Universiteit Utrecht

Game, set en maths

De wiskunde van tennis

Renske Nijssen (5680581)
Begeleider:
Dr. K. Dajani

Bachelorscriptie

21 december 2018

Inhoudsopgave

1 Inleiding

Tennis is balsport waarbij twee spelers (of vier, in het dubbelspel) het tegen elkaar opnemen. Het doel is om de bal over het net in het veld van de tegenstander te slaan, zodat deze niet meer teruggeslagen kan worden en er een punt gescoord wordt.
In dit document wordt de wiskunde achter een tenniswedstrijd nader bekeken en uitgelicht. Het artikel van Newton en Keller [1] staat hierin centraal. Het artikel is gebruikt voor de bewijzen van de kansen, die in dit document nader worden toegelicht. Echter zijn alle figuren en tabellen uniek, hiervoor is ook andere data gebruikt, namelijk die van de wedstrijden van 2018. Hierdoor is dit artikel weldegelijk van toegevoegde waarde voor de al bestaande literatuur.

In dit document wordt steeds de kans berekend voor een speler om te winnen. Om dit te kunnen doen, moet er voor iedere speler bepaald worden hoe goed deze speler is. Dit wordt gedaan met behulp van p^R, dit is de kans dat speler A een rally weet te winnen. Een rally begint met de service, waarna beide spelers een aantal keer overslaan totdat een van de twee spelers een punt weet te maken. De kans om een rally te winnen zegt dus iets over hoe goed een speler is. Is deze kans heel hoog, dan is de kans dat deze speler een wedstrijd weet te winnen ook hoog.
Wanneer iemand een hele goede service heeft, komt er al meteen druk op de ander te staan, waardoor het lastig wordt de rally nog te winnen. Vandaar dat er met het berekenen van de kansen ook steeds bekeken moet worden welke speler aan het serveren is. Over het algemeen geldt namelijk, zeker in het mannentennis, dat de kans dat je je eigen service houdt, een stuk hoger is dan de kans om de service van de tegenstander te pakken. De kans om een wedstrijd te winnen wordt dan ook gebaseerd op de kans dat een speler een punt maakt, wanneer hij zelf serveert. Dit laatste wordt aangeduidt met p_A^R, hiervoor staat de A voor het feit dat speler A aan het serveren is, de R staat voor 'rally', en de p wordt gebruikt omdat degene die serveert ook degene is die het punt maakt. Wanneer de andere speler het punt maakt, wordt dit dus aangeduid met q_A^R, waarbij $q_A^R = 1 - p_A^R$. In het document zullen heel wat verschillende notaties gebruikt worden. Het subscript zal dus steeds aangeven welke speler serveert, het superscript geeft aan wat voor iets het betreft (rally, game, set) en de p geeft de kans aan dat de speler die serveert ook wint. Als de q wordt gebruikt, wint de speler die niet serveert. Ter verduidelijking staat er in de Appendix een tabel met alle mogelijke notaties en hun betekenissen.

Omdat een wedstrijd is opgebouwd in games en sets, is ervoor gekozen om dit document ook op die manier op te bouwen. Allereerst wordt er in hoofdstuk 1 gekeken naar de kans voor een speler om een game te winnen. Een game is opgebouwd uit meerdere rally's, dus geldt, hoe groter de kans om een rally te winnen, hoe groter de kans om een game te winnen. Omdat een game verschillende uitslagen kan hebben, moeten alle mogelijke uitslagen meegenomen worden bij het berekenen van deze kans.
In hoofdstuk 2 zal de kans berekend worden om een set te winnen. Een set is opgebouwd uit games waarbij de spelers steeds om en om serveren. Het berekenen van de kans van een set zal dan ook bestaan uit alle mogelijke gamestanden, waarbij steeds rekening gehouden moet worden met welke speler aan het serveren is. Aan het einde van dit hoofstuk zal ook nog worden bewezen dat het niet van belang is welke speler de wedstrijd begint met serveren; de kans om de set te winnen blijkt hier namelijk niet van af te hangen.

Wanneer de kans om een set te winnen bekend is, kan er met behulp van deze kans berekend worden hoe groot de kans is om de wedstrijd te winnen. Bij sommige toernooien win je de wedstrijd wanneer je 2 sets weet te winnen, terwijl de mannen bij de grandslams [1] 3 sets moeten winnen om zich winnaar te mogen noemen. Deze twee wedstrijdvormen zullen dan ook beide behandeld worden.

In hoofdstuk 5 zal aan de hand van deze kans bekeken worden hoe groot de kans is om uiteindelijk een heel toernooi te winnen. Hierbij wordt er uitgegaan van een toernooi waaraan 128 spelers deelnemen. Met de kans berekent in hoofdstuk 4 kan dus voor alle mogelijke wedstrijden berekend worden hoe groot de kans is dat iedere speler weet te winnen. Met behulp van matrices worden dan voor iedere ronde de mogelijke tegenstanders en hun kansen gegeven. Uiteindelijk leidt dit tot een 128×1 kolomvector waarin ieder element de kans op toernooiwinst aangeeft voor een bepaalde speler.

Om dit te verduidelijken zal de kans berekend worden voor de halve finalisten van een toernooi om het toernooi uiteindelijk te winnen. Dus wanneer de halve finale is bereikt, wat is dan de kans om uiteindelijk het toernooi te winnen. Dit zal getest worden op twee van de vier grootste en belangrijkste tennistoernooien. Er wordt bekeken of de theoretische kans om het toernooi te winnen voor iedere halve finalist overeenkomt met het daadwerkelijke verloop van het toernooi. Dit is weergegeven in hoofdstuk 6.

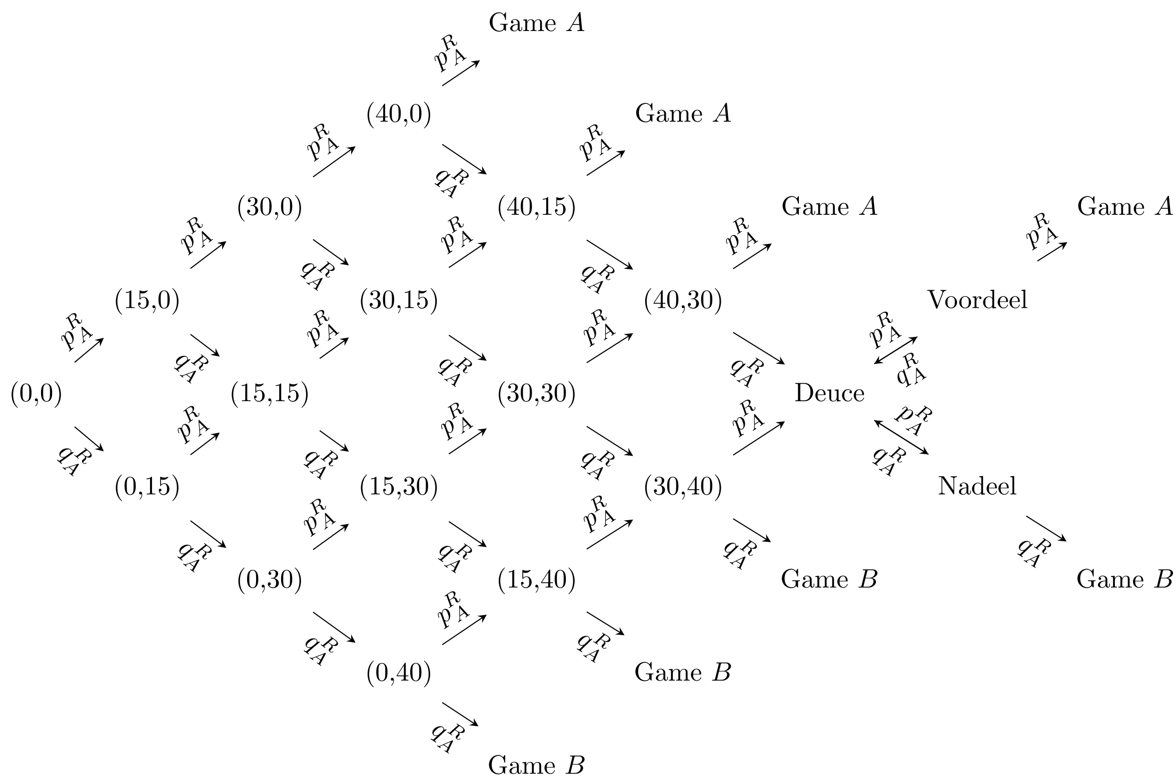
Bij het berekenen van de kansen wordt er van uitgegaan dat de punten in een tenniswedstrijd onafhankelijk gelijk verdeeld zijn (i.i.d.). In hoofdstuk 7 wordt er aan de hand van bestaande literatuur bekeken of het wel geoorloofd is om hier vanuit te gaan.

In het laatste hoofdstuk, hoofdstuk 8, wordt alles nog een keer samengevat en een conclusie getrokken over de behaalde resultaten.

[1] De grandslams zijn de vier belangrijkste en meest prestigieuze toernooien in het tennis

2 De kans op het winnen van een game

In dit hoofdstuk wordt de kans berekend dat een speler een game wint. Een game is opgebouwd uit punten. Een game wordt gewonnen wanneer er vier punten worden gemaakt, die geteld worden als $15 - 30 - 40 -$'game'. Er geldt hier echter wel dat het verschil in punten altijd minimaal twee moet zijn. Wanneer beide spelers drie punten hebben gemaakt en de stand dus 40-40 is, wordt dit aangeduid met 'deuce'. Wanneer de serveerder dan het punt pakt, wordt het 'voordeel'. Als hij dan weer een punt pakt, is het game. Echter, is het voordeel en pakt de ander de punt, wordt het weer deuce. Dit gaat dan net zo lang door, totdat iemand twee punten achter elkaar pakt na een stand van deuce. In figuur 1 wordt het verloop van een game grafisch weergegeven.



Figuur 1: Het verloop van een game waarin speler A serveert

p_A^R wordt gedefiniëerd als de kans dat speler A een rally wint wanneer A serveert. p_A^G is de kans dat speler A een game wint, terwijl A serveert. Met $p_A^G(i,j)$ wordt de kans weergegeven dat speler A i punten heeft, terwijl speler B er j heeft, wanneer speler A serveert.
Een game kan dus gewonnen worden door vier punten te scoren, terwijl je tegenstander er maar 0, 1 of 2 scoort, óf je kunt de game winnen nadat deuce is bereikt. Deze twee gevallen moeten apart worden bekeken.
Het eerste geval, waar speler A vier punten pakt, terwijl speler B er maximaal twee heeft gescoord, kan genoteerd worden door de volgende sommatie: $\sum_{j=0}^{2} p_A^G(4,j)$.

