-value  $\in (0.05, 0.1)$ .



<b>Giocatori</b>	<b>Mod1</b>	<b>Mod2</b>	<b>Mod3</b>	<b>Mod4</b>	<b>Mod5</b>	<b>Mod6</b>
Djokovic	0,516	0,92	0,879	0,762	0,607	0,285
Federer	0,445	0,757	0,442	0,395	0,393	0,179
Federer	0,482	<b>0,086</b>	0,914	0,571	0,234	0,262
Nadal	0,485	0,808	0,435	0,512	0,454	0,792
Berdych	0,853	0,212	0,367	0,32	0,925	<b>0,039</b>
Ferrer	<b>0,04</b>	<b>0,061</b>	0,374	0,766	0,106	0,931
Del Potro	0,252	<b>0,081</b>	0,141	<b>0,04</b>	<b>0,094</b>	0,263
Federer	0,823	0,515	0,772	0,635	0,705	0,654
Federer	0,357	0,397	<b>0,026</b>	<b>0,02</b>	0,421	0,584
Ferrer	0,484	0,697	<b>0,052</b>	<b>0,069</b>	0,526	0,573
Nadal	0,52	0,288	0,258	0,136	0,753	0,22
Fognini	0,414	0,211	0,575	0,991	0,707	0,411
Goffin	0,59	0,803	0,915	0,986	0,643	0,244
Tsonga	0,802	0,493	0,533	0,693	0,707	0,618
Tipsarevic	0,228	0,96	0,632	0,657	0,242	<b>0,055</b>
Dimitrov	0,51	<b>0,073</b>	<b>0,099</b>	<b>0,007</b>	<b>0,079</b>	0,379
Verdasco	0,451	<b>0,003</b>	0,918	<b>0,026</b>	0,337	0,706
Lopez	0,162	0,74	0,136	0,174	0,167	<b>0,099</b>
Seppi	0,329	0,406	0,142	0,121	0,275	<b>0,018</b>
Haase	<b>0,001</b>	0,141	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	0,616
Seppi	0,113	<b>0,019</b>	0,418	<b>0,053</b>	<b>0,013</b>	0,83
Muller	0,187	0,299	0,737	0,57	0,142	<b>0,01</b>
Struff	0,139	0,586	0,966	0,995	0,141	0,899
Kohlschreiber	0,26	<b>0,095</b>	<b>0,052</b>	0,567	0,744	0,524
Herbert	0,46	0,712	<b>0,055</b>	<b>0,1</b>	0,506	0,381
Struff	0,427	<b>0,089</b>	0,439	<b>0,091</b>	0,127	0,426
Isner	0,526	0,474	0,777	0,897	0,597	0,734
Lopez	0,646	<b>0,02</b>	0,599	<b>0,027</b>	<b>0,063</b>	0,651
Fognini	0,642	0,753	0,414	0,422	0,637	0,938
Vinolas	0,433	0,742	<b>0,065</b>	<b>0,055</b>	0,384	0,677
Murray	0,255	<b>0,006</b>	0,49	0,386	0,991	0,893
Querrey	0,744	0,99	0,79	0,771	0,777	0,968
Wawrinka	0,725	0,56	0,913	0,937	0,824	0,809
Paire	0,526	0,701	0,741	0,94	0,655	0,792
Murray	0,876	0,407	0,276	0,376	0,833	0,234
Dimitrov	0,541	0,327	0,711	0,392	0,317	0,586
Raonic	0,284	0,305	0,185	0,104	0,154	0,879
Bautista	0,782	0,942	0,829	0,987	0,675	0,58
Thiem	0,34	0,267	0,823	0,48	0,604	0,738
Goffin	<b>0,049</b>	<b>0,095</b>	0,125	0,653	0,237	0,495

Tabella 8. P-value regressione logistica, uomini

<b>Giocatori</b>	<b>Mod7</b>	<b>Mod8</b>	<b>Mod9</b>	<b>Mod10</b>	<b>Mod11</b>	<b>Mod12</b>
Djokovic	0,387	<b>0,029</b>	0,424	<b>0,099</b>	0,743	0,738
Federer	0,786	0,763	0,862	0,845	0,711	0,356
Federer	0,467	0,742	0,849	0,496	0,556	0,535
Nadal	0,462	0,166	0,637	0,236	0,763	0,382
Berdych	0,267	0,698	0,44	0,681	0,345	0,474
Ferrer	0,434	0,107	0,159	0,157	<b>0,078</b>	0,546
Del Potro	0,592	0,379	0,499	0,328	0,192	0,337
Federer	0,307	0,789	0,649	0,693	0,29	0,589
Federer	0,887	0,586	0,149	0,677	0,107	0,114
Ferrer	0,311	0,608	0,189	0,61	0,273	0,872
Nadal	0,18	0,299	0,137	0,442	0,161	0,293
Fognini	0,313	0,423	0,241	0,292	<b>0,071</b>	0,285
Goffin	0,248	0,461	0,639	0,619	0,393	0,227
Tsonga	0,573	0,933	0,649	0,936	0,829	0,957
Tipsarevic	0,936	0,688	0,944	0,712	0,971	0,689
Dimitrov	0,913	0,585	<b>0,099</b>	0,316	<b>0,058</b>	<b>0,006</b>
Verdasco	0,511	0,341	0,924	0,204	0,15	<b>0,066</b>
Lopez	0,569	0,468	0,503	0,447	0,497	0,284
Seppi	0,463	0,693	0,387	0,654	0,356	0,802
Haase	0,152	<b>0,007</b>	<b>0,001</b>	<b>0,006</b>	<b>0,002</b>	0,819
Seppi	0,29	<b>0,07</b>	0,386	<b>0,058</b>	0,162	0,141
Muller	0,756	0,517	0,851	0,389	0,721	0,179
Struff	0,288	0,364	0,56	0,292	0,373	0,457
Kohlschreiber	0,312	0,453	0,121	0,747	0,628	0,929
Herbert	0,72	0,256	<b>0,074</b>	0,212	<b>0,071</b>	0,465
Struff	0,295	0,489	0,633	0,319	0,324	0,132
Isner	0,778	0,866	0,835	0,778	0,986	0,392
Lopez	0,362	0,704	0,761	0,141	0,15	0,577
Fognini	0,691	0,663	0,503	0,517	0,373	0,386
Vinolas	0,8	0,861	<b>0,084</b>	0,777	<b>0,045</b>	0,846
Murray	0,816	0,663	0,733	0,78	0,595	<b>0,023</b>
Querrey	0,857	0,983	0,852	0,907	0,983	0,456
Wawrinka	0,911	0,629	0,681	0,758	0,889	0,226
Paire	0,359	0,602	0,445	0,661	0,609	0,207
Murray	0,715	<b>0,061</b>	0,28	0,115	0,464	0,159
Dimitrov	0,963	0,836	0,981	0,754	0,859	0,113
Raonic	0,654	0,609	0,526	0,537	0,314	0,145
Bautista	0,946	0,991	0,836	0,877	0,98	0,484
Thiem	0,572	0,156	0,624	<b>0,033</b>	0,173	0,562
Goffin	0,475	0,18	0,14	0,534	0,271	0,25

Tabella 9. P-value regressione logistica, uomini

Giocatrici	Mod1	Mod2	Mod3	Mod4	Mod5	Mod6
Kerber	0,667	0,313	0,271	0,353	0,641	0,296
Pliskova	0,528	0,614	0,223	0,191	0,446	0,623
Halep	0,434	<b>0,006</b>	0,595	0,136	0,994	<b>0,092</b>
Kuznetsvoa	0,408	0,392	0,281	0,218	0,329	0,408
Radwanska	0,423	0,88	0,11	<b>0,062</b>	0,298	0,715
Kerber	0,651	0,759	0,42	0,685	0,877	0,683
S. Williams	0,491	0,826	0,857	0,758	0,438	0,187
Sharapova	0,711	0,758	0,58	0,577	0,693	0,419
Wozniacki	<b>0,099</b>	0,584	0,981	0,98	<b>0,099</b>	0,765
Cibulkova	0,606	<b>0,05</b>	0,355	0,707	0,925	0,871
Errani	0,302	0,446	0,981	0,591	0,58	0,478
Cornet	0,492	0,498	<b>0,033</b>	<b>0,034</b>	0,444	0,552
Cibulkova	0,5	NA	0,863	0,863	0,5	0,481
Kvitova	<b>0,1</b>	NA	0,225	0,225	<b>0,1</b>	0,356
Giorgi	0,194	0,233	0,225	0,915	0,619	0,215
Pliskova	0,389	0,873	0,432	0,41	0,354	0,519
V. Williams	0,86	0,452	0,735	0,566	0,711	0,702
Stosur	0,864	0,123	0,852	0,516	0,588	0,223
Radwanska	0,659	0,373	0,221	0,601	0,989	0,711
Keys	0,321	0,471	0,108	0,14	0,358	0,571
V. Williams	0,647	0,275	0,539	0,239	0,962	0,319
S. Williams	0,809	<b>0,094</b>	0,253	0,752	0,436	0,486
Gavrilova	0,431	NA	<b>0,075</b>	<b>0,075</b>	0,431	0,547
Garcia	0,439	NA	0,674	0,674	0,439	0,796
S.Zhang	0,946	NA	0,974	0,974	0,946	0,512
Konta	0,214	NA	<b>0,054</b>	<b>0,054</b>	0,214	0,904

Tabella 10. P-value regressione logistica, donne

Giocatrici	Mod7	Mod8	Mod9	Mod10	Mod11	Mod12
Kerber	0,541	<b>0,088</b>	0,599	0,337	0,782	<b>0,066</b>
Pliskova	0,757	0,378	0,134	0,353	0,159	0,244
Halep	0,511	0,604	0,816	0,553	0,37	0,569
Kuznetsvoa	0,75	0,81	0,581	0,787	0,591	0,777
Radwanska	<b>0,048</b>	<b>0,046</b>	<b>0,049</b>	<b>0,031</b>	<b>0,035</b>	0,502
Kerber	0,179	0,441	0,185	0,391	0,12	0,515
S. Williams	0,883	0,751	0,301	0,748	0,403	0,247
Sharapova	0,144	<b>0,097</b>	0,135	0,121	0,23	<b>0,011</b>
Wozniacki	0,253	0,11	0,449	0,111	0,467	0,532
Cibulkova	0,183	0,359	0,351	0,436	0,508	0,175
Errani	0,747	0,559	0,712	0,933	0,914	0,29
Cornet	<b>0,059</b>	<b>0,001</b>	<b>0,007</b>	<b>0,001</b>	<b>0,019</b>	0,394
Cibulkova	0,128	0,411	0,219	0,411	0,219	0,829
Kvitova	0,832	0,403	0,208	0,403	0,208	0,245
Giorgi	0,262	<b>0,092</b>	0,268	0,438	0,604	0,602
Pliskova	0,598	0,4	0,614	0,611	0,778	0,667
V. Williams	0,902	0,501	<b>0,049</b>	0,683	0,187	0,632
Stosur	0,538	0,937	0,794	0,82	0,249	0,74
Radwanska	0,736	0,926	0,612	0,974	0,932	0,345
Keys	0,319	0,33	0,214	0,266	0,201	0,235
V. Williams	0,941	0,422	0,804	0,315	0,53	0,251
S. Williams	0,214	0,616	0,138	0,571	0,354	0,962
Gavrilova	<b>0,036</b>	0,117	<b>0,058</b>	0,117	<b>0,058</b>	0,346
Garcia	0,285	0,57	0,542	0,57	0,542	0,305
S.Zhang	0,325	0,54	0,286	0,54	0,286	0,789
Konta	0,878	0,319	0,261	0,319	0,261	0,961

Tabella 11. P-value regressione logistica, donne

Osservando i risultati riportati, ciò che si può in primo luogo notare è come, in diverse occasioni, i modelli suggeriscano deviazioni dal caso di indipendenza e identica distribuzione. Non sono, infatti, così isolati p-value inferiori alla soglia del 5% ed in particolare si distinguono due casi. Per quel che concerne il gruppo di giocatori maschili, è da segnalare il comportamento registrato per la coppia Seppi-Haase. In particolare, in relazione ad Haase, è possibile notare come solamente in riferimento agli stati del match Tiebreak, Set, Punto Precedente e Importance, rispettivamente Mod2, Mod6, Mod7 e Mod12, non si registrino deviazioni rispetto l'assunzione di i.i.d.. Al contrario, in tutti gli altri stati del match vengono segnalate deviazioni rispetto l'ipotesi nulla. Anche se meno accentuato, un altro caso analogo riguarda la coppia Errani-Cornet. Ancora una volta, in riferimento ai modelli stimati per Cornet, le deviazioni dall'ipotesi nulla riguardano più della metà degli stati considerati con le eccezioni di Game Point, Tiebreak, Break o Set Point e Tiebreak, Importance, rispettivamente indicati con Mod1, Mod2, Mod5, Mod6 e Mod12. Tuttavia, nonostante i due casi segnalati e i diversi p-value evidenziati, ciò che emerge in generale dalle analisi fatte è l'assenza di una precisa e sintomatica differenza rispetto al caso di indipendenza preso in esame. Le deviazioni evidenziate sembrano rispondere a casi isolati e sporadici piuttosto che sistematici. Non è infatti identificabile uno stato del match in particolare in cui sia possibile assistere ad una netta deviazione dall'ipotesi espressa sotto  $H_0$ .

In sintesi, il primo approccio implementato al fine di indagare la natura statistica dei punti giocati in una partita di tennis è stato un approccio parametrico, nello specifico la stima di una serie di modelli di regressione logistica. Una volta stimati i modelli, per ogni giocatore e per ogni stato del match, si è verificato se questi risultassero statisticamente equivalenti al modello riferito ai soli punti giocati al servizio. I p-value dei test, basati sul log rapporto di verosimiglianza, condotti hanno portato a concludere che non sono

identificabili particolari situazioni tali da suggerire deviazioni rispetto l'ipotesi nulla di indipendenza.

### **4.3. L'Approccio non Parametrico**

Le analisi parametriche rappresentano il primo dei due approcci adottati per analizzare il problema in esame. L'idea di sfruttare un approccio non parametrico nasce da una serie di considerazioni teoriche e pratiche. In prima battuta, muoversi in un contesto non parametrico permette di svincolarsi dalla necessità di definire delle assunzioni costituenti il modello, al contrario di quanto fatto precedentemente con il modello di regressione logistica. In secondo luogo, si è deciso di percorrere questa strada per tentare di differenziarsi da quanto fatto precedentemente in letteratura, proponendo un nuovo metodo per studiare il fenomeno dell'indipendenza dei punti. Sulla base di queste motivazioni si è deciso di ricorrere ad uno degli approcci non parametrici più diffusi, ovvero la simulazione Monte Carlo. L'adozione di questo approccio ha permesso, inoltre, di superare un altro problema legato ad alcune delle coppie inserite nel database, ovvero l'esigua numerosità campionaria dovuta ad un ridotto numero di incontri giocati. Da un punto di vista operativo, il primo passo è stato elaborare un algoritmo in grado di simulare un incontro di tennis partendo da una serie di input. Il primo e fondamentale elemento necessario per il funzionamento dell'algoritmo è rappresentato dalla probabilità di vincere il punto. Di conseguenza, al fine di implementare le simulazioni basate sul Metodo Monte Carlo, si è reso necessario un passaggio preliminare: stimare le probabilità di vincere un punto al servizio nei vari stati del match.

Nel seguito di questo paragrafo, dunque, verranno presentate dapprima le stime delle probabilità e, a seguire, l'illustrazione dell'algoritmo utilizzato e

l'analisi dei risultati ottenuti implementando il metodo di simulazione Monte Carlo.

#### **4.3.1. Le probabilità stimate**

Primo e fondamentale passaggio per procedere con le simulazioni è stato stimare le probabilità di vincere un punto al servizio nei vari stati del match. Nello specifico, le probabilità sono state stimate per ogni sequenza head-to-head, per ognuno dei due giocatori coinvolti nella stessa e per ognuno dei dodici stati considerati. Nel seguito si indicheranno con  $p_{A,0}$  e  $p_{B,0}$  le probabilità di vincere un punto al servizio nel caso di indipendenza, rispettivamente per il giocatore A e il giocatore B. Le diciture  $p_{A,00}$  e  $p_{B,00}$  verranno utilizzate per indicare le probabilità di vincere un punto indipendente dal servizio, rispettivamente per il giocatore A e per il giocatore B. Infine,  $p_{A,j}$  e  $p_{B,j}$  indicheranno le probabilità di vincere un punto al servizio nel j-esimo stato, per  $j=1,\dots,12$ , rispettivamente per il giocatore A e per il giocatore B.

Nelle tabelle riportate di seguito sono indicate le probabilità stimate nei vari stati del match. L'assenza di alcune probabilità è legata all'assenza di punti giocati in quello specifico stato del match. Questi casi sono relativi, per gli uomini, all'assenza di match al meglio dei cinque set o all'assenza di quinti set giocati. Per quanto concerne invece le coppie femminile, in tre casi, Cibulkova-Kvitova, Gavrilova-Garcia e Zhang-Konta, non si registrano punti giocati al tiebreak.

IP (vincere punto   servizio &   -esimo stato del match)																			
Giocatori	P <sub>00</sub>	P <sub>0</sub>	GamePoint		Tiebreak		Break- or Set-Points		Tie-break o Break- o Set-Points		Tie-break o Game-Points		Set					Punto Precedente	
			1	0	1	0	1	0	1	0	1	0	1	2	3	4	5	Vinto	Perso
Djokovic	0,50	0,64	0,63	0,641	0,64	0,64	0,641	0,637	0,643	0,636	0,63	0,64	0,631	0,63	0,65	0,65	0,67	0,63	0,65
Federer	0,50	0,64	0,63	0,645	0,66	0,64	0,622	0,643	0,622	0,644	0,63	0,65	0,66	0,63	0,63	0,62	0,65	0,64	0,64
Federer	0,49	0,62	0,61	0,628	0,46	0,63	0,626	0,622	0,605	0,626	0,6	0,63	0,651	0,59	0,64	0,53	0,69	0,63	0,61
Nadal	0,51	0,64	0,66	0,637	0,62	0,64	0,614	0,646	0,62	0,645	0,66	0,64	0,641	0,65	0,63	0,73	0,5	0,65	0,63
Berdych	0,50	0,65	0,65	0,645	0,44	0,65	0,602	0,653	0,6	0,654	0,65	0,65	0,7	0,63	0,61	0,6	-	0,63	0,67
Ferrer	0,50	0,64	0,7	0,618	0,33	0,64	0,681	0,634	0,653	0,638	0,68	0,62	0,641	0,64	0,64	0,67	-	0,65	0,63
Del Potro	0,48	0,64	0,66	0,634	0,74	0,64	0,689	0,636	0,698	0,63	0,67	0,63	0,637	0,66	0,66	0,54	0,6	0,64	0,65
Federer	0,53	0,7	0,71	0,701	0,74	0,7	0,712	0,701	0,716	0,7	0,71	0,7	0,687	0,73	0,68	0,74	0,64	0,71	0,69
Federer	0,54	0,68	0,66	0,694	0,83	0,68	0,803	0,669	0,803	0,667	0,66	0,69	0,681	0,67	0,72	-	-	0,68	0,69
Ferrer	0,56	0,59	0,61	0,581	0,67	0,59	0,687	0,576	0,678	0,577	0,61	0,58	0,554	0,64	0,55	-	-	0,57	0,61
Nadal	0,53	0,62	0,61	0,622	0,73	0,62	0,619	0,617	0,629	0,615	0,61	0,62	0,614	0,63	0,65	0,52	0,39	0,64	0,59
Fognini	0,47	0,55	0,54	0,551	0,5	0,55	0,574	0,543	0,569	0,543	0,54	0,55	0,518	0,56	0,6	0,57	0,47	0,53	0,56
Goffin	0,47	0,64	0,66	0,632	0,67	0,64	0,645	0,638	0,64	0,639	0,65	0,63	0,667	0,64	0,6	-	-	0,62	0,67
Tsonga	0,53	0,69	0,7	0,688	0,77	0,69	0,655	0,696	0,671	0,694	0,7	0,69	0,675	0,75	0,64	-	-	0,7	0,68
Tipsarevic	0,50	0,64	0,69	0,621	0,64	0,64	0,675	0,636	0,667	0,636	0,69	0,62	0,72	0,56	0,61	-	-	0,64	0,64
Dimitrov	0,50	0,66	0,63	0,67	0,67	0,68	0,471	0,672	0,459	0,691	0,58	0,69	0,69	0,62	0,65	-	-	0,66	0,66
Verdasco	0,51	0,68	0,71	0,665	0,39	0,7	0,667	0,678	0,524	0,702	0,64	0,7	0,683	0,68	0,65	-	-	0,69	0,66
Lopez	0,49	0,62	0,68	0,598	0,65	0,714	0,605	0,605	0,691	0,603	0,67	0,6	0,641	0,65	0,48	-	-	0,61	0,64
Seppi	0,52	0,63	0,59	0,641	0,5	0,63	0,531	0,639	0,534	0,641	0,59	0,64	0,579	0,64	0,72	-	-	0,61	0,65
Haase	0,49	0,58	0,48	0,628	0,36	0,59	0,394	0,608	0,405	0,61	0,48	0,63	0,573	0,58	0,6	-	-	0,55	0,61
Seppi	0,51	0,7	0,64	0,721	0,48	0,71	0,643	0,704	0,594	0,717	0,62	0,74	0,683	0,72	0,68	-	-	0,68	0,73
Muller	0,49	0,67	0,72	0,65	0,76	0,66	0,69	0,665	0,698	0,662	0,72	0,65	0,732	0,66	0,58	-	-	0,66	0,68
Struff	0,48	0,65	0,71	0,629	0,73	0,65	0,649	0,652	0,652	0,652	0,71	0,63	0,659	0,64	0,65	-	-	0,68	0,62
Kohlischreiber	0,52	0,69	0,74	0,676	0,46	0,7	0,839	0,679	0,732	0,688	0,71	0,69	0,667	0,71	0,7	-	-	0,72	0,66
Herbert	0,47	0,6	0,56	0,612	0,55	0,6	0,429	0,617	0,472	0,617	0,57	0,61	0,621	0,6	0,54	-	-	0,61	0,59
Struff	0,53	0,66	0,63	0,673	0,4	0,67	0,6	0,667	0,545	0,678	0,6	0,69	0,652	0,72	0,55	-	-	0,68	0,63
Isner	0,50	0,72	0,7	0,724	0,75	0,71	0,702	0,719	0,713	0,719	0,71	0,72	0,701	0,72	0,74	-	-	0,71	0,72
Lopez	0,50	0,72	0,71	0,727	0,61	0,74	0,694	0,725	0,64	0,739	0,68	0,75	0,725	0,75	0,67	-	-	0,74	0,71
Fognini	0,53	0,62	0,63	0,601	0,67	0,62	0,656	0,611	0,653	0,624	0,63	0,57	0,633	0,59	0,66	-	-	0,62	0,54
Vinolas	0,47	0,55	0,57	0,54	0,6	0,55	0,624	0,535	0,611	0,534	0,61	0,54	0,54	0,56	0,56	-	-	0,61	0,55
Murray	0,54	0,7	0,74	0,684	0,38	0,71	0,744	0,694	0,649	0,706	0,7	0,7	0,686	0,72	0,68	0,71	-	0,7	0,71
Querrey	0,46	0,6	0,59	0,606	0,6	0,6	0,618	0,6	0,618	0,599	0,59	0,61	0,586	0,64	0,56	0,64	-	0,61	0,6
Wawrinka	0,52	0,66	0,65	0,659	0,75	0,65	0,649	0,656	0,659	0,655	0,65	0,66	0,658	0,66	0,65	-	-	0,66	0,65
Paire	0,48	0,61	0,59	0,612	0,67	0,6	0,59	0,607	0,602	0,606	0,59	0,61	0,613	0,6	0,6	-	-	0,59	0,62
Murray	0,52	0,65	0,65	0,646	0,71	0,64	0,6	0,653	0,615	0,653	0,65	0,65	0,659	0,66	0,6	0,63	-	0,64	0,65
Dimitrov	0,48	0,6	0,59	0,61	0,53	0,61	0,588	0,606	0,572	0,609	0,58	0,61	0,611	0,6	0,61	0,56	-	0,6	0,6
Raonic	0,53	0,69	0,72	0,671	0,77	0,68	0,775	0,676	0,769	0,67	0,73	0,66	0,709	0,67	0,64	0,85	-	0,69	0,67
Bautista	0,47	0,64	0,65	0,635	645	0,64	0,625	0,641	0,64	0,639	0,65	0,63	0,593	0,68	0,7	0,4	-	0,64	0,64
Thiem	0,48	0,6	0,62	0,585	0,5	0,6	0,586	0,597	0,568	0,601	0,61	0,61	0,611	0,58	0,59	0,61	-	0,61	0,59
Goffin	0,52	0,63	0,58	0,655	0,77	0,63	0,566	0,643	0,617	0,637	0,59	0,65	0,626	0,63	0,65	0,65	-	0,64	0,62

Tabella 12. Probabilità Stimate, Uomini



IP (vincere punto   servizio & j-esimo stato del match)																				
PLAYERS	P <sub>00</sub>	P <sub>0</sub>	Punto Precedente & Game Point				Punto Precedente & BreakPoint & SetPoint				Punto Precedente & Game Point o Tiebreak				Punto Precedente & BreakPoint & SetPoint o Tiebreak				Importance	
			Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	1	0
Djokovic	0,50	0,64	0,60	0,65	0,69	0,63	0,59	0,63	0,67	0,64	0,60	0,65	0,68	0,63	0,62	0,63	0,66	0,64	0,63	0,64
Federer	0,50	0,64	0,64	0,64	0,62	0,65	0,61	0,64	0,63	0,65	0,63	0,64	0,63	0,65	0,60	0,64	0,64	0,65	0,63	0,65
Federer	0,49	0,62	0,62	0,64	0,60	0,61	0,65	0,63	0,61	0,61	0,62	0,64	0,57	0,62	0,67	0,63	0,57	0,62	0,61	0,63
Nadal	0,51	0,64	0,70	0,63	0,59	0,64	0,67	0,65	0,59	0,64	0,69	0,63	0,60	0,64	0,66	0,65	0,60	0,64	0,62	0,65
Berdych	0,50	0,65	0,63	0,63	0,69	0,66	0,53	0,64	0,64	0,67	0,63	0,63	0,69	0,66	0,52	0,64	0,65	0,67	0,63	0,66
Ferrer	0,50	0,64	0,68	0,64	0,73	0,60	0,57	0,66	0,73	0,60	0,66	0,65	0,73	0,60	0,52	0,66	0,72	0,60	0,62	0,65
Del Potro	0,48	0,64	0,67	0,62	0,65	0,65	0,70	0,63	0,69	0,64	0,67	0,62	0,68	0,64	0,67	0,63	0,71	0,63	0,66	0,64
Federer	0,53	0,70	0,71	0,71	0,69	0,69	0,75	0,71	0,68	0,69	0,72	0,71	0,68	0,69	0,77	0,70	0,66	0,69	0,69	0,71
Federer	0,54	0,68	0,64	0,71	0,70	0,68	0,84	0,67	0,78	0,67	0,64	0,70	0,70	0,68	0,85	0,67	0,77	0,67	0,75	0,67
Ferrer	0,56	0,59	0,58	0,56	0,65	0,60	0,75	0,56	0,67	0,59	0,58	0,57	0,65	0,60	0,71	0,56	0,67	0,59	0,59	0,59
Nadal	0,53	0,62	0,56	0,61	0,56	0,54	0,59	0,60	0,65	0,53	0,58	0,60	0,55	0,55	0,63	0,59	0,63	0,53	0,61	0,57
Fognini	0,47	0,55	0,48	0,49	0,55	0,49	0,42	0,49	0,57	0,49	0,46	0,50	0,55	0,49	0,38	0,50	0,57	0,49	0,47	0,51
Goffin	0,47	0,64	0,66	0,59	0,65	0,67	0,56	0,62	0,68	0,66	0,64	0,60	0,67	0,67	0,50	0,63	0,70	0,66	0,60	0,66
Tsonga	0,53	0,69	0,70	0,70	0,70	0,67	0,61	0,71	0,70	0,67	0,71	0,70	0,69	0,67	0,65	0,71	0,69	0,67	0,69	0,69
Tipsarevic	0,50	0,64	0,70	0,62	0,69	0,63	0,71	0,64	0,65	0,64	0,69	0,62	0,68	0,62	0,68	0,64	0,66	0,63	0,66	0,63
Dimitrov	0,50	0,66	0,67	0,66	0,52	0,68	0,80	0,66	0,33	0,69	0,61	0,69	0,54	0,70	0,46	0,68	0,46	0,71	0,48	0,70
Verdasco	0,51	0,68	0,76	0,66	0,56	0,67	0,67	0,69	0,67	0,66	0,70	0,69	0,50	0,70	0,52	0,72	0,52	0,68	0,57	0,70
Lopez	0,49	0,62	0,66	0,58	0,70	0,62	0,71	0,60	0,72	0,62	0,65	0,58	0,71	0,61	0,64	0,60	0,73	0,61	0,58	0,64
Seppi	0,52	0,63	0,59	0,62	0,60	0,66	0,50	0,62	0,55	0,66	0,58	0,63	0,60	0,66	0,50	0,62	0,56	0,66	0,64	0,63
Haase	0,49	0,58	0,44	0,61	0,52	0,64	0,25	0,58	0,46	0,65	0,45	0,61	0,52	0,65	0,32	0,58	0,45	0,65	0,57	0,59
Seppi	0,51	0,70	0,59	0,73	0,79	0,71	0,55	0,69	0,75	0,72	0,59	0,74	0,69	0,74	0,56	0,70	0,63	0,75	0,63	0,71
Muller	0,49	0,67	0,72	0,63	0,71	0,67	0,75	0,66	0,65	0,68	0,73	0,62	0,70	0,67	0,75	0,65	0,66	0,68	0,62	0,69
Struff	0,48	0,65	0,71	0,66	0,72	0,60	0,58	0,69	0,68	0,61	0,70	0,67	0,74	0,59	0,56	0,69	0,70	0,61	0,62	0,66
Kornischnre	0,52	0,69	0,73	0,71	0,76	0,64	0,92	0,70	0,79	0,65	0,71	0,72	0,69	0,66	0,80	0,71	0,69	0,66	0,69	0,70
Herbert	0,47	0,60	0,64	0,59	0,43	0,63	0,67	0,61	0,32	0,63	0,65	0,59	0,44	0,63	0,69	0,60	0,35	0,63	0,64	0,59
Struff	0,53	0,66	0,62	0,71	0,63	0,63	0,58	0,69	0,61	0,63	0,63	0,71	0,56	0,66	0,60	0,69	0,52	0,66	0,59	0,68
Isner	0,50	0,72	0,69	0,73	0,73	0,72	0,74	0,71	0,65	0,73	0,69	0,73	0,74	0,72	0,72	0,71	0,71	0,73	0,74	0,71
Lopez	0,50	0,72	0,73	0,74	0,67	0,71	0,69	0,74	0,70	0,71	0,71	0,75	0,63	0,74	0,66	0,75	0,63	0,73	0,71	0,73
Fognini	0,53	0,62	0,66	0,61	0,58	0,62	0,75	0,61	0,61	0,61	0,67	0,60	0,58	0,62	0,76	0,61	0,60	0,61	0,64	0,61
Vinolas	0,47	0,55	0,57	0,53	0,57	0,55	0,76	0,53	0,58	0,54	0,58	0,53	0,56	0,55	0,77	0,52	0,58	0,55	0,56	0,55
Murray	0,54	0,70	0,74	0,67	0,73	0,70	0,80	0,69	0,70	0,71	0,72	0,68	0,66	0,72	0,71	0,69	0,61	0,73	0,58	0,72
Querrey	0,46	0,60	0,59	0,61	0,59	0,60	0,73	0,60	0,59	0,60	0,58	0,62	0,61	0,59	0,61	0,61	0,62	0,59	0,58	0,61
Wawrinka	0,52	0,66	0,62	0,67	0,69	0,64	0,56	0,66	0,69	0,65	0,63	0,67	0,68	0,64	0,61	0,66	0,68	0,65	0,61	0,67
Paire	0,48	0,61	0,56	0,61	0,63	0,62	0,46	0,60	0,63	0,62	0,56	0,60	0,64	0,62	0,50	0,60	0,63	0,62	0,57	0,62
Murray	0,52	0,65	0,70	0,62	0,56	0,68	0,67	0,64	0,55	0,67	0,69	0,62	0,59	0,68	0,66	0,64	0,58	0,67	0,60	0,66
Dimitrov	0,48	0,60	0,58	0,62	0,61	0,60	0,57	0,61	0,59	0,61	0,57	0,62	0,59	0,61	0,56	0,61	0,58	0,61	0,56	0,62
Raonic	0,53	0,69	0,71	0,69	0,76	0,65	0,77	0,69	0,78	0,66	0,73	0,67	0,73	0,65	0,83	0,68	0,74	0,66	0,74	0,67
Bautista	0,47	0,64	0,64	0,64	0,66	0,64	0,53	0,65	0,67	0,64	0,67	0,62	0,63	0,65	0,67	0,63	0,63	0,65	0,61	0,65
Thiem	0,48	0,60	0,67	0,57	0,55	0,60	0,68	0,60	0,55	0,60	0,68	0,57	0,52	0,61	0,70	0,59	0,51	0,61	0,58	0,60
Goffin	0,52	0,63	0,61	0,66	0,55	0,65	0,69	0,64	0,51	0,65	0,63	0,65	0,58	0,64	0,72	0,64	0,55	0,64	0,60	0,64