Bij het tweede geval is er sprake van deuce. Om deuce te bereiken moeten beide spelers 3 punten hebben gepakt, wat wordt aangegeven met $p_A^G(3,3)$.
Wanneer deuce is bereikt, kan speler A alsnog de game pakken, door op deuce twee punten achter elkaar te pakken. Stel nu dat speler B n punten pakt, dan betekent dat dus dat speler A $n+2$ punten moet halen. Dit wordt weergegeven door $p_A^{DC}(n+2,n)$, waarbij p_A^{DC} dus de kans is dat speler A de game weet te behalen, wanneer deuce is bereikt.
Dit resulteert in de volgende formule voor de kans op het winnen van de game

$$p_A^G = \sum_{j=0}^{2} p_A^G(4,j) + p_A^G(3,3) \sum_{n=0}^{\infty} p_A^{DC}(n+2,n) \tag{1}$$

Allereerst zal $p_A^{DC}(n+2,n)$ berekend worden.
Om, als speler A zijnde, de game te pakken wanneer deuce is bereikt, moeten de laatste twee punten die gescoord zijn, allebei voor speler A zijn. Voor de punten daarvoor moet steeds gelden dat het eerst voordeel dan wel nadeel wordt en dat dan de ander het punt pakt waardoor het weer deuce wordt.
Als er n keer de stand van deuce voorkomt, zal het dus n keer voorkomen dat ofwel eerst A het punt pakt en dan B, ofwel dat eerst B het punt pakt en dan A.
Stel nu dat het j keer voorkomt dat eerst A het punt pakt en dan B $(p_A^R q_A^R)$. Dan moet er dus $n-j$ keer voorkomen dat eerst B het punt pakt en dan A $(q_A^R p_A^R)$. Dit resulteert in het volgende.

$$\begin{aligned} p_A^{DC}(n+2,n) &= \sum_{j=0}^{n} (p_A^R)^2 (p_A^R q_A^R)^j (q_A^R p_A^R)^{n-j} \binom{n}{j} \qquad (2) \\ &= (p_A^R)^2 (p_A^R q_A^R)^n \sum_{j=0}^{n} \binom{n}{j} \\ &= (p_A^R)^{n+2} (q_A^R)^n 2^n \\ &= (p_A^R)^2 (2 p_A^R q_A^R)^n \end{aligned}$$

Dan geldt dat

$$\sum_{n=0}^{\infty} p_A^{DC}(n+2,n) = \sum_{n=0}^{\infty} (p_A^R)^2 (2p_A^R q_A^R)^n = \frac{(p_A^R)^2}{1 - 2p_A^R q_A^R} = \frac{(p_A^R)^2}{(p_A^R + q_A^R)^2 - 2p_A^R q_A^R} = \frac{(p_A^R)^2}{(p_A^R)^2 + (q_A^R)^2} \tag{3}$$

Dit kan nu ingevuld worden in (1), dus

$$p_A^G = \sum_{j=0}^{2} p_A^G(4,j) + p_A^G(3,3) \frac{(p_A^R)^2}{(p_A^R)^2 + (q_A^R)^2} \tag{4}$$

Met combinarotiek kunnen nu de kansen $p_A^R(4,j)$ voor $j = 0, 1, 2$ en $p_A^R(3,3)$ worden berekend. Voor $p_A^G(4,0)$ geldt dat alle vier de punten naar speler A moeten gaan. De kans hierop is dus $p_A^G(4,0) = (p_A^R)^4$.

Voor $p_A^G(4,1)$ geldt dat van de vijf punten er vier naar speler A gaan. Echter moet gelden dat het laatste punt van de game voor speler A is. Immers, wanneer dit niet het geval is, is de game namelijk al klaar na 4 punten. Van de eerste vier punten gaan er dus drie naar speler A en eentje naar speler B, dit kan op $\binom{4}{1}$ manieren. De kans
$p_A^G(4,1) = \binom{4}{1}(p_A^R)^3(q_A^R)^1(p_A^R) = 4(p_A^R)^4(q_A^R)$

$p_A^G(4,2)$ wordt op eenzelfde wijze berekend. Er gaan weer vier punten naar speler A en in dit geval twee naar speler B. Ook hier moet gelden dat het laatste punt ten alle tijden naar speler A moet gaan. Van de eerste vijf punten zijn er dan dus nog drie voor A en twee voor B;
$p_A^G(4,2) = \binom{5}{2}(p_A^R)^3(q_A^R)^2(p_A^R) = 10(p_A^R)^4(q_A^R)^2$

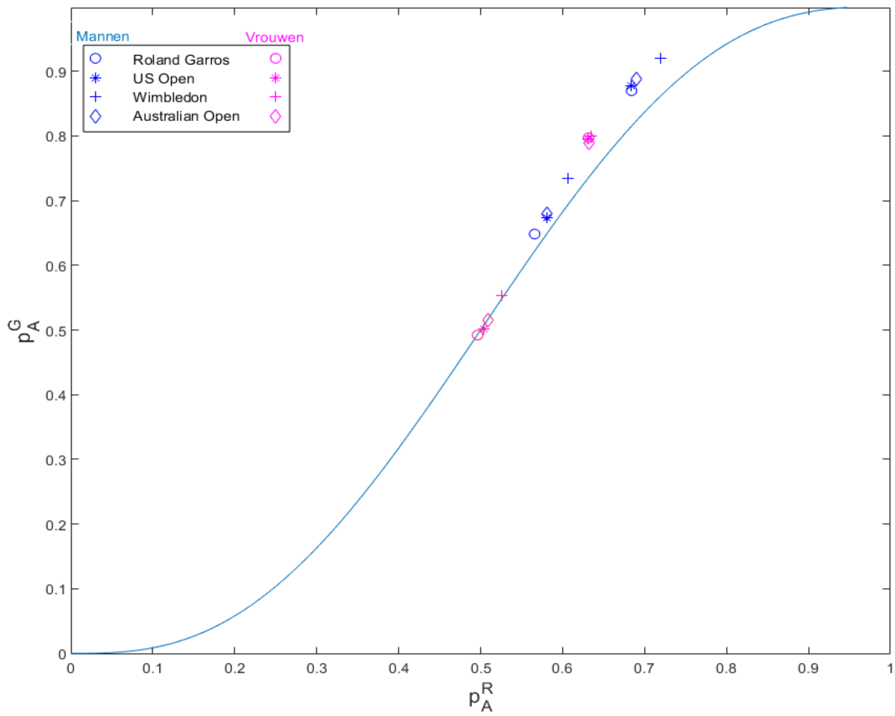
Deuce wordt het, wanneer beide speler drie punten weten te behalen. Hier zijn geen restricties op de volgorde; $p_A^G(3,3) = \binom{6}{3}(p_A^R)^3(q_A^R)^3 = 20(p_A^R)^3(q_A^R)^3$
De kans voor speler A om een game te winnen wanneer speler A zelf serveert, wordt dus gegeven door dit alles in te vullen in (4):

$$\begin{aligned} p_A^G &= \sum_{j=0}^{2} p_A^G(4,j) + p_A^G(3,3)\sum_{n=0}^{\infty} p_A^{DC}(n+2,n) \\ &= (p_A^R)^4 + 4(p_A^R)^4(q_A^R) + 10(p_A^R)^4(q_A^R)^2 + 20(p_A^R)^3(q_A^R)^3\frac{(p_A^R)^2}{(p_A^R)^2+(q_A^R)^2} \\ &= (p_A^R)^4(1+4q_A^R+10(q_A^R)^2) + 20(p_A^R q_A^R)^3\frac{(p_A^R)^2}{(p_A^R)^2+(q_A^R)^2} \end{aligned} \tag{5}$$

Dit kan ook nog geschreven worden in alleen maar termen van p_A^R, immers $q_A^R = 1 - p_A^R$. Dit verder uitwerken resulteert uiteindelijk in het volgende.

$$p_A^G = \frac{(p_A^R)^4(-8(p_A^R)^3 + 28(p_A^R)^2 - 34p_A^R + 15)}{2(p_A^R)^2 - 2p_A^R + 1} \tag{6}$$

Om te bekijken of (6) een goed beeld van de werkelijkheid geeft, is gekeken of dit overeenkomt met de data van de wedstrijden van de vier grandslams van 2018. De gebruikte data is te vinden bij [4] voor de wedstrijden van de mannen en [5] voor de wedstrijden van de vrouwen. Van al deze wedstrijden was bekend hoeveel punten een speler heeft gemaakt op zijn service en hoe vaak hij geserveerd had. Hierdoor kon bepaald worden wat p_A^R is.
$(p_A^R = \frac{\text{aantal gewonnen punten eigen service}}{\text{totaal aantal punten geserveerd}})$
Verder was ook bekend hoevaak diegene een service game gewonnen had en hoeveel games iedere speler in totaal heeft geserveerd. Hiermee kon berekend worden wat p_A^G was.
$(p_A^G = \frac{\text{aantal gewonnen service games}}{\text{totaal aantal games geserveerd}})$
Voor iedere wedstrijd was dus voor beide spelers p^R en p^G bekend. Per grandslam zijn de spelers opgedeeld in de winnaars en in de verliezers en voor beide groepen is het de gemiddelde waarde voor p_A^R en p_A^G berekend. Deze zijn in figuur 2 geplot, samen met de theoretische kans om een game te winnen, (6).



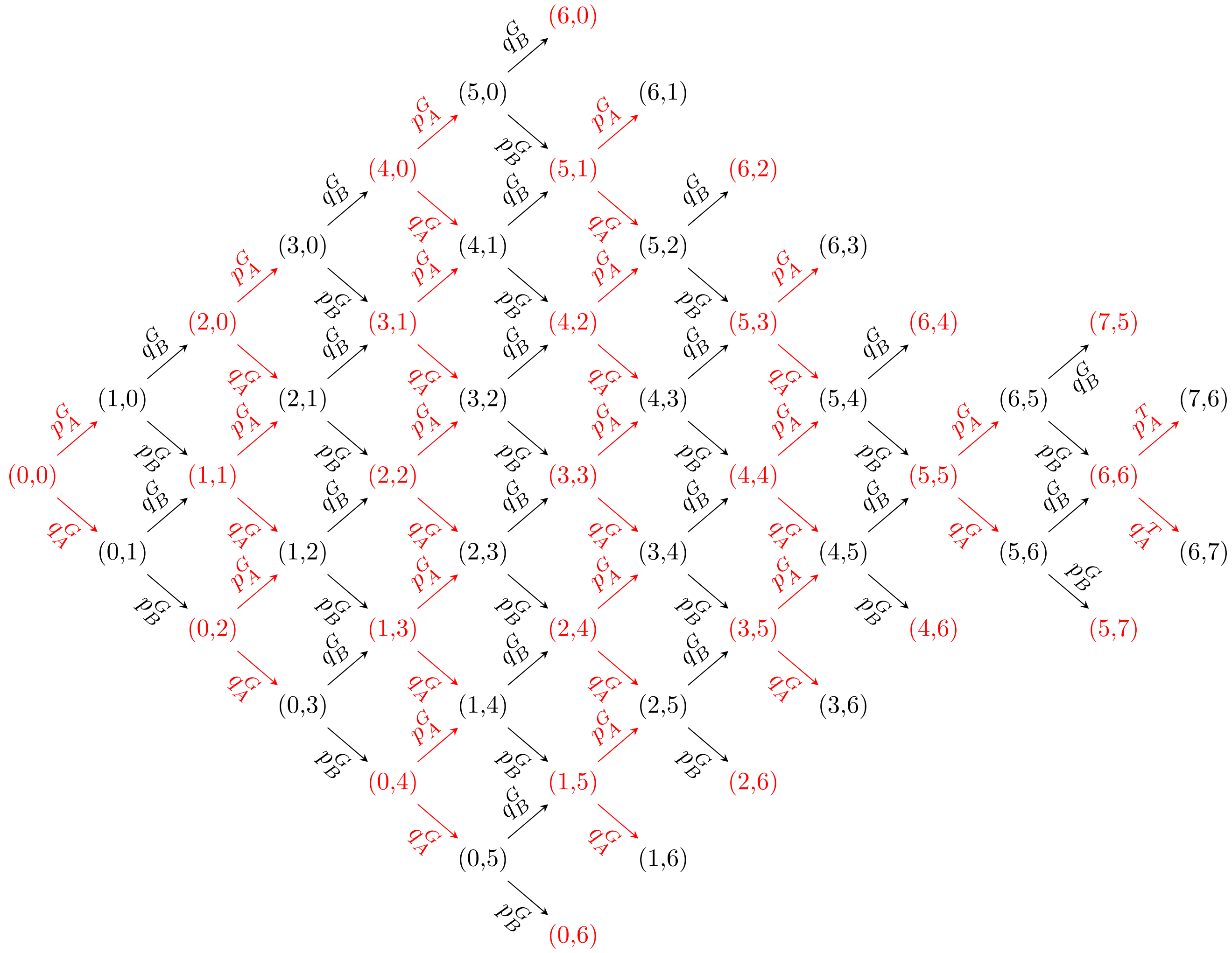
Figuur 2: De kans dat speler A de game wint, gegeven de kans dat speler A een punt wint