Tabella 13. Probabilità Stimate, Uomini



		IP (vincere punto   servizio & j-esimo stato del match)															
		GamePoint		Tiebreak		Break- or Set- Points		Tie-break or Game- or Break- Points		Tie-break or Game-Points		Set			Punto Precedente		
PLAYERS	$p_{00}$	$p_0$	1	0	1	0	1	0	1	0	1	0	1	2	3	Vinto	Perso
Kerber	0,51	0,60	0,61	0,59	0,70	0,59	0,53	0,61	0,55	0,61	0,61	0,59	0,64	0,54	0,61	0,61	0,58
Pliskova	0,49	0,57	0,59	0,56	0,62	0,56	0,63	0,56	0,62	0,55	0,59	0,56	0,58	0,56	0,55	0,56	0,57
Halep	0,52	0,58	0,61	0,57	0,18	0,59	0,55	0,59	0,50	0,60	0,58	0,58	0,55	0,59	0,66	0,57	0,60
Kuznetsova	0,48	0,54	0,51	0,55	0,42	0,54	0,48	0,55	0,47	0,55	0,51	0,55	0,56	0,52	0,52	0,53	0,55
Radwanska	0,49	0,56	0,54	0,57	0,54	0,56	0,49	0,57	0,48	0,57	0,53	0,57	0,56	0,57	0,53	0,59	0,52
Kerber	0,50	0,57	0,56	0,58	0,62	0,57	0,54	0,58	0,56	0,58	0,57	0,58	0,57	0,58	0,58	0,55	0,60
Williams	0,57	0,65	0,63	0,65	0,67	0,65	0,64	0,65	0,63	0,65	0,63	0,66	0,62	0,67	0,71	0,65	0,64
Sharapova	0,43	0,51	0,50	0,52	0,54	0,51	0,49	0,52	0,49	0,52	0,50	0,52	0,52	0,51	0,29	0,54	0,49
Wozniacki	0,53	0,60	0,65	0,58	0,67	0,60	0,60	0,60	0,60	0,60	0,65	0,58	0,60	0,60	0,58	0,58	0,62
Cibulkova	0,47	0,54	0,52	0,54	0,79	0,53	0,50	0,54	0,52	0,54	0,53	0,54	0,51	0,57	0,47	0,51	0,56
Errani	0,52	0,53	0,58	0,52	0,46	0,54	0,53	0,53	0,51	0,54	0,55	0,53	0,52	0,54	0,58	0,53	0,54
Cornet	0,48	0,50	0,52	0,49	0,57	0,50	0,59	0,48	0,59	0,47	0,52	0,49	0,49	0,49	0,57	0,46	0,54
Cibulkova	0,47	0,54	0,52	0,55	-	0,54	0,55	0,54	0,55	0,54	0,52	0,55	0,56	0,53	0,51	0,50	0,58
Kvitova	0,53	0,61	0,54	0,63	-	0,61	0,52	0,62	0,52	0,62	0,52	0,63	0,63	0,59	0,57	0,60	0,61
Giorgi	0,48	0,57	0,51	0,59	0,71	0,56	0,47	0,58	0,56	0,57	0,55	0,58	0,61	0,51	0,55	0,60	0,54
Pliskova	0,53	0,61	0,65	0,60	0,63	0,61	0,67	0,61	0,66	0,60	0,65	0,60	0,63	0,62	0,57	0,63	0,60
V. Williams	0,53	0,59	0,59	0,58	0,70	0,58	0,61	0,58	0,62	0,58	0,60	0,58	0,59	0,60	0,55	0,58	0,59
Stosur	0,48	0,53	0,54	0,53	0,78	0,53	0,54	0,53	0,57	0,53			0,51	0,52	0,62	0,52	0,55
Radwanska	0,53	0,62	0,59	0,63	0,80	0,61	0,50	0,63	0,57	0,62	0,62	0,62	0,65	0,56	0,67	0,61	0,28
Keys	0,48	0,56	0,60	0,54	0,40	0,56	0,67	0,54	0,65	0,54	0,60	0,54	0,56	0,53	0,70	0,59	0,53
V. Williams	0,47	0,60	0,61	0,59	0,47	0,60	0,56	0,60	0,54	0,61	0,60	0,60	0,56	0,65	0,59	0,60	0,60
S. Williams	0,53	0,66	0,67	0,66	0,82	0,65	0,59	0,67	0,65	0,66	0,69	0,65	0,67	0,66	0,63	0,68	0,32
Gavrilova	0,56	0,73	0,69	0,74	-	0,73	0,59	0,75	0,59	0,75	0,69	0,74	0,71	0,76	0,74	0,78	0,65
Garcia	0,44	0,58	0,55	0,59	-	0,58	0,55	0,58	0,55	0,58	0,55	0,59	0,60	0,55	0,65	0,61	0,55
S.Zhang	0,46	0,51	0,51	0,51	-	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,53	0,49	0,49	0,48	0,53
Konta	0,54	0,59	0,53	0,61	-	0,59	0,46	0,61	0,46	0,61	0,53	0,61	0,59	0,58	0,59	0,59	0,59

Tabella 14. Probabilità Stimate, Uomini

IP (vincere punto   servizio & j-esimo stato del match)																				
PLAYERS	p <sub>00</sub>	p <sub>0</sub>	Punto Precedente & Game Point				Punto Precedente & BreakPoint & SetPoint				Punto Precedente & Game Point o Tiebreak				Punto Precedente & BreakPoint & SetPoint o Tiebreak				Importance	
			Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	Vinto & 1	Vinto & 0	Perso & 1	Perso & 0	1	0
Kerber	0,51	0,60	0,70	0,57	0,49	0,61	0,61	0,61	0,50	0,60	0,68	0,58	0,53	0,60	0,56	0,62	0,55	0,59	0,52	0,62
Pliskova	0,49	0,57	0,63	0,53	0,54	0,59	0,77	0,54	0,56	0,58	0,63	0,52	0,54	0,59	0,74	0,54	0,57	0,57	0,61	0,55
Halep	0,52	0,58	0,62	0,54	0,59	0,60	0,50	0,57	0,58	0,60	0,61	0,55	0,54	0,62	0,46	0,58	0,52	0,62	0,61	0,58
Kuznetsova	0,48	0,54	0,52	0,54	0,50	0,56	0,53	0,53	0,46	0,57	0,51	0,54	0,51	0,56	0,47	0,54	0,48	0,57	0,55	0,54
Radwanska	0,49	0,56	0,61	0,59	0,44	0,55	0,65	0,59	0,43	0,55	0,61	0,59	0,43	0,56	0,64	0,59	0,42	0,55	0,54	0,57
Kerber	0,50	0,57	0,56	0,55	0,56	0,61	0,60	0,55	0,52	0,62	0,58	0,54	0,56	0,61	0,66	0,54	0,52	0,62	0,55	0,58
Williams	0,57	0,65	0,61	0,66	0,65	0,64	0,50	0,66	0,71	0,63	0,61	0,67	0,65	0,64	0,53	0,66	0,69	0,63	0,70	0,64
Sharapova	0,43	0,51	0,58	0,52	0,41	0,51	0,70	0,53	0,43	0,50	0,58	0,52	0,42	0,51	0,63	0,53	0,44	0,50	0,57	0,48
Wozniacki	0,53	0,60	0,66	0,54	0,63	0,62	0,65	0,57	0,57	0,63	0,66	0,54	0,63	0,62	0,64	0,57	0,58	0,63	0,58	0,61
Cibulkova	0,47	0,54	0,53	0,50	0,51	0,58	0,49	0,51	0,50	0,58	0,54	0,50	0,53	0,57	0,51	0,51	0,52	0,57	0,50	0,55
Errani	0,52	0,53	0,61	0,50	0,54	0,54	0,64	0,52	0,50	0,56	0,55	0,52	0,55	0,54	0,48	0,53	0,52	0,55	0,49	0,55
Cornet	0,48	0,50	0,35	0,51	0,69	0,47	0,37	0,47	0,68	0,49	0,38	0,50	0,67	0,47	0,43	0,46	0,65	0,49	0,53	0,49
Cibulkova	0,47	0,54	0,50	0,50	0,53	0,59	0,67	0,49	0,52	0,59	0,50	0,50	0,53	0,59	0,67	0,49	0,52	0,59	0,55	0,54
Kvitova	0,53	0,61	0,56	0,63	0,52	0,64	0,69	0,60	0,45	0,65	0,56	0,63	0,52	0,64	0,69	0,60	0,45	0,65	0,65	0,59
Giorgi	0,48	0,57	0,62	0,60	0,37	0,59	0,71	0,60	0,40	0,57	0,63	0,59	0,47	0,57	0,69	0,59	0,51	0,55	0,60	0,56
Pliskova	0,53	0,61	0,71	0,59	0,56	0,61	0,75	0,61	0,60	0,60	0,69	0,59	0,60	0,60	0,69	0,62	0,64	0,59	0,63	0,61
V. Williams	0,53	0,59	0,54	0,60	0,67	0,56	0,37	0,61	0,73	0,55	0,56	0,60	0,65	0,56	0,46	0,60	0,71	0,56	0,56	0,59
Stosur	0,48	0,53	0,53	0,51	0,55	0,55	0,67	0,51	0,53	0,55	0,56	0,50	0,55	0,55	0,80	0,50	0,53	0,55	0,52	0,54
Radwanska	0,53	0,62	0,61	0,61	0,57	0,64	0,44	0,62	0,53	0,64	0,62	0,60	0,61	0,63	0,55	0,61	0,59	0,64	0,67	0,60
Keys	0,48	0,56	0,58	0,60	0,64	0,49	0,65	0,58	0,68	0,49	0,57	0,61	0,64	0,49	0,61	0,59	0,68	0,49	0,62	0,54
V. Williams	0,47	0,60	0,66	0,57	0,55	0,61	0,63	0,60	0,54	0,61	0,65	0,57	0,53	0,62	0,60	0,60	0,52	0,62	0,55	0,61
S. Williams	0,53	0,66	0,70	0,68	0,62	0,64	0,76	0,68	0,50	0,66	0,70	0,67	0,66	0,62	0,75	0,68	0,57	0,65	0,66	0,66
Gavrilova	0,56	0,73	0,75	0,79	0,55	0,68	0,69	0,79	0,46	0,68	0,75	0,79	0,55	0,68	0,69	0,79	0,46	0,68	0,67	0,74
Garcia	0,44	0,58	0,58	0,63	0,50	0,57	0,40	0,62	0,57	0,55	0,58	0,63	0,50	0,57	0,40	0,62	0,57	0,55	0,54	0,60
S.Zhang	0,46	0,51	0,53	0,46	0,49	0,55	0,67	0,47	0,47	0,55	0,53	0,46	0,49	0,55	0,67	0,47	0,47	0,55	0,49	0,51
Konta	0,54	0,59	0,58	0,59	0,46	0,64	0,47	0,60	0,44	0,63	0,58	0,59	0,46	0,64	0,47	0,60	0,44	0,63	0,59	0,59

Tabella 15. Probabilità Stimate, Uomini

Da una prima osservazione di queste tabelle è estremamente complesso trarre indicazioni circa eventuali differenze tra i vari stati del match. Tuttavia, è comunque possibile recuperare qualche informazione di carattere generale e ottenere un primo confronto tra uomini e donne. In primo luogo si può notare come, sia per gli uomini che per le donne, la probabilità di vincere un punto è sempre maggiore nel caso in cui si consideri il servizio rispetto al caso in cui vengano considerati tutti i punti giocati. Da questo punto di vista i dati rispecchiano quanto era lecito attendersi sottolineando come il servizio rappresenti un ruolo chiave nel tennis dando ulteriore valore alla definizione degli stati cui si è fatto ricorso. Se questo comportamento risulta comune per uomini e donne, non altrettanto uniforme è l'importanza in termini quantitativi che la componente del servizio riveste nei due contesti. Considerando infatti la differenza tra  $p_0$  e  $p_{00}$  si può notare come questa risulti più marcata nel caso delle coppie maschili rispetto a quello delle coppie femminili:

	$P_{00}$		$P_0$		$P_0 - P_{00}$	
	Uomini	Donne	Uomini	Donne	Uomini	Donne
<b>Min</b>	0.461	0.433	0.548	0.499	0.028	0.015
<b>Max</b>	0.562	0.567	0.723	0.729	0.221	0.17
<b>Range</b>	0.101	0.134	0.175	0.230	0.193	0.155
<b>Media</b>	0.502	0.499	0.642	0.578	0.139	0.782

Tabella 16. Confronto tra  $p_0$  e  $p_{00}$

La tabella di cui sopra, mette in luce le differenze tra il contesto maschile e il contesto femminile. Ad eccezione di qualche isolato caso, analizzando le coppie femminili emerge che la probabilità di vincere un punto al servizio è mediamente inferiore rispetto a quanto osservato per gli uomini. Osservando infine, le probabilità riportate in Tabella 14 e 15 relative alle donne, si può notare come i casi in cui il servizio acquisisce maggior rilevanza sono relativi a due particolari giocatrici. Il primo, e per certi versi sorprendente, caso riguarda

la Gavrilova la cui probabilità di vincere un punto al servizio contro la rivale Garcia si attesta al 73 %. Decisamente meno sorprendente è, invece, il caso di Serena Williams la cui probabilità di vincere un punto al servizio è del 65 % e del 66 % rispettivamente contro Maria Sharapova e la sorella Venus Williams. Questi primi semplici confronti, che permettono di denotare al meglio le differenze tra le due realtà considerate, trovano inoltre riscontro anche in letteratura. Già Klaassen e Magnus, infatti, avevano messo in luce come la componente del servizio assumesse un ruolo minore se riferita alla realtà femminile.

Sempre sulla scia di quanto proposto in *Analyzing Wimbledon* dai due autori sopra citati si è deciso di valutare le coppie considerate sulla base di un nuovo parametro, definito come:

$$\delta = p_A - p_B.$$

Il parametro  $\delta$  permette, sulla base della definizione data da Klaassen e Magnus di valutare la differenza di qualità tra differenti giocatori. I due argomentano infatti che la probabilità di vincere un punto al servizio non dipenda solo dalla capacità al servizio, *serv*, di un giocatore ma anche dall'abilità nel ricevere dell'avversario, *rec*. In questo senso:

$$p_A = \text{serv}_A - \text{rec}_B \text{ e } p_B = \text{serv}_B - \text{rec}_A.$$

Sulla base di queste definizioni, l'interpretazione di  $\delta$  risulta immediata: tanto maggiore sarà la differenza in termini qualitativi tra due giocatori, tanto maggiori saranno le capacità al servizio e in ricezione dell'uno rispetto all'altro e di conseguenza, tanto più diverso da zero risulterà essere il valore del parametro. Se poi, sarà il giocatore A ad avere un vantaggio in termini di qualità il parametro assumerà valore positivo, viceversa nel caso in cui risultasse essere B il giocatore con la più alta probabilità di vincere un punto al servizio.

Giocatori	Delta
Djokovic vs Federer	-0,004
Federer – Nadal	-0,019
Berdych – Ferrer	0,007
Del Potro – Federer	-0,06
Federer – Ferrer	0,092
Nadal – Fognini	0,069
Goffin – Tsonga	-0,052
Tipsarevic – Dimitrov	-0,018
Verdasco – Lopez	0,057
Seppi – Haase	0,046
Seppi – Muller	0,031
Struff – Kohlschreiber	-0,041
Herbert – Struff	-0,06
Isner – Lopez	-0,005
Fognini – Vinolas	0,068
Murray – Querrey	0,097
Murray – Dimitrov	0,05
Wawrinka – Paire	0,043
Raonic – Bautista	0,046
Thiem – Goffin	-0,039
<b>Media</b>	<b>0,0154</b>

Giocatrici	Delta
Kerber - Pliskova	0,031
Halep - Kuznetsova	0,043
Radwanska - Kerber	-0,015
Williams - Sharapova	0,135
Wozniacki - Cibulkova	0,063
Errani- Cornet	0,035
Cibulkova - Kvitova	-0,068
Giorgi - Pliskova	-0,045
V. Williams - Stosur	0,055
Radwanska - Keys	0,058
V. Williams - S. Williams	-0,065
Gavrilova - Garcia	0,149
S. Zhang - Konta	-0,081
<b>Media</b>	<b>0,023</b>

Tabella 17. Delta

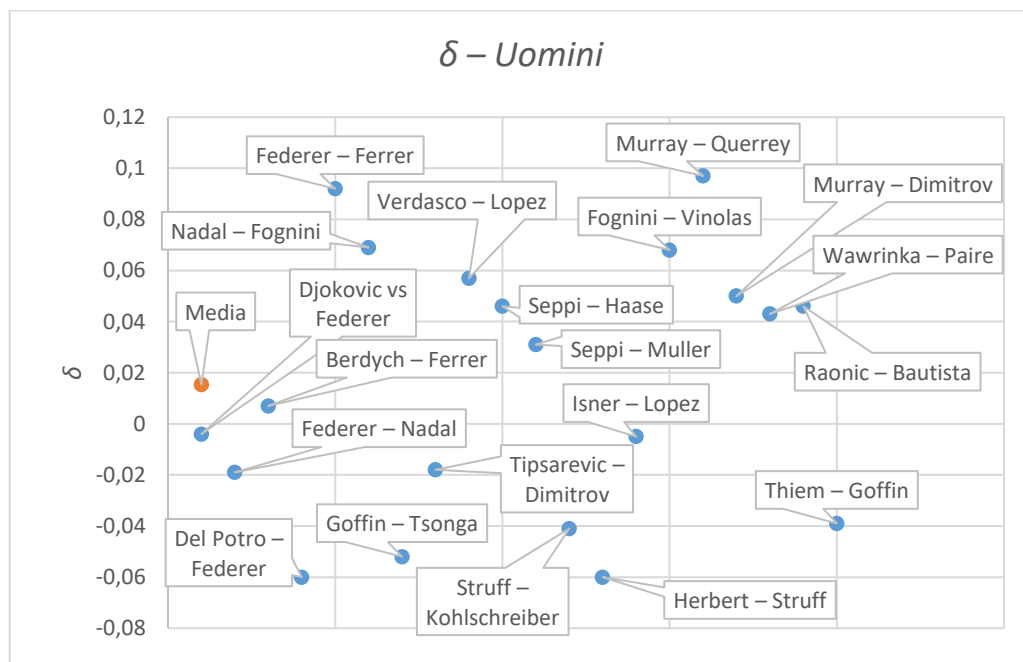


Figura 3.  $\delta$  - Uomini

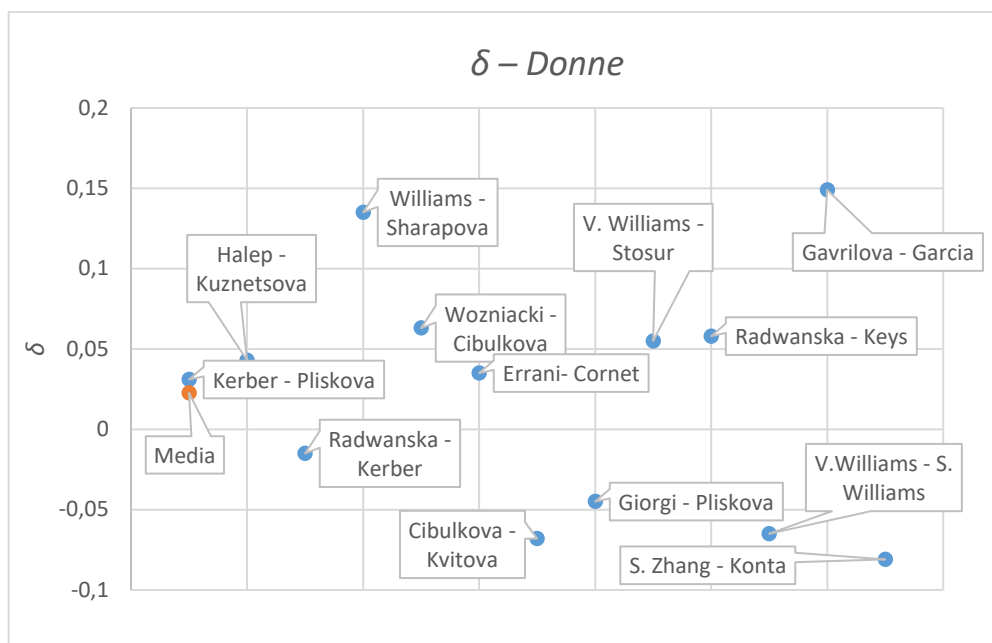


Figura 4.  $\delta$  - Donne

I due grafici e la tabella riportati permettono, in sintesi, di avere un'indicazione relativa ai giocatori considerati in termini di qualità relativa rispetto all'avversario. Alla luce di queste nuove informazioni trovano conferme le analisi preliminari svolte sui ranking ribadendo così quanto già espresso in sede di presentazione dei dati. Le coppie considerate coinvolgono infatti sia giocatori comparabili in termini di qualità, sia che essi siano al vertice delle classifiche o meno, che giocatori tra cui sia osservabile un divario.

Per concludere, partendo dalla definizione di stati del match sono state stimate le probabilità di vincere un punto al servizio nelle varie situazioni di gioco. Questo passaggio risulterà fondamentale nel prosieguo dell'elaborato dato che  $p_{A,j}$  e  $p_{B,j}$  rappresentano gli input necessari all'algoritmo utilizzato per le simulazioni Monte Carlo. Sulla base delle probabilità stimate è stato inoltre possibile argomentare circa alcune differenze tra le coppie considerate, per lo più tra uomini e donne, adducendo ulteriori elementi utili nel chiarire i dati a disposizione.

### 4.3.2. Simulazioni Monte Carlo

L'approccio non parametrico sotto forma di simulazione Monte Carlo rappresenta, assieme alla definizione di stato del match, una delle novità di questo elaborato rispetto a quanto già presentato in letteratura. Concettualmente, il metodo Monte Carlo permette di generare un nuovo campione statistico, partendo da un campione di riferimento, selezionando per  $n$  volte con prefissata probabilità un elemento dal campione stesso. In questo modo il nuovo, o i nuovi campioni ottenuti, seppur fittizi avranno la stessa distribuzione del campione originario. Partendo dunque dalla definizione data, sono state considerate le sequenze dei punti vinti e persi al servizio sotto l'ipotesi nulla di indipendenza e in riferimento a ciascuno stato del match. Per evitare un onere computazionale troppo oneroso da sostenere si è deciso di non seguire la strada del puro ricampionamento punto per punto ma si è deciso di considerare le probabilità di vincere un punto al servizio, stimate per ciascuno dei dodici stati del match considerati. Ancora una volta, per ogni sequenza head-to-head, per ognuno dei giocatori in essa coinvolti e per ciascuno degli stati del match, si è voluto verificare che le probabilità di vincere un set,  $p_i^S$  per  $i =$  giocatore A, B, e un match,  $p_i^M$  per  $i =$  giocatore A, B non dipendessero dallo stato del match relativo. Ovvero, è stato considerato un sistema d'ipotesi del tipo:

$$H_0: p_{i,0}^S = p_{i,j}^S \quad H_1: p_{i,0}^S \neq p_{i,j}^S;$$

$$H_0: p_{i,0}^M = p_{i,j}^M \quad H_1: p_{i,0}^M \neq p_{i,j}^M;$$

per  $i = A, B$  e  $j = 1, \dots, 12$ .

Per rispondere ad una tale esigenza è stato elaborato un algoritmo in grado di simulare il completo svolgimento di un incontro di tennis partendo dalla probabilità di vincere il punto. Riproducendo, dunque, punto per punto lo

svolgimento di un incontro è stato possibile sfruttare le diverse probabilità stimate a seconda dei giocatori coinvolti e dello stato del match relativo al punto corrente. È importante sottolineare come, in sede di simulazione, non si sia tenuto conto di chi fosse il primo giocatore a servire in un match e in un set; in questo senso è stato arbitrariamente definito come primo giocatore a servire il giocatore A. Questa scelta si fonda su quanto dimostrato da Klaassen e Magnus: se è vero che il servizio è determinante in termini di probabilità di vincere un game, altrettanto non si può affermare circa l'importanza del primo servizio di un set e di un match. L'impatto di questa situazione può dunque essere tralasciato.

Sfruttando un siffatto algoritmo, per ogni sequenza head-to-head composta da un numero  $m$  di incontri, per ogni giocatore coinvolto e per i dodici stati del match sono state giocate tramite simulazione 2000 sequenze ciascuna delle quali costituita da  $m$  incontri sfruttando le probabilità stimate sotto  $H_0$ , nello specifico  $p_{A,0}$  e  $p_{B,0}$ , e sotto  $H_1$ , in particolare  $p_{A,j}$  e  $p_{B,j}$  per  $j = 1, \dots, 12$ . Dopo aver condotto le simulazioni si è proceduto con l'ottenere le stime Monte Carlo delle probabilità di vincere un set e un match:

$$\hat{p}_{i,0}^S, \hat{p}_{i,j}^S \quad \text{e} \quad \hat{p}_{i,0}^M, \hat{p}_{i,j}^M;$$

per  $i = A, B$  e  $j = 1, \dots, 12$ .

Inoltre, l'aver replicato 2000 volte le simulazioni di sequenze composte da  $m$  incontri ha permesso di poter lavorare non solo con le stime Monte Carlo delle probabilità ma anche con le loro distribuzioni:

$$\hat{p}_{i,0,r}^S, \hat{p}_{i,j,r}^S \quad \text{e} \quad \hat{p}_{i,0,r}^M, \hat{p}_{i,j,r}^M;$$

per  $i = A, B$  e  $j = 1, \dots, 12$  e  $r = 1, \dots, 2000$ .

Sfruttando, infine, le distribuzioni Monte Carlo e i quantili empirici, di livello 0.025 e 0.975, calcolati sulla distribuzione ottenuta sotto l'ipotesi nulla, è stato



possibile procedere alla verifica dell'ipotesi di indipendenza e identica distribuzione.

Il grafico che segue, riferito alla distribuzione di probabilità di vincere il match nel caso Federer-Nadal, è esemplificativo del procedimento seguito per condurre le analisi. In particolare, come esempio è stata considerato lo stato del match relativo ai soli Game Points. Nel grafico sono riportati, oltre alla distribuzione Monte Carlo sotto  $H_0$ , i quantili empirici e la probabilità stimata di vincere il punto nello stato del match preso come esempio.