In figuur 2 is goed te zien dat de punten redelijk goed op de lijn liggen. Verder zijn er vier groepjes te zien. Het bovenste groepje is de data van de winnaars van de mannen, het groepje net daaronder van de winnaars bij de vrouwen. Het derde groepje zijn de mannelijke verliezers en het onderste groepje bevat de verliezers bij de vrouwen. Het valt dus op dat de mannen in alle gevallen een grotere kans hebben op een punt te pakken op hun eigen service dan de vrouwen, dit is te verklaren doordat mannen vaak een stuk harder serveren en zij daardoor hun tegenstander meteen onder druk zetten, waardoor het lastig is om het punt te pakken.

Door de vier grandslams met elkaar te vergelijken zie je eigenlijk steeds hetzelfde fenomeen. Het plusje dat de Wimbledon data aangeeft ligt steeds het hoogste. Dan volgen de US Open en Australian Open die steeds erg dicht bij elkaar liggen en daaronder is steeds het rondje van Roland Garros te zien.
Dit is te verklaren door de ondergrond waar de vier toernooien op gespeeld worden. Wimbledon wordt gespeeld op gras, wat een erg snelle ondergrond is. Hierdoor zijn de rally's vaak erg kort en is degene met een goede service vaak in het voordeel. De US en Australian open worden beide op hardcourt gespeeld, vandaar dat de data dus ook erg dicht bij elkaar ligt. Het laatste grandslam, Roland Garros, wordt gespeeld op gravel. Dit is de minst snelle ondergrond, waardoor de rally's vaak juist erg lang zijn. Hierdoor is het makkelijker voor een tegenstander om de service terug te brengen en de rally toch nog te winnen.

3 De kans op het winnen van een set

Nu de kans op het winnen van een game bekend is, kan ook de kans op het winnen van een set worden berekend. Een set is namelijk opgebouwd uit games. De twee spelers service steeds om de beurt, totdat een van de twee spelers 6 games heeft behaald. Echter geldt hier ook dat het verschil in games minstens twee moet zijn. Wanneer de stand in games (6-6) is, wordt er een tiebreak gespeeld om te bepalen wie de set wint.
In een set serveren beide spelers steeds om de beurt een game. Bij het berekenen van de kans om de set te winnen, moet er dus ook steeds meegenomen worden welke speler die desbetreffende game aan het serveren is. Over het algemeen geldt namelijk dat de kans op het winnen van een game die je zelf serveert een stuk hoger is dan een game waarin je niet zelf hoeft te serveren.



Figuur 3: Het verloop van een set. Met rood zijn de games aangegeven die speler A serveert, met zwart de games die speler B serveert.

In figuur 3 is te zien hoe een set verloopt die speler A begint met serveren. De kans voor speler A om een set te winnen die hij zelf begonnen is met serveren, wordt aangeduid met p_A^S.
Een set win je dus wanneer je 6 games weet te behalen. Ook hier geldt weer dat het verschil in games minimaal 2 moet zijn. Wanneer de stand 6-5 is, moet er dus sowieso nog een game gespeeld worden. Wanneer de speler die voorstond deze game wint, wint diegene de set met 7-5. Wanneer de andere speler de game weet te behalen, zal de stand 6-6 worden. Dan wordt een tiebreak gespeeld. Dit is een speciale game waarin je 7 punten moet halen, ook weer met een verschil van minstens twee punten.
Bij het berekenen van de kans op het winnen van de set moet dus weer gebruik worden gemaakt van gevalsonderscheiding. $p_A^S(i,j)$ is de kans dat speler A i games heeft gehaald, terwijl speler B er j wist te verdienen, dit in de set die speler A is begonnen met serveren. p_A^T is de kans dat indien er een tiebreak wordt gespeeld, speler A deze weet te winnen, ook weer wanneer speler A de tiebreak begonnen is met serveren. Omdat er bij het aanvangen van een tiebreak altijd 12, en dus een even aantal games is gespeeld, zal degene die de set is begonnen met serveren, ook altijd degene zijn die de tiebreak mag beginnen met serveren.
De kans dat speler A de set wint, wanneer hij zelf begonnen is met serveren, wordt dus gegeven door:

$$p_A^S = \sum_{j=0}^{4} p_A^S(6,j) + p_A^S(7,5) + p_A^S(6,6)p_A^T \tag{7}$$

Allereerst wordt gekeken naar de kans om de set winnen in het geval dat speler B ten hoogste 4 games weet te halen. Dit is dus de sommatie bij (7). Dit is lastiger dan het bereken van de kans om een game te winnen, omdat er hier rekening moet worden gehouden met welke speler serveert.
De regel bij tennis is dat de spelers om de beurt één game serveren. Er wordt dus gekenen naar de kans dat speler A wint wanneer hij begint met serveren. Het serveer schema zal dus $ABABABAB$ zijn; A serveert wanneer de som van de tot dan toe gespeelde games even is, speler B bij serveert wanneer die som oneven is. Bij de even games is de kans dat speler A de game wint dus p_A^G, terwijl deze kans bij de oneven games q_B^G is. Om de kans $p_A^G(i,j)$ te berekenen moet er dus steeds gekeken worden naar de stand van de voorafgaande game. Er wordt dus gebruik gemaakt van achterwaartse recursie. Hier moeten dus twee gevallen worden onderscheiden.
Voor $0 \leq i, j \leq 6$.

- $i - 1 + j$ is even
  In dat geval gaat speler A serveren en is de kans dat speler A die game wint p_A^G. De kans dat speler B de game wint is q_A^G. Er zijn twee manieren om een stand van (i,j) te bereiken.
  Of de stand voorafgaand van de game was $(i-1,j)$ en speler A wint de game: $p_A^S(i-1,j)p_A^G$.
  Of de stand was $(i,j-1)$ en speler B wint de game: $p_A^S(i,j-1)q_A^S$.
  Dus geldt:

$$\begin{aligned} p_A^S(i,j) &= p_A^S(i-1,j)p_A^G + p_A^G(i,j-1)q_A^G \\ &\textbf{laat } i-1 \textbf{ term weg als } j=6, i \leq 5 \\ &\textbf{laat } j-1 \textbf{ term weg als } i=6, j \leq 5 \end{aligned} \tag{8}$$

- $i-1+j$ oneven
  In dit geval gaat speler B serveren en is de kans dat speler A die game wint q_B^G. De kans dat speler B de game wint is p_B^G. Ook hier zijn weer twee mogelijkheden om een stand van (i,j) te bereiken. Of de stand voorafgaand aan de game was $(i-1,j)$ en speler A wint de game: $p_A^S(i-1,j)q_B^G$.
  Of de stand was $(i,j-1)$ en speler B wint de game: $p_A^S(i,j-1)p_B^S$.
  Dus geldt:
  $$p_A^S(i,j) = p_A^S(i-1,j)q_B^G + p_A^G(i,j-1)p_B^G \tag{9}$$

De begincondities bij deze recursies zijn:

$$p_A^S(0,0) = 1; \qquad p_A^S(i,j) = 0 \text{ als } i < 0, \text{ of } j < 0. \tag{10}$$

Het geval van een set die in 7-5 eindigt moet dus apart bekeken worden. Bij een stand van 6-5 gaat speler B serveren. De kans voor speler A om die game te winnen is dus q_B^G, terwijl die kans voor speler B p_B^G is.

$$p_A^S(7,5) = p_A^S(6,5)q_B^G; \qquad p_A^S(5,7) = p_A^S(5,6)p_B^G \tag{11}$$

In de Appendix is de uitwerking voor (8)-(11) te vinden.

Met het behulp van deze setstanden, kunnen de rest van de mogelijke setstanden berekend worden. Immers, door $p_A^G \leftrightarrow q_A^G$ en $p_B^G \leftrightarrow q_B^G$ om te draaien in $p_A^S(i,j)$, krijg je $p_A^S(j,i)$. De kans op een tiebreak kun je uitdrukken in "1-'de kans op geen tiebreak'". Dus:

$$p_A^S(6,6) = 1 - \big[\sum_{i=0}^{4}(p_A^S(i,6) + p_A^S(6,i)) + p_A^S(7,5) + p_A^S(5,7)\big] \tag{12}$$

3.1 Kans op het winnen van een tiebreak

Een tiebreak is een speciale game om ervoor te zorgen dat sets niet eindeloos doorgaan. Degene die het eerste zeven punten scoort, die wint de tiebreak. Echter, ook hier geldt dat dat met minstens twee punten verschil moet zijn. In een tiebreak serveert eerst de ene speler eenmaal, daarna serveren de spelers om de beurt steeds twee keer. Omdat de tiebreak bij een 6-6 stand plaatsvindt, is de speler die de set is begonnen met serveren, ook degene die de tiebreak mag beginnen met serveren. In dit geval zal het dus speler A zijn. De serveer volgorde zal dus $A, BB, AA, BB, AA, ..$ zijn.
De tiebreak kan dus weer worden opgedeeld in twee sommaties:

$$p_A^T = \sum_{j=0}^{5} p_A^T(7,j) + p_A^T(6,6)\sum_{n=0}^{\infty} p_A^T(n+2,n) \tag{13}$$

Het deel achter de eerste sommatie geeft de mogelijkheid weer dat de tiebreak is afgelopen wanneer speler A 7 punten heeft behaald en speler B er maximaal vijf heeft behaald. Het tweede deel geeft de mogelijkheid weer dat de stand van 6-6 is bereikt. In dit geval gaan de spelers net zo lang door totdat een van de twee, twee punten achter elkaar heeft gescoord. Omdat beide spelers serveren in de tiebreak moet er dus meegenomen worden welke speler er serveert. Daarom worden ook hier terugwaartse recursieve formules gebruikt. Ook hier moet namelijk steeds gekeken worden naar het voorafgaande punt.
Ook hier is weer sprake van gevalsonderscheiding:

- als $i-1+j = 0, 3, 4, ..., 4n-1, 4n, ..$ is speler A aan de beurt om te serveren:
  Voor $i \leq i, j \leq 7$:

$$\begin{aligned} p_A^T(i,j) &= p_A^T(i-1,j)p_A^R + p_A^T(i,j-1)q_A^R \\ &\textbf{laat } i-1 \textbf{ weg als } i=7, j \leq 6 \\ &\textbf{laat } j-1 \textbf{ weg als } j=7, i \leq 6 \end{aligned} \tag{14}$$

- als $i-1+j = 1, 2, 5, 6, .., 4n+1, 4n+2, ...$ gaan speler B serveren:
  Voor $i \leq i, j \leq 7$:

$$\begin{aligned} p_A^T(i,j) &= p_A^T(i-1,j)q_B^R + p_A^T(i,j-1)p_B^R \\ &\textbf{laat } i-1 \textbf{ weg als } i=7, j \leq 6 \\ &\textbf{laat } j-1 \textbf{ weg als } j=7, i \leq 6 \end{aligned} \tag{15}$$

Met begincondities:

$$p_A^T(0,0) = 1, \qquad p_A^T(i,j) = 0 \text{ als } i < 0, \text{ of } j < 0 \tag{16}$$

De uitwerkingen van (14)-(16) worden in de Appendix gegeven.

Ook hier geldt dat als $p_A^T(j,i)$ gegeven wordt door $p_A^R \leftrightarrow q_A^R$ en $p_B^R \leftrightarrow q_B^R$ om te draaien in $p_A^T(i,j)$. Dan geldt ook:

$$p_A^T(6,6) = 1 - \Big[\sum_{i=0}^{5}(p_A^T(i,7) + p_A^T(7,i))\Big] \tag{17}$$

Wanneer de stand van 6-6 optreedt in een tiebreak, gaan de spelers net zo lang door, totdat een van de twee spelers twee punten meer heeft dan de ander.
Bij 6-6 is speler nog eenmaal aan de beurt om te serveren, daarna speler B twee keer, dan speler A twee keer, etc. Oftewel $AB, BA, AB, BA, AB, ...$ Dus eigenlijk kunnen we de tiebreak opdelen in steeds twee punten waarin ze beide een keer serveren. De ene keer serveert A als eerste, daarna B.
Om de tiebreak en daarmee de set, te winnen, is het voor speler A dus nodig om (van zo'n setje van twee) een keer zijn eigen service te houden én om een keer de service van speler B te pakken. De kans hierop is $p_A^R q_B^R$. Zolang er geen winnaar is, zijn er twee mogelijkheden. Of ze pakken beide hun eigen service, dit gebeurt met kans $p_A^R p_B^R$, of ze pakken allebei de service van de ander, met kans $q_A^R q_B^R$. Van de n setjes van twee zullen er dus j setjes zijn waarbij ze beide hun eigen service houden en $n-j$ waarbij ze beide de service van de ander pakken. Waarbij $0 \leq j \leq n$ Dit leidt tot de volgende vergelijking:

$$\begin{aligned} p_A^T(n+2,n) &= \sum_{j=0}^{n}(p_A^R p_B^R)^j (q_A^R q_B^R)^{n-j}\binom{n}{j} p_A^R q_B^R \\ &= p_A^R q_B^R \sum_{j=0}^{n}(p_A^R p_B^R)^j (q_A^R q_B^R)^{n-j}\binom{n}{j} \\ &= (p_A^R p_B^R + q_A^R q_B^R)^n p_A^R q_B^R \end{aligned} \tag{18}$$

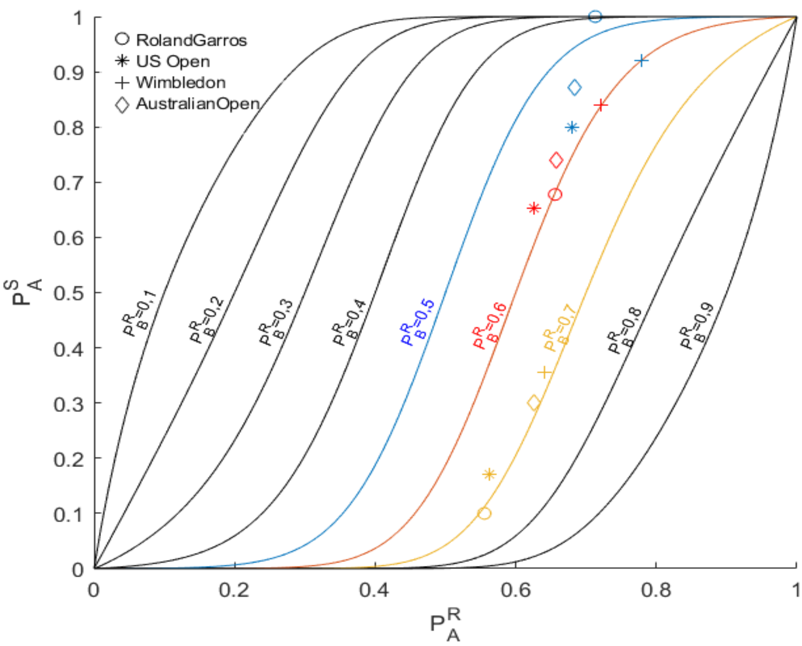
Door (14)-(18) in te vullen in (13) kan p_A^T berekend worden. Door deze, samen met (8)-(12) dan weer in te vullen in (7) wordt de kans p_A^S gegeven.

Ook bij deze formule wordt weer gekeken of dit overeenkomt met de werkelijkheid. Hiervoor is weer data gebruikt van de vier grandslams.
De grafieken zijn geplot voor bepaalde waarden van p_B^R. Aangezien het professionele tenniswedstrijden betreft, waar alleen de 128 allerbeste tennissers aan mee mogen doen, komt het vrijwel nooit voor dat er waarden van $p_B^R < 0.4$ of $p_B^R > 0.8$ in een wedstrijd plaatsvinden. De focus is dus gelegd op waarde rond de 0.5, 0.6 en 0.7.
Bij iedere grandslam is gezocht naar wedstrijden waarbij de kans voor een van de twee spelers om een punt te maken, tussen de $p_B^R = 0.50 \pm 0.01$, $p_B^R = 0.60 \pm 0.01$ of $p_B^R = 0.70 \pm 0.01$ lagen. In tabel 1 is te zien hoeveel wedstrijden dit er bij de mannen per toernooi waren.

Waarden p_B^R	Roland Garros	US Open	Wimbledon	Australian Open
0.50 ± 0.01	6	11	5	7
0.60 ± 0.01	27	18	20	19
0.70 ± 0.01	10	17	21	15

Tabel 1: Aantal wedstrijden per grandslam bij de mannen waarbij $p_B^R = 0.50 \pm 0.01$, $p_B^R = 0.60 \pm 0.01$ of $p_B^R = 0.70 \pm 0.01$

Voor al deze wedstrijden is voor de tegenstander, speler A, berekend wat de kans is om een punt en om een game te winnen, respectievelijk p_A^R en p_A^S. Dit laatste is gedaan door het aantal gewonnen sets door speler A te delen door het totaal aantal gespeelde sets.
Daarna is per toernooi en per waarde van p_B^R het gemiddelde voor p_A^R en p_A^S berekend. Dit is geplot in figuur 4, tezamen met de theoretische kans om een set te winnen, gegeven een aantal verschillende waarden van p_B^R.



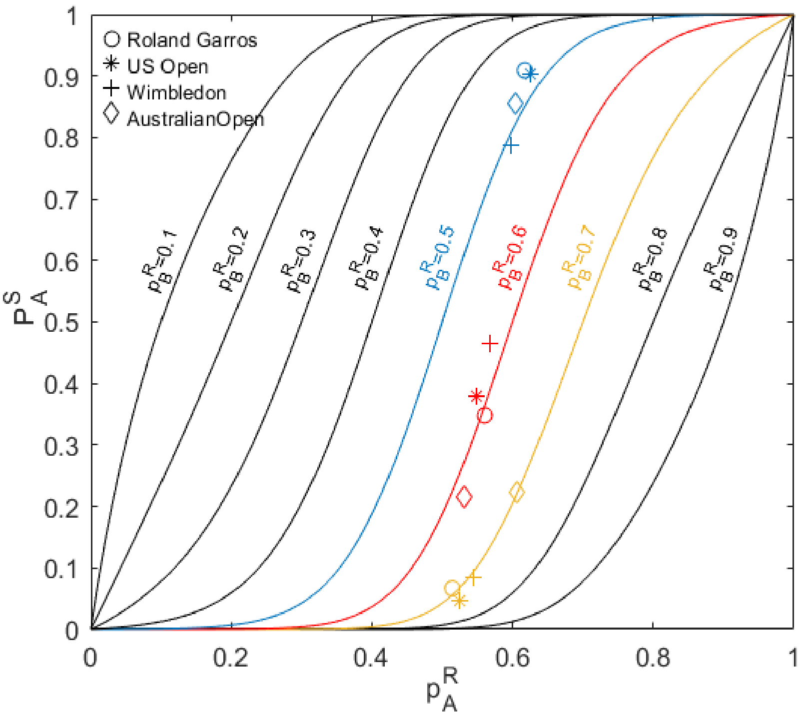
Figuur 4: de kans p_A^S gegeven p_A^R voor verschillende waarden van p_B^R. Data van wedstrijden van de mannen waarbij $p_B^R = 0.50 \pm 0.01$ (blauw), $p_B^R = 0.60 \pm 0.01$ (rood), $p_B^R = 0.70 \pm 0.01$ (oranje)

Wat opvalt in figuur 4 is dat de punten waarbij $p_B^R = 0.60 \pm 0.01$ en $p_B^R = 0.70 \pm 0.01$ erg goed op de theoretisch voorspelde lijn liggen. Dit is niet het geval voor $p_B^R = 0.50 \pm 0.01$. Dit laatste kan wellicht verklaard worden door het beperkt aantal wedstrijden waarbij dit het geval is, zoals te zien in tabel 1. Bij ieder toernooi is het aantal wedstrijden waarbij $p_B^R = 0.50 \pm 0.01$ relatief erg laag. Dit kan dus van invloed zijn op de kansen. Vandaar dat de voorspelling niet helemaal overeenkomt.

Dit alles is ook voor de vrouwen gedaan, wat te zien is in tabel 2 en figuur 5.