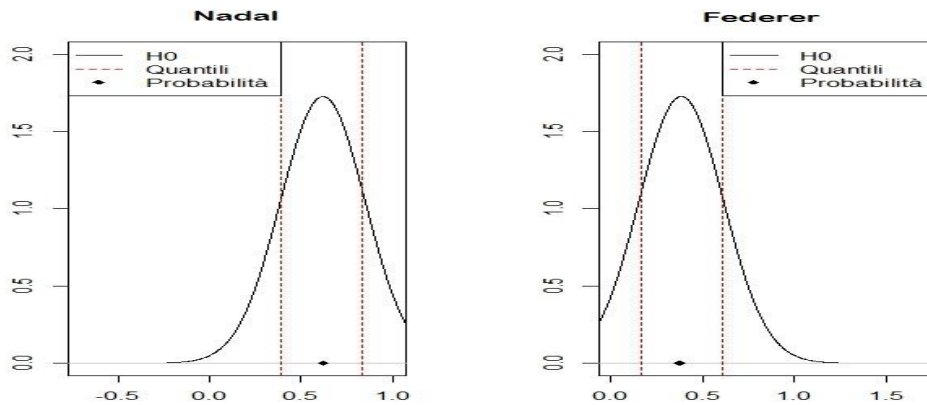


Figura 5. Distribuzioni Monte Carlo Federer-Nadal

In riferimento a quanto illustrato in precedenza, si è voluta testare l'ipotesi di indipendenza valutando se le probabilità stimate sui dati originali risultassero appartenenti alla regione definita dai quantili della distribuzione Monte Carlo ottenuta sotto  $H_0$ . Detto altrimenti, per i diversi stati del match considerati sono stati calcolati i p-value empirici definiti come:

$$p - \text{value}_{\text{emp}} = \hat{p}_{i,j,r} > p_{i,j}$$

per  $i = A, B$  e  $j = 1, \dots, 12$  e  $r = 1, \dots, 2000$

Nelle pagine successive sono riportate le tabelle contenenti i p-value empirici calcolati in relazione alle distribuzioni Monte Carlo per le probabilità di vincere set e match, sia per gli uomini che per le donne. In rosso sono stati evidenziati i p-value inferiori al 5% sintomatici di deviazioni dall'ipotesi di indipendenza.

P-VALUE EMPIRICO - SET - UOMINI							
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente
Djokovic	0,508	0,518	0,554	0,517	0,484	0,218	0,484
Federer	0,492	0,483	0,446	0,483	0,516	0,782	0,516
Federer	0,532	0,607	0,761	0,607	0,496	0,547	0,514
Nadal	0,469	0,394	0,239	0,394	0,504	0,454	0,486
Berdych	0,499	0,341	0,317	0,567	0,485	0,793	0,341
Ferrer	0,502	0,659	0,684	0,434	0,504	0,207	0,659
Del Potro	0,505	0,563	0,386	0,458	0,531	0,623	0,421
Federer	0,495	0,438	0,614	0,542	0,469	0,378	0,579
Federer	0,525	0,525	0,406	0,585	0,525	0,613	0,585
Ferrer	0,476	0,476	0,595	0,416	0,476	0,387	0,416
Nadal	0,316	0,381	0,202	0,172	0,202	0,285	0,381
Fognini	0,684	0,620	0,798	0,828	0,798	0,715	0,620
Goffin	0,455	0,538	0,538	0,455	0,455	0,455	0,391
Tsonga	0,546	0,463	0,463	0,546	0,546	0,546	0,609
Tipsarevic	0,471	0,542	0,399	0,399	0,523	0,523	0,523
Dimitrov	0,529	0,459	0,602	0,602	0,477	0,477	0,477
Verdasco	0,509	0,388	0,595	0,595	0,388	0,509	0,595
Lopez	0,492	0,612	0,405	0,405	0,612	0,492	0,405
Seppi	0,521	0,658	0,445	0,521	0,590	0,521	0,521
Haase	0,479	0,343	0,556	0,479	0,410	0,479	0,479
Seppi	0,507	0,486	0,718	0,614	0,576	0,507	0,486
Muller	0,493	0,515	0,283	0,387	0,425	0,493	0,515
Struff	0,440	0,440	0,229	0,440	0,440	0,440	0,440
KohlSchreiber	0,560	0,560	0,772	0,560	0,560	0,560	0,560
Herbert	0,454	0,454	0,357	0,454	0,454	0,454	0,454
Struff	0,547	0,547	0,643	0,547	0,441	0,547	0,547
Isner	0,466	0,466	0,209	0,387	0,325	0,547	0,466
Lopez	0,534	0,534	0,792	0,614	0,675	0,454	0,534
Fognini	0,516	0,411	0,516	0,523	0,411	0,516	0,548
Vinolas	0,485	0,590	0,485	0,478	0,590	0,485	0,453
Murray	0,453	0,615	0,770	0,615	0,615	0,615	0,453
Querrey	0,548	0,385	0,231	0,385	0,385	0,385	0,548
Wawrinka	0,528	0,528	0,470	0,528	0,528	0,528	0,576
Paire	0,472	0,472	0,531	0,472	0,472	0,472	0,425
Murray	0,478	0,475	0,388	0,475	0,506	0,520	0,477
Dimitrov	0,523	0,526	0,612	0,526	0,495	0,480	0,523
Raonic	0,473	0,590	0,467	0,467	0,514	0,395	0,514
Bautista	0,528	0,411	0,533	0,533	0,486	0,605	0,486
Thiem	0,463	0,590	0,641	0,426	0,428	0,542	0,463
Goffin	0,537	0,410	0,359	0,574	0,572	0,459	0,537

Tabella 18. P-value Empirici, Set - Uomini

P-VALUE EMPIRICO - SET - UOMINI					
Giocatori	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Djokovic	0,345	0,543	0,373	0,506	0,595
Federer	0,655	0,457	0,628	0,494	0,405
Federer	0,468	0,597	0,585	0,482	0,533
Nadal	0,533	0,403	0,416	0,518	0,468
Berdych	0,436	0,228	0,499	0,228	0,444
Ferrer	0,565	0,772	0,502	0,773	0,557
Del Potro	0,461	0,421	0,325	0,617	0,505
Federer	0,539	0,579	0,675	0,384	0,495
Federer	0,585	0,585	0,585	0,525	0,525
Ferrer	0,416	0,416	0,416	0,476	0,476
Nadal	0,455	0,482	0,447	0,537	0,381
Fognini	0,546	0,518	0,554	0,463	0,620
Goffin	0,455	0,391	0,391	0,455	0,538
Tsonga	0,546	0,609	0,609	0,546	0,463
Tipsarevic	0,399	0,542	0,399	0,542	0,399
Dimitrov	0,602	0,459	0,602	0,459	0,602
Verdasco	0,666	0,509	0,666	0,509	0,595
Lopez	0,335	0,492	0,335	0,492	0,405
Seppi	0,537	0,590	0,521	0,590	0,521
Haase	0,464	0,410	0,479	0,410	0,479
Seppi	0,375	0,614	0,614	0,576	0,614
Muller	0,626	0,387	0,387	0,425	0,387
Struff	0,440	0,604	0,355	0,604	0,527
Kohlschreiber	0,560	0,397	0,646	0,397	0,474
Herbert	0,454	0,303	0,454	0,357	0,560
Struff	0,547	0,698	0,547	0,643	0,441
Isner	0,387	0,325	0,325	0,325	0,466
Lopez	0,614	0,676	0,676	0,676	0,534
Fognini	0,548	0,516	0,548	0,514	0,578
Vinolas	0,453	0,485	0,453	0,486	0,423
Murray	0,453	0,615	0,453	0,451	>0,99
Querrey	0,548	0,385	0,548	0,549	<0,001
Wawrinka	0,557	0,528	0,557	0,576	0,528
Paire	0,443	0,472	0,443	0,425	0,472
Murray	0,600	0,388	0,478	0,389	>0,99
Dimitrov	0,401	0,612	0,523	0,612	<0,001
Raonic	0,510	0,467	0,467	0,550	>0,99
Bautista	0,490	0,533	0,533	0,451	<0,001
Thiem	0,463	0,542	0,395	0,426	0,463
Goffin	0,537	0,459	0,605	0,574	0,537

Tabella 19. P-value Empirici, Set - Uomini

P-VALUE EMPIRICO - SET - DONNE												
Giocatrici	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Kerber	0,467	0,335	0,528	0,528	0,339	0,528	0,546	0,546	0,467	0,546	0,467	0,434
Pliskova	0,533	0,666	0,472	0,472	0,661	0,472	0,454	0,454	0,533	0,454	0,533	0,566
Halep	0,509	0,569	0,673	0,509	0,515	0,509	0,509	0,515	0,569	0,509	0,509	0,515
Kuznetsova	0,492	0,432	0,327	0,492	0,486	0,492	0,492	0,486	0,432	0,492	0,492	0,486
Radwanska	0,502	0,398	0,553	0,555	0,450	0,542	0,630	0,653	0,653	0,712	0,653	0,502
Kerber	0,498	0,602	0,448	0,446	0,550	0,458	0,370	0,347	0,347	0,288	0,347	0,498
S. Williams	0,595	0,620	0,595	0,595	0,595	0,595	0,595	0,595	0,693	0,620	0,620	0,538
Sharapova	0,406	0,380	0,380	0,380	0,406	0,406	0,406	0,406	0,307	0,380	0,380	0,462
Wozniacki	0,492	0,609	0,617	0,446	0,546	0,446	0,492	0,446	0,496	0,432	0,492	0,492
Cibulkova	0,509	0,392	0,345	0,554	0,455	0,554	0,509	0,554	0,504	0,569	0,509	0,509
Errani	0,498	0,302	0,565	0,498	0,302	0,565	0,562	0,498	0,295	0,498	0,295	0,650
Cornet	0,503	0,698	0,436	0,454	0,698	0,436	0,439	0,503	0,706	0,503	0,706	0,351
Cibulkova	0,444	0,579	NA	0,444	0,579	0,444	0,441	0,441	0,509	0,371	0,509	0,525
Kvitova	0,557	0,421	NA	0,557	0,421	0,557	0,559	0,559	0,492	0,630	0,492	0,475
Giorgi	0,543	0,311	0,462	0,462	0,391	0,543	0,462	0,462	0,391	0,462	0,391	0,462
Pliskova	0,457	0,689	0,539	0,539	0,609	0,457	0,539	0,539	0,609	0,539	0,609	0,539
V. Williams	0,530	0,530	0,620	0,530	0,530	0,530	0,530	0,530	0,681	0,530	0,741	0,530
Stosur	0,470	0,470	0,381	0,470	0,470	0,470	0,470	0,470	0,320	0,470	0,259	0,470
Radwanska	0,575	0,379	0,421	0,575	0,421	0,731	0,465	0,575	0,379	0,575	0,379	>0,999
Keys	0,426	0,622	0,580	0,426	0,580	0,269	0,535	0,426	0,622	0,426	0,622	<0,001
V. Williams	0,460	0,460	0,604	0,497	0,396	0,396	0,396	0,396	0,451	0,396	0,346	0,497
S. Williams	0,540	0,540	0,396	0,503	0,604	0,605	0,605	0,605	0,549	0,604	0,654	0,503
Gavrilova	0,405	0,405	NA	0,405	0,729	0,729	0,729	0,729	0,729	0,729	0,729	0,405
Garcia	0,596	0,596	NA	0,596	0,271	0,271	0,271	0,271	0,271	0,271	0,271	0,596
S.Zhang	0,479	0,582	NA	0,373	0,545	0,479	0,479	0,479	0,545	0,373	0,545	0,479
Konta	0,521	0,419	NA	0,627	0,455	0,521	0,521	0,521	0,455	0,627	0,455	0,521

Tabella 20. P-value Empirici, Set - Donne

P-VALUE EMPIRICO - MATCH - UOMINI							
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente
Djokovic	0,505	0,505	0,505	0,505	0,505	0,157	0,505
Federer	0,495	0,495	0,495	0,495	0,495	0,843	0,495
Federer	0,573	0,573	0,747	0,573	0,573	0,211	0,573
Nadal	0,428	0,428	0,254	0,428	0,428	0,789	0,428
Berdych	0,591	0,343	0,343	0,591	0,591	0,814	0,343
Ferrer	0,409	0,658	0,658	0,409	0,409	0,187	0,658
Del Potro	0,399	0,631	0,399	0,399	0,399	0,631	0,399
Federer	0,602	0,370	0,602	0,602	0,602	0,370	0,602
Federer	0,745	0,352	0,352	0,745	0,352	0,352	0,745
Ferrer	0,256	0,649	0,649	0,256	0,649	0,649	0,256
Nadal	0,519	0,519	0,166	0,166	0,166	0,519	0,519
Fognini	0,481	0,835	0,481	0,835	0,835	0,481	0,481
Goffin	0,500	0,500	0,500	0,500	0,500	0,500	0,500
Tsonga	0,501	0,501	0,501	0,501	0,501	0,501	0,501
Tipsarevic	0,529	0,529	0,529	0,529	0,529	0,529	0,529
Dimitrov	0,471	0,471	0,471	0,471	0,471	0,471	0,471
Verdasco	0,305	0,305	0,741	0,741	0,305	0,305	0,741
Lopez	0,695	0,695	0,260	0,260	0,695	0,695	0,260
Seppi	0,470	0,789	0,470	0,470	0,470	0,470	0,470
Haase	0,530	0,212	0,530	0,530	0,530	0,530	0,530
Seppi	0,406	0,406	0,754	0,754	0,406	0,406	0,406
Muller	0,595	0,595	0,246	0,246	0,595	0,595	0,595
Struff	0,350	0,350	0,350	0,350	0,350	0,350	0,350
Kohlschreiber	0,651	0,651	0,651	0,651	0,651	0,651	0,651
Herbert	0,649	0,232	0,232	0,649	0,649	0,232	0,232
Struff	0,351	0,769	0,769	0,351	0,351	0,769	0,769
Isner	0,438	0,438	0,185	0,438	0,438	0,735	0,438
Lopez	0,563	0,563	0,816	0,563	0,563	0,266	0,563
Fognini	0,653	0,351	0,653	0,653	0,351	0,351	0,653
Vinolas	0,347	0,650	0,347	0,347	0,650	0,650	0,347
Murray	0,717	0,717	0,717	0,717	0,717	0,717	0,717
Querrey	0,283	0,283	0,283	0,283	0,283	0,283	0,283
Wawrinka	0,484	0,484	0,484	0,484	0,484	0,484	0,484
Paire	0,517	0,517	0,517	0,517	0,517	0,517	0,517
Murray	0,469	0,469	0,469	0,469	0,469	0,469	0,469
Dimitrov	0,532	0,532	0,532	0,532	0,532	0,532	0,532
Raonic	0,639	0,639	0,639	0,639	0,639	0,227	0,639
Bautista	0,362	0,362	0,362	0,362	0,362	0,774	0,362
Thiem	0,447	0,447	0,447	0,447	0,447	0,447	0,447
Goffin	0,554	0,554	0,554	0,554	0,554	0,554	0,554