Waarden p^R	Roland Garros	US Open	Wimbledon	Australian Open
0.50 ± 0.01	22	17	22	16
0.60 ± 0.01	23	22	23	17
0.70 ± 0.01	10	7	10	6

Tabel 2: Aantal wedstrijden per grandslam bij de mannen waarbij $p_B^R = 0.50 \pm 0.01$, $p_B^R = 0.60 \pm 0.01$ of $p_B^R = 0.70 \pm 0.01$



Figuur 5: de kans p_A^S gegeven p_A^R voor verschillende waarden van p_B^R. Data van wedstrijden van de vrouwen waarbij $p_B^R = 0.50 \pm 0.01$ (blauw), $p_B^R = 0.60 \pm 0.01$ (rood), $p_B^R = 0.70 \pm 0.01$ (oranje)

Bij de wedstrijden van de vrouwen, zoals te zien in figuur 5, liggen de punten allemaal erg dicht bij de lijn, zelfs als het aantal wedstrijden beperkt is. De voorspelde kansen komen hier dus erg goed overeen met de werkelijkheid.

3.2 Beginnen met serveren of retourneren

In deze paragraaf wordt bewezen dat het voor de kans om de set te winnen niet uitmaakt of je begint met serveren of niet. Wanneer het niet uitmaakt, moet de kans dat speler A de set wint wanneer speler A begint met serveren (p_A^S) gelijk zijn aan de kans dat speler A de set wint wanneer speler B de set begint met serveren (q_B^S).
De formule voor p_A^S is gegeven door

$$p_A^S = \sum_{j=0}^{4} p_A^S(6,j) + p_A^S(7,5) + p_A^S(6,6)p_A^T$$

zoals gegeven in (7).
De formule voor q_B^S wordt gegeven door

$$q_B^S = \sum_{j=0}^{4} p_B^S(6,j) + p_B^S(5,7) + p_B^S(6,6)q_B^T \tag{19}$$

$p_B^S(j,i)$ kan verkregen worden door $p_A^G \leftrightarrow q_B^G$ en $p_B^G \leftrightarrow q_A^G$. Immers, een voorsprong van $(6,1)$ voor speler A is hetzelfde als een achterstand van $(6,1)$ voor speler B. Het enige waar rekening mee gehouden moet worden is welke speler er serveert. Beide kansen p_A^G en q_B^G geven de kans aan dat speler A de game wint, alleen bij de eerste serveert A en bij de tweede serveert B.

Wanneer $p_A^G \to q_B^G$ en $p_B^G \to q_A^G$ in de formules (A.1) en (A.6) wordt omgedraaid, is het meteen duidelijk dat

$$p_A^S(6,0) = p_B^S(0,6) \tag{20}$$

$$p_A^S(7,5) = p_B^S(5,7) \tag{21}$$

Dit is ook logisch, immers bij een stand van $(6,0)$ in het voordeel van speler A wil dit zeggen dat speler A drie games heeft gewonnen op zijn eigen service en drie games heeft gewonnen op de service van speler B. Welke speler dan begonnen is met serveren is niet van belang.
Hetzelfde geldt bij een $(7,5)$ winst voor speler A. Hierbij hebben zowel speler A als speler B er van de eerste tien games er vijf gewonnen. En de laatste twee games heeft speler A allebei gewonnen. Van deze twee games hebben speler A en B er beide 1 geserveerd. Ook hier maakt het dus niet uit wie er is begonnen met serveren.

Verder is het ook duidelijk dat

$$p_A^S(6,6) = p_B^S(6,6) \tag{22}$$

Er moet dus alleen nog aangetoond worden dat

$$\sum_{j=1}^{4} p_A^S(6,j) = \sum_{j=1}^{4} p_B^S(j,6) \tag{23}$$

en dat

$$p_A^T = q_B^T \tag{24}$$

Om (23) te bewijzen, wordt de sommatie opgesplitst en wordt bewezen dat

$$p_A^S(6,1) + p_A^S(6,2) = p_B^S(1,6) + p_B^S(2,6) \tag{25}$$

en

$$p_A^S(6,3) + p_A^S(6,4) = p_B^S(3,6) + p_B^S(4,6) \tag{26}$$

Hiervoor worden de formules (A.1)-(A.6) uit de Appendix gebruikt en wordt $q_A^G = 1 - p_A^G$ en $q_B^G = 1 - p_B^G$ gesubstitueerd. Verder worden ook hier p_A^G en p_B^G door respectievelijk q_B^G en q_A^G vervangen zodat $p_B^S(j,i)$ uit $p_A^S(i,j)$ verkregen kan worden. Door dit helemaal uit te werken kunnen de formules alsvolgt geschreven worden.

$$p_A^S(6,2n-1) + p_A^S(6,2n) = \sum_{i=0}^{6+2n} \sum_{j=0}^{6+2n} a_{ij}^S(n)(p_A^G)^i(p_B^G)^j \tag{27}$$

$$p_B^S(6,2n-1) + p_B^S(6,2n) = \sum_{i=0}^{6+2n} \sum_{j=0}^{6+2n} b_{ij}^S(n)(p_A^G)^i(p_B^G)^j \tag{28}$$

voor $n = 1, 2$. Door dit helemaal uit te werken blijkt dat $a_{ij}^S(n) = b_{ij}^S(n)$. Deze coëfficienten zijn gegeven in de Appendix, (A.13)-(A.14).
Nu moet er dus alleen nog bewezen worden dat $p_A^T = q_B^T$. Dit wordt op ongeveer dezelfde wijze gedaan. De formule van p_A^T wordt gegeven in (13) en

$$q_B^T = \sum_{j=0}^{5} p_B^T(j,7) + p_B^T(6,6) \sum_{n=0}^{\infty} p_B^T(n,n+2) \tag{29}$$

Ook hier wordt $p_B^T(j,i)$ uit (29) verkregen door $p_A^R \leftrightarrow q_B^R$ en $p_B^R \leftrightarrow q_A^R$ om te draaien in (A.7)-(A.12) & (17)-(18). Uit deze laatste twee volgt dan meteen dat $p_A^T(6,6) = p_B^T(6,6)$ en $p_A^T(n+2,n) = p_B^T(n,n+2)$. Rest alleen nog te bewijzen dat

$$\sum_{j=0}^{5} p_A^T(7,j) = \sum_{j=0}^{5} p_B^T(j,7) \tag{30}$$

Ook deze vergelijking wordt weer in stukjes verdeeld en per stuk bewezen.

$$p_A^T(7,0) + p_A^T(7,1) = p_B^T(0,7) + p_B^T(1,7) \tag{31}$$

$$p_A^T(7,2) + p_A^T(7,3) = p_B^T(2,7) + p_B^T(3,7) \tag{32}$$

$$p_A^T(7,4) + p_A^T(7,5) = p_B^T(4,7) + p_B^T(5,7) \tag{33}$$

Hiervoor worden (A.7)-(A.12) gebruikt en worden $q_A^R = 1 - p_A^R$ en $q_B^R = 1 - p_B^R$. Hieruit volgt weer dat

$$p_A^T(7,2n) + p_A^T(7,2n+1) = \sum_{i=0}^{4} \sum_{j=0}^{4} a_{ij}^T(n)(p_A^R)^i(p_B^R)^j \tag{34}$$

$$p_B^T(2n,7) + p_A^T(2n+1,7) = \sum_{i=0}^{4} \sum_{j=0}^{4} a_{ij}^T(n)(p_A^R)^i(p_B^R)^j \tag{35}$$

voor $n = 0, 1, 2$. Ook hier blijkt dan weer dat de coëfficienten $a_{ij}^T(n) = b_{ij}^T(n)$. Deze coëfficienten zijn weergegeven in de Appendix, (A.15)-(A.16)

4 Kans op het winnen van de wedstrijd

In dit hoofdstuk wordt p_A^W berekend, de kans dat speler A de wedstrijd wint, wanneer speler A de wedstrijd begint met serveren. Doordat in de vorige paragraaf, paragraaf 3.2 is bewezen dat het niet uitmaakt wie van de twee spelers begint met serveren, wordt het berekenen van deze kans een stuk gemakkelijker. Wanneer het namelijk uit zou maken welke speler zou beginnen met serveren, zou er steeds per set bekeken moeten worden wie er begint en wie er eindigt met serveren. Dit levert een groot aantal mogelijkheden op die allemaal meegenomen zouden moeten worden.
Aangezien het dus niet uitmaakt wie begonnen is met serveren, zijn dus alleen de kansen nodig dat speler A dan wel speler B de set wint. De kans dat speler A de set wint, wordt gegeven door (7), de kans dat speler B de set wint, is dan 1-'(7)'. Immers, ofwel speler A wint de set, ofwel speler B wint hem, dus moeten deze kansen samen 1 zijn.

Bij de mannen en vrouwen is er een verschil in het aantal sets dat gewonnen moet worden om de wedstrijd te winnen. Bij de grandslams geldt namelijk bij de mannen 'best of 5'. Oftewel, degene die het eerste drie sets heeft, wint. Bij de vrouwen geldt 'best of drie' en is de winnaar degene die als eerste twee sets heeft gewonnen. De kans om de wedstrijd te winnen, hangt dus ook af van het aantal sets dat gewonnen moet worden.

Er is niet alleen een verschil in mannen en vrouwen, er is ook een klein verschil in toernooien. Bij de US Open en alle andere toernooien die door het jaar heen worden gespeeld, wordt er in de laatste set ook een tiebreak gespeeld, terwijl er bij de andere drie grandslams in de laatste set net zolang wordt doorgespeeld, totdat iemand een verschil van twee games heeft. Hier is het dus mogelijk dat de laatste set in een gamestand van bijvoorbeeld (29-27) eindigt. In dit document wordt de vorm gebruikt waarbij ook in de laatste set sprake is van een tiebreak. Dit omdat dit de meestvoorkomende vorm is. Als vervolgonderzoek zou die andere mogelijkheid uitgebreid besproken kunnen worden.

4.1 Best of 3

Allereerst wordt de kans voor de vrouwen berekend, waarbij 'best of 3' van toepassing is. Deze kans wordt genoteerd met p_A^{WV}.
Voor de kans om een wedstrijd te winnen in dit geval zijn er eigenlijk maar twee mogelijkheden. Ofwel wint speler A met een stand van (2,0) in sets, of het wordt (2,1). In dit laatste geval moet gelden dat het (1,1) in setstanden is geweest, want anders is het na 2 sets al klaar. (1,1) kan het op twee manieren worden, ofwel wint eerst speler A een set en daarna speler B, of net andersom. De kans p_A^{WV} wordt dus gegeven door:

$$p_A^{WV} = (p_A^S)^2 + 2(p_A^S)^2 p_B^S \tag{36}$$

4.2 Best of 5

Van de mannen wordt er dus wat meer gevraagd en moeten er drie sets gewonnen worden. Deze kans is dan ook iets complexer en wordt genoteerd met p_A^{WM}
Een wedstrijd bij de mannen kan gewonnen worden met een stand van $(3,0)$, $(3,1)$ of $(3,2)$. Bij deze laatste twee standen moet gelden dat de laatste set sowieso door de uiteindelijke winnaar gewonnen moet zijn, anders had de laatste set niet meer gespeeld hoeven worden.