Tabella 21. P-value Empirici, Match – Uomini

P-VALUE EMPIRICO - MATCH - UOMINI					
Giocatori	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Djokovic	0,409	0,542	0,409	0,542	0,542
Federer	0,591	0,459	0,591	0,459	0,459
Federer	0,382	0,573	0,573	0,382	0,573
Nadal	0,618	0,428	0,428	0,618	0,428
Berdych	0,343	0,343	0,591	0,343	0,591
Ferrer	0,658	0,658	0,409	0,658	0,409
Del Potro	0,399	0,399	0,399	0,631	0,399
Federer	0,602	0,602	0,602	0,370	0,602
Federer	0,352	0,745	0,745	0,352	0,352
Ferrer	0,649	0,256	0,256	0,649	0,649
Nadal	0,519	0,519	0,519	0,519	0,166
Fognini	0,481	0,481	0,481	0,481	0,835
Goffin	0,500	0,500	0,500	0,500	0,500
Tsonga	0,501	0,501	0,501	0,501	0,501
Tipsarevic	0,529	0,529	0,529	0,529	0,178
Dimitrov	0,471	0,471	0,471	0,471	0,823
Verdasco	0,741	0,305	0,741	0,305	0,741
Lopez	0,260	0,695	0,260	0,695	0,260
Seppi	0,470	0,470	0,470	0,470	0,470
Haase	0,530	0,530	0,530	0,530	0,530
Seppi	0,406	0,754	0,754	0,406	0,406
Muller	0,595	0,246	0,246	0,595	0,595
Struff	0,350	0,776	0,350	0,776	0,350
Kohlschreiber	0,651	0,225	0,651	0,225	0,651
Herbert	0,649	0,232	0,649	0,232	0,649
Struff	0,351	0,769	0,351	0,769	0,351
Isner	0,438	0,438	0,438	0,438	0,438
Lopez	0,563	0,563	0,563	0,563	0,563
Fognini	0,653	0,653	0,653	0,351	0,653
Vinolas	0,347	0,347	0,347	0,650	0,347
Murray	0,717	0,717	0,717	0,717	>0,99
Querrey	0,283	0,283	0,283	0,283	<0,001
Wawrinka	0,484	0,484	0,484	0,484	ist
Paire	0,517	0,517	0,517	0,517	0,517
Murray	0,720	0,469	0,469	0,469	>0,99
Dimitrov	0,281	0,532	0,532	0,532	<0,001
Raonic	0,639	0,639	0,639	0,639	>0,99
Bautista	0,362	0,362	0,362	0,362	<0,001
Thiem	0,447	0,447	0,447	0,447	0,447
Goffin	0,554	0,554	0,554	0,554	0,554

Tabella 22. P-value Empirici, Match – Uomini

P-VALUE EMPIRICO - MATCH - DONNE												
Giocatrici	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Kerber	0,543	0,230	0,543	0,543	0,543	0,543	0,543	0,543	0,543	0,543	0,543	0,543
Pliskova	0,457	0,770	0,457	0,457	0,457	0,457	0,457	0,457	0,457	0,457	0,457	0,457
Halep	0,385	0,677	0,677	0,385	0,385	0,385	0,385	0,385	0,677	0,385	0,385	0,385
Kuznetsova	0,616	0,323	0,323	0,616	0,616	0,616	0,616	0,616	0,323	0,616	0,616	0,616
Radwanska	0,540	0,540	0,540	0,540	0,540	0,540	0,540	0,540	0,540	0,763	0,540	0,540
Kerber	0,460	0,460	0,460	0,460	0,460	0,460	0,460	0,460	0,460	0,237	0,460	0,460
S. Williams	0,624	0,624	0,624	0,624	0,624	0,624	0,624	0,624	0,624	0,624	0,624	0,624
Sharapova	0,377	0,377	0,377	0,377	0,377	0,377	0,377	0,377	0,377	0,377	0,377	0,377
Wozniacki	0,363	0,667	0,667	0,363	0,667	0,363	0,363	0,363	0,667	0,363	0,363	0,363
Cibulkova	0,638	0,333	0,333	0,638	0,333	0,638	0,638	0,638	0,333	0,638	0,638	0,638
Errani	0,390	0,390	0,711	0,390	0,390	0,711	0,711	0,390	0,390	0,390	0,390	0,711
Cornet	0,610	0,610	0,289	0,610	0,610	0,289	0,289	0,610	0,610	0,610	0,610	0,289
Cibulkova	0,373	0,373	NA	0,373	0,373	0,373	0,373	0,373	0,373	0,373	0,373	0,373
Kvitova	0,627	0,627	NA	0,627	0,627	0,627	0,627	0,627	0,627	0,627	0,627	0,627
Giorgi	0,310	0,310	0,310	0,310	0,310	0,310	0,310	0,310	0,310	0,310	0,310	0,310
Pliskova	0,691	0,691	0,691	0,691	0,691	0,691	0,691	0,691	0,691	0,691	0,691	0,691
V. Williams	0,667	0,667	0,667	0,667	0,667	0,667	0,667	0,667	0,667	0,667	0,667	0,667
Stosur	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334
Radwanska	0,376	0,376	0,376	0,376	0,376	0,795	0,376	0,376	0,376	0,376	0,376	0,998
Keys	0,624	0,624	0,624	0,624	0,624	0,206	0,624	0,624	0,624	0,624	0,624	0,003
V. Williams	0,489	0,489	0,489	0,489	0,489	0,489	0,489	0,489	0,489	0,489	0,489	0,489
S. Williams	0,512	0,512	0,512	0,512	0,512	0,512	0,512	0,512	0,512	0,512	0,512	0,512
Gavrilova	0,866	0,866	NA	0,866	0,866	0,866	0,866	0,866	0,866	0,866	0,866	0,866
Garcia	0,134	0,134	NA	0,134	0,134	0,134	0,134	0,134	0,134	0,134	0,134	0,134
S.Zhang	0,551	0,551	NA	0,551	0,551	0,551	0,551	0,551	0,551	0,551	0,551	0,551
Konta	0,450	0,450	NA	0,450	0,450	0,450	0,450	0,450	0,450	0,450	0,450	0,450

Tabella 23. P-value Empirici, Match – Donne



I risultati dei test mostrano chiaramente come il risultato di set e match non venga influenzato dalle differenti probabilità di vincere il punto calcolate in relazione ai diversi stati del match considerati. In altre parole, ciò che l'approccio basato sulle simulazioni sembra suggerire è una chiara evidenza verso la conferma dell'ipotesi nulla di indipendenza dei punti. I pochi e isolati casi in cui si manifestano deviazioni rispetto all'ipotesi da verificare appaiono dunque come situazioni sporadiche e occasionali piuttosto che sintomatiche di una tendenza generale nei dati. I risultati presentati confermano e danno maggior forza alle analisi di stampo parametrico condotte in precedenza.

Le indicazioni fornite dai p-value empirici, seppur chiare e precise, non possono comunque ritenersi sufficienti a fornire un'indicazione definitiva riguardante i possibili effetti dei singoli stati del match rispetto le probabilità di vincere set o match. Allo scopo di ottenere indicazioni più chiare e precise si è deciso di confrontare le distribuzioni Monte Carlo ottenute sotto ipotesi di indipendenza e sotto assunzione di deviazione rispetto all'ipotesi nulla, al fine di valutarne eventuali differenze statisticamente significative. I criteri sulla base dei quali è stato effettuato il confronto sono due: in primo luogo è stato condotto un test non parametrico per il confronto di distribuzioni. Il test deputato a questo scopo è il test di Kolmogorov-Smirnov. Lo scopo del test è di valutare la bontà di adattamento in relazione al calcolo della distanza fra la vera funzione di ripartizione dei dati,  $\hat{F}_n(\cdot)$ , e la funzione di ripartizione ipotizzata,  $F_0(\cdot)$ . La statistica test si presenta dunque nella forma:

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)|.$$

Valori grandi della statistica  $D_n$ , portano a concludere che le distribuzioni risultano tra loro differenti.

Per quel che concerne il secondo confronto, data la distribuzione Monte Carlo relativa al j-esimo stato del match, per  $j = 1, \dots, 12$ , è stata considerata la frazione di valori eccedenti le bande di confidenza del 5% costruite sulla distribuzione Monte Carlo ottenuta sotto l'ipotesi nulla. In questo modo, dunque, frazioni di

punti inferiori al 5% risultano sintomatiche di assenza di deviazioni dall'ipotesi di indipendenza.

Al fine di fornire un esempio grafico relativo al confronto tra distribuzioni, si consideri quanto riportato di seguito. I grafici in figura mostrano un confronto riguardante la coppia Giorgi-Pliskova relativo alle distribuzioni Monte Carlo ottenute sotto  $H_0$  e sotto una specifica  $H_1$ , in questo caso relativa allo stato del match riguardante Break- o Set- Points. Sono state considerate le distribuzioni relative alla probabilità di vincere un set e un match al fine di valutare anche graficamente eventuali differenze, non solo tra le giocatrici, ma anche tra le due situazioni dell'incontro.

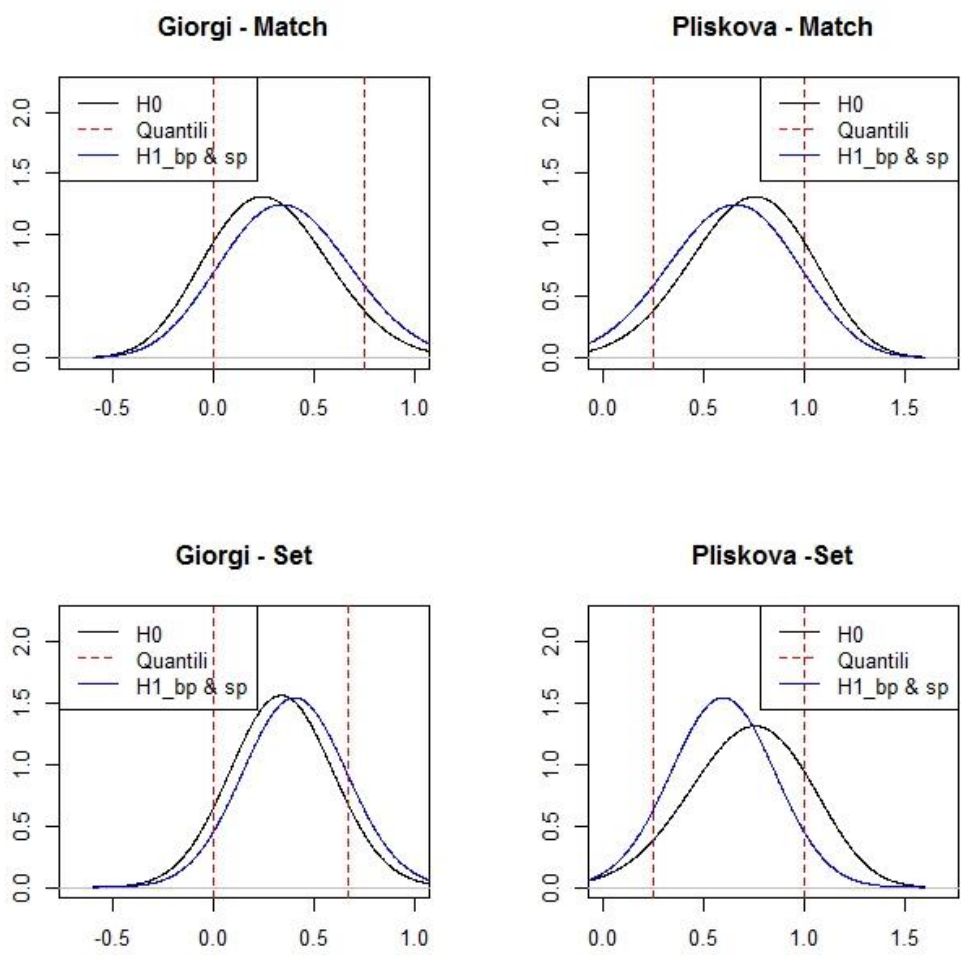


Figura 6. Confronto Distribuzioni Monte-Carlo

A conclusione delle analisi vengono riportate le tabelle relative, rispettivamente, ai p-value dei test di Kolmogorov-Smirnov e alle frazioni dei valori eccedenti la soglia del 5% in relazione al procedimento precedentemente illustrato.

Per quel che riguarda le tabelle relative ai test di Kolmogorov-Smirnov, al fine di facilitarne l'interpretazione e la lettura, sono stati evidenziati in rosso i p-value inferiori al 5%. Ovvero, sono stati evidenziati quei casi in cui le distribuzioni sono risultate essere statisticamente diverse tra loro. Dall'analisi delle tabelle di seguito riportate, si può concludere che in diversi casi il test porta a rifiutare l'ipotesi nulla, soprattutto nel caso delle distribuzioni Monte Carlo relative ai set. È bene, tuttavia, sottolineare che il test risulta particolarmente sensibile anche a piccole differenze tra le distribuzioni. Un esempio di questo comportamento è dato dal grafico "*Giorgi-Set*", riportato in Figura 6. Il p-value del test di Kolmogorov-Smirnov per il caso in questione risulta  $<0.001$  indicando come le distribuzioni siano tra loro statisticamente diverse nonostante una prima analisi grafica possa far dubitare di ciò. Anche per ovviare a questo si è deciso di procedere con la valutazione della frazione di osservazioni, calcolate in riferimento alle distribuzioni Monte Carlo ottenute sotto  $H_1$ , eccedenti i quantili 0.0025 e 0.975 della distribuzione ottenuta tramite simulazione sotto  $H_0$ . Nelle tabelle che seguono sono stati evidenziati in rosso quei valori, superiori al 5%, indicanti una deviazione significativa rispetto all'ipotesi nulla di riferimento. Dall'analisi dei risultati si può notare come i casi i cui si assiste a queste deviazioni, soprattutto nel caso delle distribuzioni riferite ai match, rappresentino l'eccezione piuttosto che la regola. Anche le analisi non parametriche, infine, sembrano confermare l'assunzione di indipendenza dei punti. Così come suggerito dal modello di regressione logistica, anche in questo caso non si riscontrano chiari e marcati effetti degli stati considerati sulla probabilità di vincere set e match.

Giocatori	KOLMOGOROV SMIRNOV - SET - UOMINI											
	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Djokovic Federer	0,002	0,390	<0,001	<0,001	0,257	<0,001	<0,001	<0,001	0,001	<0,001	0,970	<0,001
Federer Nadal	0,185	<0,001	<0,001	<0,001	0,665	0,012	0,212	0,010	<0,001	<0,001	0,349	0,022
Berdych Ferrer	0,770	<0,001	<0,001	<0,001	0,029	<0,001	<0,001	<0,001	<0,001	0,586	<0,001	0,011
Del Potro Federer	0,310	<0,001	<0,001	0,002	0,045	<0,001	<0,001	0,008	0,002	<0,001	<0,001	0,665
Federer Ferrer	0,692	0,960	<0,001	<0,001	0,095	<0,001	0,003	0,257	0,226	<0,001	0,413	0,883
Nadal Fognini	0,310	0,000	<0,001	0,075	0,000	<0,001	<0,001	<0,001	0,016	<0,001	0,863	0,000
Goffin Tsonga	0,919	0,042	0,001	0,919	0,718	0,088	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Tipsarevic Dimitrov	0,111	<0,001	<0,001	<0,001	0,226	0,274	0,150	<0,001	<0,001	<0,001	0,001	0,000
Verdasco Lopez	0,370	<0,001	<0,001	<0,001	<0,001	<0,001	0,001	<0,001	0,999	<0,001	<0,001	0,172
Seppi Haase	0,795	<0,001	0,018	0,024	<0,001	0,020	0,103	0,150	<0,001	0,172	<0,001	0,064
Seppi Muller	0,198	0,004	<0,001	<0,001	0,002	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Struff Kohlschreiber	0,198	<0,001	<0,001	<0,001	0,744	0,075	0,391	0,129	<0,001	<0,001	<0,001	0,241
Herbert Struff	0,718	0,008	<0,001	0,795	0,172	0,292	0,018	0,212	<0,001	0,129	<0,001	<0,001
Isner Lopez	0,436	0,919	<0,001	<0,001	<0,001	0,001	<0,001	<0,001	<0,001	0,000	<0,001	<0,001
Fognini Vinolas	0,413	<0,001	0,560	0,508	<0,001	0,064	0,129	0,011	0,665	0,020	0,139	0,013
Murray Querrey	0,161	0,001	<0,001	0,665	0,744	<0,001	0,413	0,011	0,795	0,948	<0,001	0,000
Murray Dimitrov	0,172	0,009	<0,001	0,006	0,534	0,883	0,035	<0,001	<0,001	0,103	<0,001	<0,001
Wawrinka Benoit	0,718	0,226	0,001	0,612	0,919	0,744	0,006	0,029	0,059	0,024	0,007	0,460
Raonic Bautista	0,024	<0,001	<0,001	<0,001	0,665	<0,001	0,139	0,120	0,000	0,001	0,150	<0,001
Thiem Goffin	0,172	<0,001	<0,001	<0,001	<0,001	0,005	0,139	0,639	0,054	<0,001	<0,001	0,274

Tabella 24. P-value Kolmogorov-Smirnov, Set - Uomini

KOLMOGOROV SMIRNOV - MATCH - UOMINI												
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Djokovic Federer	0,002	0,998	<0,001	0,999	0,198	<0,001	0,018	<0,001	0,001	<0,001	0,999	<0,001
Federer Nadal	0,436	<0,001	<0,001	<0,001	0,665	0,000	0,770	0,013	<0,001	0,001	0,198	0,985
Berdych Ferrer	>0,999	<0,001	<0,001	<0,001	0,795	<0,001	<0,001	0,002	<0,001	0,970	<0,001	0,139
Del Potro Federer	0,935	<0,001	<0,001	0,003	0,035	<0,001	0,032	0,172	0,005	<0,001	<0,001	0,007
Federer Ferrer	0,770	0,819	<0,001	0,013	0,960	<0,001	0,042	0,919	0,484	0,003	0,902	0,045
Nadal Fognini	>0,999	0,001	<0,001	0,111	0,002	<0,001	<0,001	0,054	0,744	0,002	0,413	<0,001
Goffin Tsonga	0,978	0,075	0,020	0,960	1,000	0,082	<0,001	0,003	0,960	0,586	<0,001	<0,001
Tipsarevic Dimitrov	0,413	<0,001	<0,001	<0,001	0,902	0,005	0,560	<0,001	0,001	<0,001	0,003	<0,001
Verdasco Lopez	0,436	<0,001	0,003	0,001	<0,001	<0,001	0,007	<0,001	0,993	<0,001	0,004	0,329
Seppi Haase	>0,999	<0,001	0,010	0,292	<0,001	<0,001	0,902	0,612	0,001	0,329	<0,001	0,996
Seppi Muller	0,639	0,103	<0,001	<0,001	0,075	<0,001	<0,001	<0,001	<0,001	<0,001	0,005	0,045
Struff Kohlschreiber	0,150	<0,001	<0,001	0,001	1,000	0,970	0,863	0,560	<0,001	<0,001	<0,001	0,005
Herbert Struff	0,560	0,045	<0,001	1,000	0,460	0,001	0,172	0,198	<0,001	0,257	<0,001	0,161
Isner Lopez	0,999	0,993	<0,001	<0,001	<0,001	<0,001	0,054	<0,001	<0,001	<0,001	<0,001	<0,001
Fognini Vinolas	0,985	<0,001	>0,999	0,935	<0,001	<0,001	0,948	0,212	>0,999	0,534	0,484	0,948
Murray Querrey	>0,999	0,035	<0,001	>0,999	>0,999	<0,001	>0,999	>0,999	>0,999	>0,999	0,003	<0,001
Murray Dimitrov	0,257	0,257	<0,001	0,050	0,612	<0,001	0,484	<0,001	<0,001	0,718	<0,001	<0,001
Wawrinka Benoit	0,770	0,770	0,015	>0,999	>0,999	0,744	0,004	0,103	0,069	0,120	0,007	0,639
Raonic Bautista	0,129	<0,001	0,004	<0,001	>0,999	<0,001	0,586	0,212	<0,001	0,042	0,883	<0,001
Thiem Goffin	0,391	<0,001	<0,001	0,011	0,009	<0,001	0,795	0,534	0,069	<0,001	0,007	0,002

Tabella 25. P-value Kolmogorov-Smirnov, Match – Uomini

KOLMOGOROV SMIRNOV - SET - DONNE												
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Kerber Pliskova	<0,001	<0,001	0,111	0,139	<0,001	0,391	0,001	<0,001	0,050	0,029	0,015	<0,001
Halep Kuznetsova	0,241	<0,001	<0,001	0,935	0,770	0,005	0,370	0,883	<0,001	0,150	0,842	0,534
Radwanska Kerber	0,391	<0,001	0,004	<0,001	0,000	0,111	0,000	0,000	<0,001	<0,001	0,000	0,349
S. Williams Sharapova	0,985	0,032	0,978	0,639	0,935	0,000	0,960	0,935	<0,001	0,413	<0,001	0,198
Wozniacki Cibulkova	0,004	<0,001	0,000	0,001	0,161	<0,001	0,001	<0,001	0,935	0,000	0,013	0,005
Errani Cornet	0,985	<0,001	0,012	0,948	<0,001	<0,001	0,009	0,013	<0,001	0,111	<0,001	<0,001
Cibulkova Kvitova	0,919	<0,001	NA	0,010	0,009	0,024	0,018	0,005	0,095	<0,001	0,095	<0,001
Giorgi Pliskova	0,006	<0,001	0,560	0,948	0,001	<0,001	0,978	>0,99	<0,001	0,744	<0,001	0,863
V. Williams Stosur	0,150	0,082	<0,001	0,349	0,508	<0,001	0,460	0,999	<0,001	0,064	<0,001	0,902
Radwanska Keys	0,586	<0,001	<0,001	0,185	<0,001	<0,001	0,050	0,129	<0,001	0,050	<0,001	<0,001
V. Williams S. Williams	0,310	0,883	<0,001	0,172	<0,001	<0,001	<0,001	<0,001	0,026	<0,001	<0,001	0,612
Gavrilova Garcia	0,883	<0,001	NA	0,970	0,059	<0,001	<0,001	<0,001	0,095	<0,001	0,007	<0,001
S. Zhang Konta	0,391	<0,001	NA	0,005	<0,001	0,744	0,508	0,003	0,010	<0,001	0,050	0,198

Tabella 26. P-value Kolmogorov-Smirnov, Set – Donne

KOLMOGOROV SMIRNOV - MATCH - DONNE												
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Kerber Pliskova	0,001	<0,001	0,349	0,902	<0,001	<0,001	0,045	0,006	0,088	0,241	0,018	0,005
Halep Kuznetsova	0,212	0,003	0,000	0,902	0,902	<0,001	0,586	0,902	<0,001	0,413	0,665	0,999
Radwanska Kerber	0,484	<0,001	0,022	0,004	0,001	<0,001	0,000	<0,001	<0,001	<0,001	<0,001	0,370
S. Williams Sharapova	>0,99	0,329	>0,99	0,770	0,999	<0,001	0,999	1,000	<0,001	>0,99	<0,001	0,985
Wozniacki Cibulkova	0,020	<0,001	<0,001	0,003	0,993	<0,001	0,026	0,002	0,863	0,000	0,198	0,002
Errani Cornet	0,948	<0,001	0,082	>0,99	<0,001	<0,001	0,054	0,069	<0,001	0,185	<0,001	<0,001
Cibulkova Kvitova	>0,99	<0,001	NA	0,020	0,292	0,026	0,009	0,069	0,534	<0,001	0,560	<0,001
Giorgi Pliskova	0,038	<0,001	0,935	0,902	0,018	>0,99	>0,99	>0,99	<0,001	0,795	<0,001	0,863
V. Williams Stosur	0,993	0,310	0,005	0,960	0,935	<0,001	0,819	0,612	<0,001	0,241	<0,001	0,960
Radwanska Keys	0,819	<0,001	<0,001	0,088	<0,001	<0,001	0,436	0,370	<0,001	0,095	<0,001	<0,001
V. Williams S. Williams	0,819	0,902	<0,001	0,069	0,002	<0,001	<0,001	<0,001	0,003	<0,001	<0,001	<0,001
Gavrilova Garcia	>0,99	0,042	NA	>0,99	>0,99	<0,001	<0,001	0,000	0,508	<0,001	>0,99	0,007
S. Zhang Konta	0,883	<0,001	NA	0,257	<0,001	0,003	>0,99	0,310	0,008	0,003	0,020	0,212

Tabella 27. P-value Kolmogorov-Smirnov, Match - Donne

FRAZIONE DEI PUNTI OLTRE LE BANDE DI CONFIDENZA 5% - SET - UOMINI												
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Djokovic Federer	0,056	0,050	0,060	0,052	0,057	0,058	0,139	0,075	0,053	0,075	0,053	0,052
Federer Nadal	0,050	0,055	0,095	0,056	0,051	0,042	0,055	0,053	0,069	0,046	0,053	0,055
Berdych Ferrer	0,045	0,061	0,065	0,046	0,043	0,093	0,053	0,046	0,108	0,039	0,110	0,040
Del Potro Federer	0,041	0,048	0,061	0,044	0,045	0,028	0,053	0,041	0,044	0,062	0,056	0,047
Federer Ferrer	0,046	0,045	0,047	0,053	0,042	0,035	0,046	0,049	0,056	0,063	0,048	0,043
Nadal Fognini	0,045	0,044	0,042	0,044	0,043	0,344	0,067	0,050	0,047	0,043	0,040	0,034
Goffin Tsonga	0,035	0,033	0,027	0,032	0,031	0,037	0,045	0,019	0,028	0,020	0,017	0,008
Tipsarevic Dimitrov	0,037	0,033	0,047	0,052	0,031	0,028	0,040	0,048	0,028	0,063	0,038	0,052
Verdasco Lopez	0,020	0,010	0,036	0,034	0,011	0,013	0,027	0,053	0,022	0,057	0,015	0,031
Seppi Haase	0,028	0,059	0,025	0,027	0,036	0,024	0,023	0,035	0,032	0,027	0,035	0,030
Seppi Muller	0,021	0,017	0,068	0,045	0,031	0,020	0,020	0,016	0,042	0,035	0,027	0,035
Struff Kohlschrei	0,024	0,039	0,086	0,036	0,029	0,027	0,028	0,027	0,013	0,047	0,011	0,030
Herbert Struff	0,016	0,022	0,034	0,013	0,017	0,014	0,029	0,018	0,041	0,017	0,030	0,011
Isner Lopez	0,051	0,043	0,126	0,052	0,061	0,032	0,048	0,058	0,071	0,083	0,084	0,044
Fognini Vinolas	0,043	0,050	0,043	0,048	0,053	0,035	0,045	0,048	0,048	0,043	0,049	0,049
Murray Querrey	0,048	0,050	0,109	0,053	0,056	0,063	0,049	0,055	0,052	0,050	0,048	>,0,999
Murray Dimitrov	0,044	0,043	0,047	0,042	0,037	0,041	0,042	0,049	0,055	0,040	0,050	>,0,999
Wawrinka Benoit	0,046	0,051	0,053	0,048	0,060	0,053	0,049	0,051	0,055	0,052	0,054	0,059
Raonic Bautista	0,032	0,055	0,029	0,033	0,036	0,007	0,034	0,035	0,028	0,031	0,044	>,0,999
Thiem Goffin	0,039	0,042	0,051	0,052	0,041	0,035	0,041	0,036	0,037	0,040	0,030	0,037

Tabella 28. Frazione di punti oltre le bande di confidenza, Set – Uomini



FRAZIONE DEI PUNTI OLTRE LE BANDE DI CONFIDENZA 5% - MATCH - UOMINI												
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Djokovic Federer	0,037	0,045	0,043	0,043	0,044	0,149	0,049	0,033	0,030	0,032	0,026	0,030
Federer Nadal	0,030	0,030	0,048	0,032	0,030	0,055	0,027	0,027	0,033	0,017	0,027	0,029
Berdych Ferrer	0,014	0,013	0,012	0,024	0,014	0,129	0,012	0,015	0,021	0,012	0,028	0,014
Del Potro Federer	0,004	0,005	0,014	0,010	0,005	0,001	0,010	0,010	0,011	0,016	0,004	0,011
Federer Ferrer	0,011	0,010	0,004	0,012	0,007	0,004	0,008	0,007	0,014	0,024	0,011	0,006
Nadal Fognini	0,012	0,008	0,006	0,012	0,010	0,497	0,035	0,008	0,010	0,010	0,020	0,002
Goffin Tsonga	0,006	0,005	0,006	0,006	0,006	0,011	0,010	0,003	0,008	0,005	0,003	0,001
Tipsarevic Dimitrov	0,027	0,023	0,044	0,054	0,024	0,020	0,030	0,051	0,012	0,057	0,022	0,114
Verdasco Lopez	0,003	0,003	0,008	0,006	0,001	0,001	0,006	0,011	0,004	0,014	0,004	0,005
Seppi Haase	0,006	0,020	0,004	0,010	0,013	0,005	0,005	0,006	0,011	0,008	0,012	0,010
Seppi Muller	0,006	0,004	0,027	0,016	0,008	0,004	0,006	0,004	0,011	0,012	0,009	0,007
Struff Kohlschrei	0,009	0,016	0,037	0,014	0,008	0,007	0,009	0,007	0,005	0,016	0,003	0,008
Herbert Struff	0,002	0,006	0,006	0,004	0,004	0,005	0,006	0,003	0,006	0,003	0,003	0,004
Isner Lopez	0,014	0,013	0,035	0,015	0,014	0,048	0,012	0,016	0,023	0,026	0,020	0,016
Fognini Vinolas	0,009	0,004	0,007	0,007	0,006	0,004	0,011	0,012	0,006	0,013	0,006	0,011
Murray Querrey	0,002	0,003	0,022	0,004	0,004	0,007	0,003	0,003	0,003	0,003	0,003	0,998
Murray Dimitrov	0,004	0,004	0,003	0,006	0,006	0,004	0,008	0,010	0,003	0,004	0,003	>0,999
Wawrinka Benoit	0,003	0,005	0,006	0,006	0,005	0,008	0,005	0,005	0,006	0,007	0,006	0,006
Raonic Bautista	0,013	0,024	0,011	0,010	0,013	0,000	0,015	0,018	0,009	0,011	0,026	>0,999
Thiem Goffin	0,017	0,006	0,006	0,022	0,019	0,009	0,015	0,015	0,010	0,021	0,015	0,019