- De kans op $(3,0)$ wordt gegeven door $(p_A^S)^3$, alle drie de sets moeten naar speler A.
- De kans op $(3,1)$ wordt gegeven door $3(p_A^S)^2 p_B^S p_A^S$; van de eerste drie sets, gaan er twee naar speler A en eentje naar speler B, dit kan op $\binom{3}{2} = 3$ manieren.
- De kans op $(3,2)$ wordt gegeven door $6(p_A^S)^2 (p_B^S)^2 p_A^S$; van de eerste vier sets gaan er twee naar speler A en twee naar speler B en de laatste set gaan naar speler A.

De totale kans om een wedstrijd te winnen in het geval van 'best of 5' wordt dus gegeven door:

$$p_A^{WM} = 3(p_A^S)^3 + 3(p_A^S)^3 p_B^S + 6(p_A^S)^3 (p_B^S)^2 \tag{37}$$

5 Kans op het winnen van het toernooi

In dit hoofdstuk wordt de kans gegeven voor een speler om een toernooi te winnen. Hier wordt uit gegaan van een knock out systeem waaraan $128 = 2^7$ speler meedoen. Iedere speler wordt genummerd door $i = 1, ..., 128$. De nummering vindt zo plaats dat speler 1 in de eerste ronde tegen speler 2 speelt. Voor de verliezer is het toernooi ten einde, terwijl de winnaar uitkomt tegen de winnaar van speler 3/4. De verliezer is weer uitgeschakeld en de winnaar gaat het opnemen tegen de winnaar van 5/6/7/8, etc.
De kans dat speler i een wedstrijd van speler j wint, wordt gegeven door p_{ij}^W. Deze kans kan berekend worden aan de hand van (36) in het geval van 'best of 3' en (37) in het geval van 'best of 5'. Om de kans te bereken op het winnen van het toernooi, wordt de kolomvector $\mathbf{p}^{(n)} \in R^{1\times128}$ gegeven

$$\mathbf{p}^{(n)} = \begin{pmatrix} p_1^{(n)} \\ p_2^{(n)} \\ p_3^{(n)} \\ \vdots \\ p_{128}^{(n)} \end{pmatrix} \tag{38}$$

Hierbij is $p_i^{(n)}$ de conditionele kans dat speler i de wedstrijd in de $n$de ronde weet te winnen, gegeven de kans dat het deze ronde heeft weten te behalen (en dus in de (n-1)de ronde heeft gewonnen).
De kans om een bepaalde ronde te behalen wordt gegeven door een recursieve formule; de kans dat ronde n wordt behaald maal de kans dat de wedstrijd in de $n$de ronde gewonnen wordt gegeven het feit dat deze ronde is behaald. Ofwel

$$\mathbf{p}^{(0)} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \qquad \mathbf{p}^{(n)} = \mathbf{P}_n \mathbf{p}^{(n-1)} \qquad (n = 1, .., 6) \tag{39}$$

Hierbij is $\mathbf{P}_n$ een 128×128 blokdiagonale matrix, bestaande uit 2^{7-n} blokken. Dit aantal komt overeen met het aantal wedstrijden dat er gespeeld moet worden. In de $n$de ronde worden er namelijk 2^{7-n} wedstrijden gespeeld. Ieder blok staat dus ook voor één wedstrijd. Ieder blok en dus iedere wedstrijd wordt gelabeld door $\mathbf{P}_n^{(k)}$, met $1 \leq k \leq 2^{7-n}$. k geeft hierbij het nummer aan van de wedstrijd van die ronde.
Hoe verder het toernooi vordert, hoe lager het aantal wedstrijden. Immers, per ronde halveert het aantal deelnemers. In de eerste ronde zal de matrix $\mathbf{P}^{(n)}$ dus uit 2^7 wedstrijden bestaan, terwijl er in de vierde ronde nog maar 2^4 wedstrijden zijn. Dit is dan ook terug te zien in het aantal blokken. Per ronde halveert het aantal blokken waaruit de matrix is opgebouwd. De matrix ziet er dan ook alsvolgt uit.

$$
\mathbf{P}_n = \begin{bmatrix}
\mathbf{P}_n^{(1)} & 0 & 0 & \cdots & 0 \\
0 & \mathbf{P}_n^{(2)} & 0 & \cdots & 0 \\
0 & 0 & \mathbf{P}_n^{(3)} & \cdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \mathbf{P}_n^{(2^{7-n})}
\end{bmatrix} \tag{40}
$$

Hier is $\mathbf{P}_n^{(k)}$ een $2^n \times 2^n$ off-diagonale blokmatrix.
Iedere blok uit (40) representeert dus een wedstrijd. Voor iedere wedstrijd zijn er echter steeds meerdere spelers mogelijk. De eerste ronde staat voor iedereen vast, hier zal namelijk altijd speler $2k - 1$ tegen speler $2k$ spelen (voor $k = 1, .., 64$). De ronde die volgt is echter afhankelijk van wie de eerste ronde door is gekomen. Per ronde verdubbelt het aantal mogelijke tegenstanders. Voor speler 1 geldt bijvoorbeeld dat deze het in de eerste ronde opneemt tegen speler 2. In de tweede ronde moet hij het opnemen tegen de winnaar van speler3/speler 4. De derde ronde zal gespeeld worden tussen de winnaar van deze wedstrijd en de winnaar van 5/6/7/8. Aangezien de matrix waarden heeft voor alle mogelijke wedstrijden die gespeeld worden, worden de blokken van (40) dus iedere ronde verdubbelt. Verder geldt dat wanneer de matrix een waarden heeft voor $P_{i,j}$, deze ook een waarde moet hebben voor $P_{j,i}$. (Hierbij staat $P_{i,j}$ voor de kans dat speler i van speler j wint.)

Vandaar dat $\mathbf{P}_n^{(k)}$ ook weer is opgebouwd uit twee blokken. Het ene blok representeert de kansen dat speler α van speler β wint en het andere blok de kans dat speler β wint van α. Wat leidt tot de volgende matrix.

$$
\mathbf{P}_n^{(k)} = \begin{bmatrix}
0 & \mathbf{P}_{\alpha,\beta}^{(n,k)} \\
\mathbf{P}_{\beta,\alpha}^{(n,k)} & 0
\end{bmatrix} \tag{41}
$$

waarbij $\alpha = (k-1)2^n + 1$, $\beta = k2^n$.
De $\mathbf{P}_{\alpha,\beta}^{(n,k)}$ en $\mathbf{P}_{\beta,\alpha}^{(n,k)}$ zijn dan weer $2^{n-1} \times 2^{n-1}$ matrices van de vorm.

$$
\mathbf{P}_{\alpha,\beta}^{(n,k)} = \begin{bmatrix}
P_{\alpha,\beta+1-2^{n-1}} & \cdots & P_{\alpha,\beta-1} & P_{\alpha,\beta} \\
P_{\alpha+1,\beta+1-2^{n-1}} & \cdots & P_{\alpha+1,\beta-1} & P_{\alpha+1,\beta} \\
\vdots & \ddots & \vdots & \vdots \\
P_{\alpha+2^{n-1}-1,\beta+1-2^{n-1}} & \cdots & P_{\alpha+2^{n-1}-1,\beta-1} & P_{\alpha+2^{n-1}-1,\beta}
\end{bmatrix} \tag{42}
$$

$$
\mathbf{P}_{\beta,\alpha}^{(n,k)} = \begin{bmatrix}
P_{\beta+1-2^{n-1},\alpha} & \cdots & P_{\beta+1-2^{n-1},\alpha+2^{n-1}-2} & P_{\beta+1-2^{n-1},\alpha+2^{n-1}-1} \\
P_{\beta+2-2^{n-1},\alpha} & \cdots & P_{\beta+2-2^{n-1},\alpha+2^{n-1}-2} & P_{\beta+2-2^{n-1},\alpha+2^{n-1}-1} \\
\vdots & \ddots & \vdots & \vdots \\
P_{\beta,\alpha} & \cdots & P_{\beta,\alpha+2^{n-1}-2} & P_{\beta,\alpha+2^{n-1}-1}
\end{bmatrix} \tag{43}
$$

De waarden van de matrix, $P_{i,j}$, de kans dat speler i van speler j wint, kunnen berekend worden aan de hand van de formules (36) of (37), afhankelijk van de wedstrijdvorm.
De matrix (42) geeft alle mogelijkheden voor de $k$de wedstrijd in de $n$de ronde. In deze matrix is voor iedere wedstrijd $P_{i,j}$ gegeven. In matrix (43) wordt voor iedere wedstrijd de waarden $P_{j,i}$ gegeven, dus de kans dat de andere speler wint.

Door (41)-(44) in te vullen in (40) kan voor iedere speler berekend worden wat de kans is om de wedstrijd in ronde n te winnen, gegeven dat ronde n is bereikt. De uiteindelijke kans om het toernooi te winnen is dan het product van al deze conditionele kansen. Dit leidt tot de volgende vector $\mathbf{p}^{TW}$:

$$\mathbf{p}^{TW} \equiv \begin{pmatrix} p_1^{TW} \\ p_2^{TW} \\ p_3^{TW} \\ \vdots \\ p_{128}^{TW} \end{pmatrix} = \begin{pmatrix} \prod_{n=1}^{7} p_1^{(n)} \\ \prod_{n=1}^{7} p_2^{(n)} \\ \prod_{n=1}^{7} p_3^{(n)} \\ \vdots \\ \prod_{n=1}^{7} p_{128}^{(n)} \end{pmatrix} \tag{44}$$

Samenvattend, de kans om een toernooi te winnen kan dus berekend worden aan de hand van de recursieve formule gegeven bij (39). Hierbij wordt de kans om de $n$de ronde te winnen, gegeven het feit dat de $n$de ronde is bereikt, gegeven door $\mathbf{P}_n$. Dit is een 128×128 matrix opgebouwd uit 2^{7-n} blokken. Hierin staat ieder blok voor één specifieke wedstrijd. Deze matrix is gegeven in (40).
Voor iedere wedstrijd zijn er steeds meerdere mogelijke spelers. Hoe verder in het toernooi, hoe meer mogelijke spelers er zijn. Ieder blok uit (40) is weer opgebouwd uit kleinere blokken. Deze blokken geven alle mogelijkheden combinaties van spelers voor die wedstrijd weer. Met behulp van (42) en (43) worden al deze mogelijkheden voor een wedstrijd weergegeven.
Uiteindelijk kan dan voor iedere speler i de kans om het toernooi te winnen, p_i^{TW}, gegeven worden. Dit is namelijk het product van alle conditionele kansen om een bepaalde ronde te behalen, zoals gegeven in (44).

Om de matrices en de werking daarvan iets duidelijker te maken, zullen de gevallen waarbij$n = 1$ en $n = 2$ helemaal uitgewerkt worden.