Tabella 29. Frazione di punti oltre le bande di confidenza, Match – Uomini

FRAZIONE DEI PUNTI OLTRE LE BANDE DI CONFIDENZA 5% - SET - DONNE												
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Kerber Pliskova	0,048	0,069	0,057	0,047	0,063	0,043	0,053	0,058	0,046	0,049	0,050	0,074
Halep Kuznetsova	0,031	0,040	0,045	0,034	0,035	0,019	0,028	0,039	0,041	0,030	0,029	0,030
Radwanska Kerber	0,043	0,039	0,047	0,036	0,037	0,040	0,051	0,051	0,060	0,076	0,048	0,037
S. Williams Sharapova	0,028	0,029	0,028	0,024	0,024	0,013	0,021	0,022	0,066	0,033	0,062	0,021
Wozniacki Cibulkova	0,060	0,059	0,064	0,059	0,061	0,036	0,053	0,065	0,065	0,050	0,061	0,055
Errani Cornet	0,038	0,076	0,037	0,034	0,064	0,038	0,040	0,039	0,074	0,031	0,078	0,059
Cibulkova Kvitova	0,036	0,027	NA	0,033	0,025	0,027	0,034	0,025	0,026	0,032	0,018	0,014
Giorgi Pliskova	0,014	0,049	0,014	0,017	0,027	0,011	0,016	0,016	0,037	0,021	0,034	0,017
V. Williams Stosur	0,035	0,044	0,047	0,043	0,044	0,084	0,050	0,040	0,058	0,044	0,081	0,045
Radwanska Keys	0,019	0,003	0,011	0,023	0,012	0,057	0,015	0,025	0,002	0,024	0,005	>0,999
V. Williams S. Williams	0,020	0,023	0,016	0,020	0,023	0,029	0,038	0,033	0,024	0,024	0,043	0,017
Gavrilova Garcia	0,017	0,011	NA	0,017	0,011	0,042	0,046	0,039	0,026	0,047	0,026	0,007
S. Zhang Konta	0,013	0,003	NA	0,022	0,008	0,014	0,016	0,018	0,011	0,026	0,009	0,015

Tabella 30. Frazione di punti oltre le bande di confidenza, Set – Donne

FRAZIONE DEI PUNTI OLTRE LE BANDE DI CONFIDENZA 5% - MATCH - DONNE												
Giocatori	GamePoint	Break- or Set-Points	Tiebreak	Tie-break o Game-Points	Tie-break o Break- o Set-Points	Set	Punto Precedente	Punto Precedente & Game Point	Punto Precedente & BreakPoint & SetPoint	Punto Precedente & Game Point o Tiebreak	Punto Precedente & BreakPoint & SetPoint o Tiebreak	Importance
Kerber Pliskova	0,009	0,001	0,009	0,008	0,003	0,008	0,013	0,013	0,007	0,010	0,007	0,014
Halep Kuznetsova	0,002	0,005	0,005	0,005	0,002	0,000	0,001	0,005	0,005	0,002	0,002	0,001
Radwanska Kerber	0,032	0,031	0,038	0,036	0,029	0,049	0,044	0,052	0,053	0,068	0,046	0,027
S. Williams Sharapova	0,011	0,010	0,014	0,009	0,011	0,001	0,007	0,007	0,028	0,011	0,027	0,009
Wozniacki Cibulkova	0,008	0,017	0,010	0,006	0,011	0,002	0,006	0,004	0,009	0,001	0,007	0,005
Errani Cornet	0,016	0,005	0,023	0,015	0,003	0,025	0,020	0,007	0,002	0,009	0,001	0,054
Cibulkova Kvitova	0,004	0,003	NA	0,007	0,002	0,003	0,006	0,006	0,003	0,005	0,001	0,000
Giorgi Pliskova	0,006	0,021	0,008	0,004	0,009	0,008	0,007	0,005	0,015	0,008	0,008	0,010
V. Williams Stosur	0,012	0,016	0,021	0,013	0,015	0,037	0,016	0,008	0,038	0,016	0,050	0,013
Radwanska Keys	0,003	0,001	0,002	0,002	0,000	0,020	0,003	0,003	0,000	0,004	0,001	0,963
V. Williams S. Williams	0,013	0,010	0,006	0,010	0,014	0,021	0,023	0,021	0,014	0,015	0,027	0,007
Gavrilova Garcia	0,004	0,003	NA	0,007	0,007	0,022	0,018	0,019	0,011	0,020	0,008	0,003
S. Zhang Konta	0,023	0,006	NA	0,029	0,016	0,021	0,020	0,030	0,017	0,034	0,016	0,023

Tabella 31. Frazione di punti oltre le bande di confidenza, Match – Donne



# Conclusione

È chiaro, anche a chi ha meno affinità con la pratica sportiva, che il risultato finale di un evento sportivo sia determinato da altre componenti che esulano i meri aspetti tecnici e tattici. È manifesto, infatti, che in una sfida tra diversi atleti un ruolo chiave sia determinato dalle emozioni e dall'influenza che esse possono avere sull'una e sull'altra parte. Da questo punto di vista, un caso emblematico di tutto ciò è rappresentato senza ombra di dubbio dal tennis. In uno sport come il tennis, in cui la sfida assume i connotati di uno scontro diretto tra due giocatori, è lecito pensare che la componente mentale ed emotiva giochi un ruolo cardine fino anche ad influenzare in modo sostanziale il risultato di una partita. Ciò che, infine, è lecito aspettarsi è che quanto più un giocatore sia in grado di mediare e gestire le fasi emotive che occorrono durante un match, tanto più egli sarà in grado di affermarsi come un vincente.

Questo elaborato prende le mosse dalle considerazioni di cui sopra e si propone di verificare da un punto di vista statistico l'impatto che certe situazioni di partita hanno sul risultato finale di un incontro di tennis. L'interrogativo dal quale si è partiti riguarda dunque la proprietà di indipendenza e identica distribuzione dei punti giocati in un incontro di tennis. Declinando l'analisi statistica anche in un'ottica sportiva, ciò che si è cercato di fare è stato verificare l'impatto che certe situazioni di gioco hanno sulla probabilità di vincere un set o un match. Per far ciò, il punto di partenza è stato l'individuazione di queste particolari situazioni definite nel corso dell'elaborato, stati del match. Una volta definiti gli stati è stata calcolata la probabilità di vincere un punto al servizio nelle situazioni di indipendenza e nelle altre varie situazioni considerate. Al fine di valutare possibili deviazioni dall'ipotesi nulla in esame sono state implementate delle analisi sia di stampo parametrico, in termini di modello di regressione logistica, che di stampo non

parametrico, sotto forma di simulazioni Monte Carlo. I risultati ottenuti con i due differenti approcci risultano essere coerenti tra loro. In entrambi i casi, infatti, ciò che emerge è una generale conferma dell'ipotesi nulla di indipendenza e identica distribuzione. Non mancano situazioni in cui si assiste a deviazioni rispetto all'ipotesi di indipendenza e identica distribuzione ma, questi sembrano essere riconducibili a deviazioni occasionali piuttosto che decise e sintomatiche tendenze espresse dai dati.

Le analisi proposte, seppur basate su approcci e dati differenti rispetto a quanto già precedentemente proposto in letteratura, confermano i risultati degli autori che precedentemente avevano affrontato il problema. In questo senso si può infine concludere che, nonostante non manchino casi di deviazione rispetto all'ipotesi nulla, l'assunzione di indipendenza e identica distribuzione per i punti giocati al servizio in una partita di tennis rappresenta una valida approssimazione. Detto altrimenti, la probabilità di vincere un set o un match non risulta essere influenzata dalle varie situazioni che occorrono nell'arco di un incontro, indistintamente per gli uomini e per le donne, sia che si considerino i giocatori al vertice delle classifiche sia che si valutino giocatori di livello minore.







# Appendice A

Lo scopo di questa appendice è quello di fornire un breve compendio relativo alle regole del tennis e ad alcuni termini tecnici utilizzati nel corso dell'elaborato.

## **Regole**

Una partita di tennis è costituita da set. Ogni set, è a sua volta composto da game ed ogni game è costituito da punti. Una partita di singolo è disputata tra due giocatori, e, per definire quale dei due servirà primo nel match si esegue un lancio di moneta. Da questo momento in poi, il servizio si alterna ad ogni game.

Un punto si gioca con il giocatore al servizio che gioca la palla nel campo dell'avversario il quale è chiamato a rispondere. Se il giocatore al servizio commette un fallo di servizio, ha diritto ad un secondo servizio. Se anche questo servizio non dovesse risultare regolare, il punto verrà assegnato al giocatore in ricezione. Lo scambio inizia a seguito di una battuta regolare e termina non appena un giocatore riesca a far rimbalzare la palla per almeno due volte nel campo dell'avversario.

Un game si dichiara concluso nel momento in cui uno dei due giocatori riesce a vincere quattro punti, contati come 0, 15, 30, 40. Quando il punteggio si trova in situazione di parità sul 40-40 si verifica un caso noto come deuce. Il giocatore che vincerà il punto successivo conquisterà il primo vantaggio e dovrà conquistarne un secondo per vincere il game. In caso contrario si tornerebbe nella situazione di deuce.

Il tiebreak è un particolare tipo di game, vinto dal primo giocatore che raggiunge almeno sette punti con un vantaggio di almeno due sull'avversario. A differenza di un qualsiasi altro game, i punti nel tiebreak sono contati a

partire da 1,2,3,... Il primo giocatore a servire nel tiebreak è il giocatore che ha ricevuto nel game precedente. Dopo, il primo servizio, il turno di battuta varia ogni due punti giocati.

Un set si dichiara concluso quando un giocatore vince sei game con uno scarto di almeno due rispetto all'avversario. Nel caso di parità sul 5-5, il set può finire 7-5 giocando dunque due game normali, 7-6 giocando due game e un tiebreak oppure protrarsi fino a che un giocatore non vince due game successivi, nei casi in cui le regole del torneo non prevedano un tiebreak sul 6-6.

Un match consiste di tre o cinque set. Per le donne tutti gli incontri sono disputati al meglio dei tre set, mentre per gli uomini il numero di set giocati varia a seconda del torneo.

Ogni torneo si differenzia, infine, per la superficie di gioco e per il numero di punti che assegna in termini di classifica di ranking.

## **Termini**

**Break:** game vinto dal giocatore che riceve.

**BreakPoint:** punto che, se vinto da chi riceve, porta ad un Break.

**GamePoint:** punto che, se vinto dal giocatore in vantaggio, porta a vincere il game.

**Grande Slam:** insieme dei quattro maggiori tornei: Australian Open, French Open, meglio noto come Roland Garros, Wimbledon, US Open.

**MatchPoint:** punto che, se vinto dal giocatore in vantaggio, porta a vincere la partita.

**Ranking:** classifica ufficiale dei giocatori, aggiornata settimanalmente, basata sui punti accumulati dai giocatori durante l'anno. Il miglior piazzamento è il numero 1.

**SetPoint:** punto che, se vinto dal giocatore in vantaggio, porta a vincere il set.

**Tiebreak:** particolare tipo di game giocato sul punteggio di 6 game pari, per decretare il vincitore del set

# Bibliografia

- Azzalini A. (2001). *Inferenza Statistica. Una presentazione basata sul concetto di verosimiglianza*. Springer.
- Barnett T. and Clarke S. (2005). *Combining player statistics to predict outcomes of tennis matches*. *Journal of Management Mathematics* (2005) 16, 113-120.
- Barnett T. J. (2006). *Mathematical modelling in hierarchical games with specific reference to tennis*. PhD thesis, Swinburne University of Technology, Melbourne
- Clowes S., Graeme C., Tomljanovic L. (2013). *Dynamic evaluation of conditional probabilities of winning a tennis match*. *Proceedings of the Sixth Australian Conference on Mathematics and Computers in Sport, 2002*, pp. 112 - 118
- Knight G. and O'Donoghue P. (2012). *The probability of winning break points in Grand Slam men's tennis*. *European Journal of Sport Science*, Vol. 12, Iss. 6 (2012), pp. 462-468.
- Lewis M. (2003). *Moneyball: The art of winning an unfair game*. W.W. Norton & Company.
- Madurska A. M. (2012), *A set by set analysis method for predicting the outcomes of professional singles tennis matches*. Imperial College London.
- Magnus Jan R. and Klaassen F. J. G. M. (1999). *On the advantage of serving first in a tennis set: four years at Wimbledon*. *Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 48, Iss. 2 July 1999, pp 247-256.
- Magnus Jan R. and Klaassen F. J. G. M. (1999). *The effect of New Balls in tennis: four years at Wimbledon*. *Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 48, Iss. 2 July 1999, pp 239-246.

- Magnus Jan R. and Klaassen F. J. G. M. (1999). *The final set in a tennis match: four years at Wimbledon*. Journal of Applied Statistics, Vol. 26, Iss. 4 1999, pp 461-468.
- Magnus Jan R. and Klaassen F. J. G. M. (2001). *Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model*. Journal of the American Statistical Association, 96:454, 500:509
- Magnus Jan R. and Klaassen F. J. G. M. (2014). *Analyzing Wimbledon. The power of statistics*. Oxford University Press, New York.
- McHale I. and Morton A. (2011). *A Bradley-Terry type model for forecasting match results*. International Journal of Forecasting, 619-630.
- Morris C. (1977). *The most important points in tennis*. In Optimal Strategies in Sport, eds. S.P. Ladany and R. Machol, Amsterdam: North-Holland, 131:140.
- O'Donoghue P. (2013). *The most important points in Grand Slam single tennis*. Research Quarterly for Exercise and Sport vol. 72.
- Pace L. e Salvan A. (2012). *Introduzione alla Statistica II. Inferenza, verosimiglianza, modelli*. Cedam.
- Paulden T. (2016, June). *Smashing the racket*. Significance, Vol. 3, Iss. 3, pp 16-21.
- Pollard G. and Pollard G. (2011). *Applying statistical test for the independence of points in tennis*. The 11th Australasian Conference on Maths and Computers in Sport (11M&CS), Melbourne.
- Siegel S., Castellan J (1992). *Nonparametric Statistics for the Behavioural Sciences*. McGraw-Hill.
- Spedicato G. A., Kang T. S., Yalamanchi S. B. and Yadav D. (2016). *The marchovchain package: a package for easily handling discrete markov chains in R*. URL <https://CRAN.R-project.org/package=markovchain>