Allereerst $n = 1$, dan

$$\mathbf{P_1} = \begin{bmatrix} P_1^{(1)} & 0 & 0 & \cdots & 0 \\ 0 & P_1^{(2)} & 0 & \cdots & 0 \\ 0 & 0 & P_1^{(3)} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & P_1^{(64)} \end{bmatrix} \tag{45}$$

waarbij $\mathbf{P}_1^{(k)}$ een 2×2 matrix is, namelijk

$$\mathbf{P}_1^{(k)} = \begin{bmatrix} 0 & P_{2k-1,2k} \\ P_{2k,2k-1} & 0 \end{bmatrix} \tag{46}$$

Expliciet:

$$\mathbf{P}_1^{(1)} = \begin{bmatrix} 0 & P_{1,2} \\ P_{2,1} & 0 \end{bmatrix}, \mathbf{P}_1^{(2)} = \begin{bmatrix} 0 & P_{3,4} \\ P_{4,3} & 0 \end{bmatrix}, ..., \mathbf{P}_1^{(64)} = \begin{bmatrix} 0 & P_{127,128} \\ P_{128,127} & 0 \end{bmatrix} \tag{47}$$

In de eerste ronde zijn de enige wedstrijden die plaatsvinden namelijk: speler $2k$ tegen speler $2k - 1$ (voor $k = 1, .., 64$, er zijn immers 64 wedstrijden in de eerste ronde).

In de tweede ronde zullen de winnaars van de eerste ronde tegen elkaar spelen. Zo zal de winnaar van speler 1/2 tegen de winnaar 3/4 spelen. Er zijn 32 wedstrijden en voor iedere wedstrijd zijn vier mogelijkheden, voor $k = 0, .., 32$:

- speler $4k - 3$ tegen speler $4k - 1$
- speler $4k - 3$ tegen speler $4k$
- speler $4k - 2$ tegen speler $4k - 1$
- speler $4k - 2$ tegen speler $4k$

Er zijn dus 32 wedstrijden wat leidt tot 32 blokken. Elk van deze blokken is weer opgebouwd uit twee blokken van 2×2. Dit omdat er voor iedere speler twee mogelijke tegenstanders zijn in deze ronde.
Dus, voor $n = 2$:

$$\mathbf{P_2} = \begin{bmatrix} P_2^{(1)} & 0 & 0 & \cdots & 0 \\ 0 & P_2^{(2)} & 0 & \cdots & 0 \\ 0 & 0 & P_2^{(3)} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & P_2^{(32)} \end{bmatrix} \tag{48}$$

Waarbij

$$\mathbf{P}_2^{(k)} = \begin{bmatrix} 0 & 0 & P_{4k-3,4k-1} & P_{4k-3,4k} \\ 0 & 0 & P_{4k-2,4k-1} & P_{4k-2,4k} \\ P_{4k-1,4k-3} & P_{4k-1,4k-2} & 0 & 0 \\ P_{4k,4k-3} & P_{4k,4k-2} & 0 & 0 \end{bmatrix} \tag{49}$$

Oftewel,

$$\mathbf{P}_2^{(1)} = \begin{bmatrix} 0 & 0 & P_{1,3} & P_{1,4} \\ 0 & 0 & P_{2,3} & P_{2,4} \\ P_{3,1} & P_{4,1} & 0 & 0 \\ P_{4,1} & P_{3,2} & 0 & 0 \end{bmatrix}, \quad \mathbf{P}_2^{(2)} = \begin{bmatrix} 0 & 0 & P_{5,7} & P_{5,8} \\ 0 & 0 & P_{6,7} & P_{6,8} \\ P_{7,5} & P_{7,6} & 0 & 0 \\ P_{8,5} & P_{8,6} & 0 & 0 \end{bmatrix}, ...,$$

$$\mathbf{P}_2^{(32)} = \begin{bmatrix} 0 & 0 & P_{125,127} & P_{125,128} \\ 0 & 0 & P_{126,127} & P_{126,128} \\ P_{127,125} & P_{127,126} & 0 & 0 \\ P_{128,125} & P_{128,126} & 0 & 0 \end{bmatrix} \tag{50}$$

5.1 Wat is de kans op het winnen van het toernooi, wanneer de halve finale al is bereikt.

In deze paragraaf wordt er berekend wat de kans is voor alle halve finalisten om het toernooi te winnen. Aangezien het voor de halve finalisten wordt gedaan, wordt er dus vanuit gegaan dat ze de halve finale al hebben bereikt. De spelers worden hier weer genummerd van 1 t/m 4.[2]
Dit betekent dus dat de vectoren met de kans om de kwartfinale, de halve finale en de finale te winnen, gegeven wordt door respectievelijk $\mathbf{p}^{(0)}$, $\mathbf{p}^{(1)}$ en $\mathbf{p}^{(2)}$, waarbij:

$$\mathbf{p}^{(0)} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{p}^{(n)} = \begin{pmatrix} p_1^{(n)} \\ p_2^{(n)} \\ p_3^{(n)} \\ p_4^{(n)} \end{pmatrix} \text{ voor } n = 1, 2 \tag{51}$$

De matrices $\mathbf{P}_1$ en $\mathbf{P}_2$ worden gegeven door

$$\mathbf{P}_1 = \begin{bmatrix} 0 & P_{12} & 0 & 0 \\ P_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & P_{34} \\ 0 & 0 & P_{43} & 0 \end{bmatrix} \tag{52}$$

$$\mathbf{P}_2 = \begin{bmatrix} 0 & 0 & P_{13} & P_{14} \\ 0 & 0 & P_{23} & P_{24} \\ P_{31} & P_{32} & 0 & 0 \\ P_{41} & P_{42} & 0 & 0 \end{bmatrix} \tag{53}$$

De kans dat speler i de halve finale wint en dus de finale haalt, is de $i$de component van

$$\mathbf{p}^{(1)} = \mathbf{P_1 p^{(0)}} = \begin{bmatrix} 0 & P_{12} & 0 & 0 \\ P_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & P_{34} \\ 0 & 0 & P_{43} & 0 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} P_{12} \\ P_{21} \\ P_{34} \\ P_{43} \end{pmatrix} \tag{54}$$

De kans dat speler i de finale wint, wanneer hij die bereikt, wordt gegeven door:

$$\mathbf{p}^{(2)} = \mathbf{P_2 p^{(1)}} = \mathbf{P}_2 = \begin{bmatrix} 0 & 0 & P_{13} & P_{14} \\ 0 & 0 & P_{23} & P_{24} \\ P_{31} & P_{32} & 0 & 0 \\ P_{41} & P_{42} & 0 & 0 \end{bmatrix} \begin{pmatrix} P_{12} \\ P_{21} \\ P_{34} \\ P_{43} \end{pmatrix} = \begin{pmatrix} P_{13}P_{34} + P_{14}P_{43} \\ P_{23}P_{34} + P_{24}P_{43} \\ P_{31}P_{12} + P_{32}P_{12} \\ P_{41}P_{12} + P_{42}P_{21} \end{pmatrix} \tag{55}$$

[2]Dit slaat echter niet op dezelfde nummers 1 t/m 4 als in de vorige paragraaf. De 1 is hier de eerste halve finalist, de 2 is voor de tweede halve finalist etc.

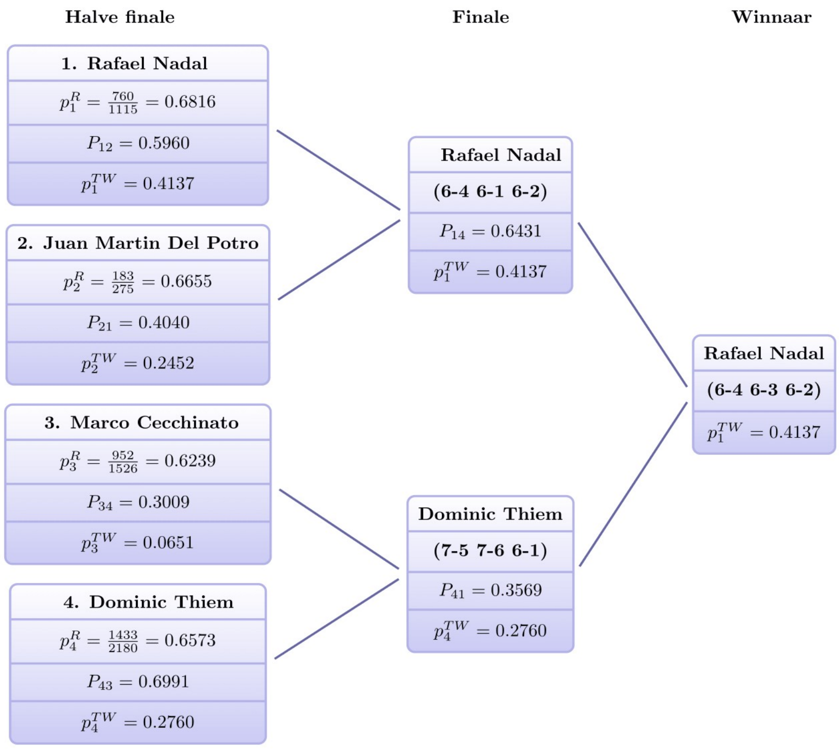
De totale kans voor speler i om het toernooi te winnen wanneer de halve finale al is bereikt, wordt gegeven door de i de component van de vector:

$$\mathbf{p}^{TW} = \begin{pmatrix} P_{12}(P_{13}P_{34} + P_{14}P_{43}) \\ P_{21}(P_{23}P_{34} + P_{24}P_{43}) \\ P_{34}(P_{31}P_{12} + P_{32}P_{21}) \\ P_{43}(P_{41}P_{12} + P_{42}P_{21}) \end{pmatrix} \tag{56}$$

Immers, zo geldt bijvoorbeeld voor speler 1, dat allereerst de halve finale gewonnen moet worden. De kans hierop is P_{12}. Daarnaast geldt dat ook de finale gewonnen moet worden. Wanneer de finale is bereikt, zijn er twee mogelijke tegenstanders voor speler 1, namelijk speler 3 of speler 4. In het geval dat speler 3 de finale haalt, moet hij dus van speler 4 hebben gewonnen, dit gebeurt met kans P_{34}. De kans dat speler 1 wint van speler 3, is P_{13}. Om de kans te bereken gegeven dat speler 1 in de finale tegen speler 3 moet en hij weet te winnen, is dus $P_{34}P_{13}$. De eerste twee kansen geven de mogelijkheid aan dat de finale speler 1 tegen speler 3 wordt. De laatste kans is die dat speler 1 van speler 3 wint.
De andere mogelijkheid was dat speler 4 de finale zou halen en speler 1 deze zou winnen. De kans dat speler 4 de finale haalt, is P_{43}, de kans dat speler 1 daarin wint, is P_{14}.
De uiteindelijk kans is dus $P_{12}(P_{13}P_{34} + P_{14}P_{43})$; speler 1 moet de finale halen én hij moet hierin winnen van óf speler 3, óf speler 4.

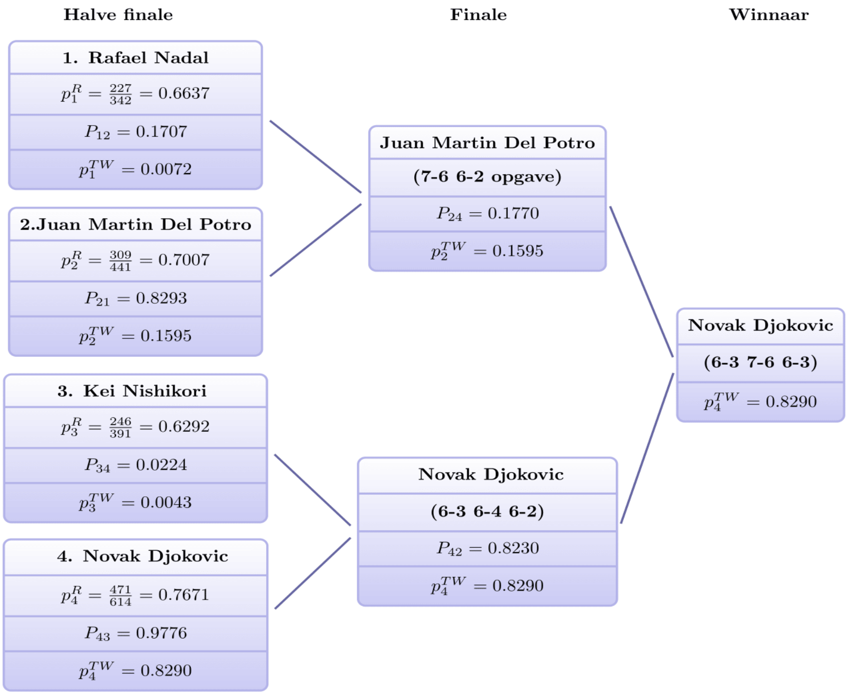
6 Halve finales Roland Garros en US Open

In dit hoofdstuk wordt gekeken of er met de voorafgaande hoofdstukken een voorspelling gedaan kan worden wie de winnaar van het toernooi wordt, wanneer de halve finalisten al bekend zijn. Er wordt dus steeds berekend wat de kans is voor ieder van de vier halve finalisten om uiteindelijk het toernooi te winnen.
Echter, om zo'n voorspelling te kunnen doen, moet voor iedere speler i de kans p_i^R bekend zijn. Dit is dus de kans om een punt op eigen service te winnen. Voorafgaand aan ieder grand slam worden door iedere speler een aantal toernooien gespeeld. Dit om in vorm te komen voor het toernooi en gewend te raken aan de ondergrond. De toernooien voorafgaand aan de grand slams, de zogenaamde ATP (mannen), of WTA (vrouwen) toernooien, worden altijd op dezelfde ondergrond gespeeld als het eerstvolgende grandslam.
De prestaties van deze toernooien kunnen dus een goed beeld geven van de vorm van de spelers in aanloop naar de grand slams. Daarnaast geldt vaak dat de top spelers redelijk gelijkwaardige tegenstanders hebben. Daarom is er voor gekozen deze toernooien te gebruiken om een voorspelling te maken van p_i^R voor iedere speler.
De data van de toernooien voorafgaand aan de grandslams wordt dus gebruikt. Aan de hand van alle wedstrijden die in deze toernooien zijn gespeeld wordt berekend wat de p_i^R is voor iedere speler. Dit wordt gedaan door het totaal aantal gewonnen punten op eigen service in al deze wedstrijden bij elkaar op te tellen en te delen door het totale aantal gespeelde punten op eigen service in deze wedstrijden.
Newton en Keller gebruiken in hun artikel [1] de data tot aan de halve finale van de grand slams zelf. In dit document is daar echter niet voor gekozen omdat de tegenstanders van iedere speler in dat geval verschillend zijn. Immers, afhankelijk van de plaatsing van de spelers hebben ze een makkelijke, of moeilijke loting. Dit geeft dus een minder goed beeld van de kwaliteiten van de speler dan wanneer de voorafgaande toernooien worden meegenomen.
Wanneer voor iedere speler i de kans p_i^R bekend is, kan met behulp van (36) of (37) berekend worden wat de kans is om de wedstrijd te winnen, P_{ij}. Dan kan met (56) berekend worden wat de kans is om het toernooi te winnen, p_i^{TW}. Dit is gedaan voor zowel Roland Garros als US Open, voor de mannen en voor de vrouwen. Dit levert de volgende resultaten op:



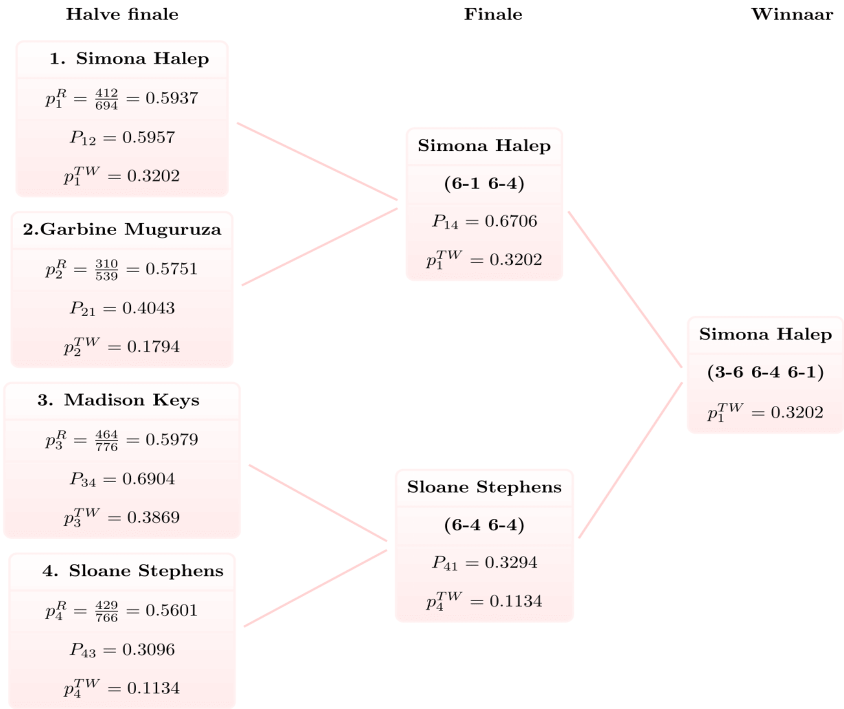
Figuur 6: Roland Garros mannen

In figuur 6 zijn de halve finalisten van het mannen toernooi van Roland Garros te zien. Rafael Nadal blijkt de grootste kans te hebben om het toernooi te winnen, immers $p_1^{TW} = 0.4137$. Nadal blijkt ook de winnaar te zijn van het toernooi, dus deze voorspelling komt uit. Verder geldt eigenlijk voor alle kansen dat ze overeenkomen met de werkelijke winnaars. Zo zal volgens de voorspelling de finale gespeeld worden tussen Rafael Nadal en Dominic Thiem, wat ook daadwerkelijk is gebeurd. De kans voor Nadal om dan de finale te winnen bleek 0.6431 te zijn. Hij heeft de finale ook redelijk makkelijk gewonnen.



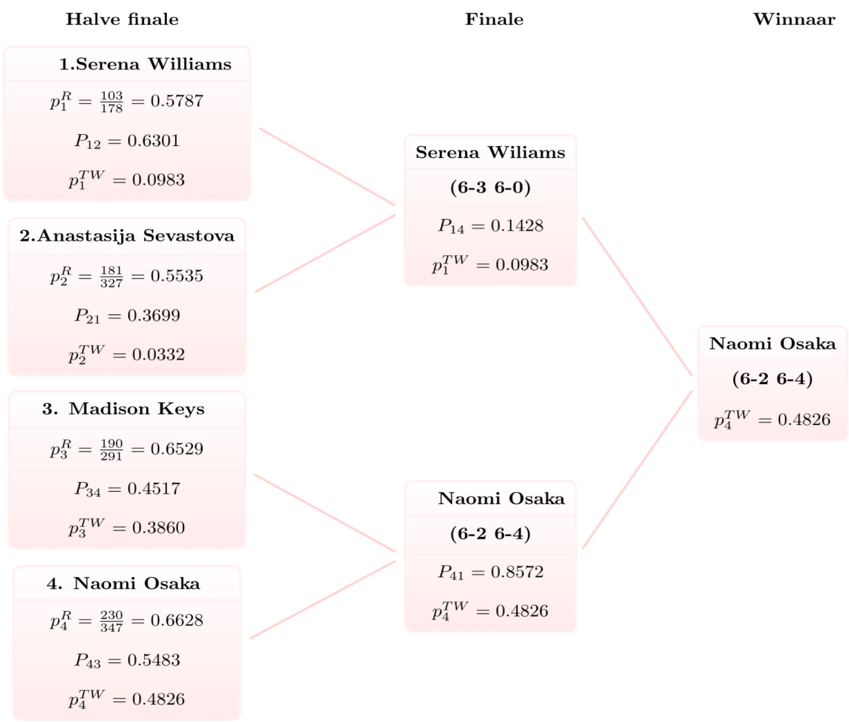
Figuur 7: US Open mannen

In figuur 7 zijn de halve finalisten van de US Open te zien. Ook hier blijkt de voorspelling te kloppen. Gegeven de halve finalisten blijkt Novak Djokovic veruit de grootste kanshebber om het toernooi te winnen, uitgaande van zijn prestaties in de toernooien ter voorbereiding op de US Open.
Zijn percentage p_4^R blijkt veruit de hoogste van de vier. Zijn goede vorm heeft hij dus ook in het toernooi weten door te zetten, waardoor hij het toernooi heeft gewonnen.



Figuur 8: Roland Garros vrouwen

In figuur 8 zijn de halve finales van Roland Garros van de dames te zien. Hier valt op dat de kansen om een punt te winnen, p_i^R van de vier dames vrij dicht bij elkaar liggen. Madison Keys blijkt de favoriet de zijn voor toernooiwinst. Dit komt vooral omdat zij het in de halve finale makkelijker zou moeten hebben dan Simona Halep, want p_1^R en p_3^R liggen erg dicht bij elkaar. Het bleek toch niet zo makkelijk te zijn voor Keys, want de halve finale verloor ze van Sloane Stephens.
Toen de finale bekend was, bleek Simona Halep de grote favoriet om te winnen. P_{14} is namelijk 0.6706, terwijl P_{41} 0.3294 is. Sloane Stephens was niet in staat om nog een keer te stunten en Simona Halep wist inderdaad, zoals verwacht, de finale te winnen.



Figuur 9: US Open vrouwen

Als laatste werden ook de halve finalisten van US Open bij de vrouwen onder de loep genomen. Dit is te zien in figuur 9.
Hier blijkt de voorspelling, net als bij de mannen, uit te komen. Osaka had het volgens de voorspelling in de halve finale nog redelijk moeilijk; de kans om deze te winnen was $P_{43} = 0.5483$, maar in de finale zou ze vrij gemakkelijk over Williams heen moeten lopen. De kans dat ze de finale zou winnen is namelijk $P_{41} = 0.8572$

7 Zijn de punten in een tenniswedstrijd onafhankelijk gelijk verdeeld?

Zoals in de inleiding vermeld is er in dit document vanuit gegaan dat de punten in een tenniswedstrijd onafhankelijk gelijk verdeeld zijn. Door dit aan te nemen, zijn de kansen p_A^R en p_B^R in een wedstrijd constant. Dit hoeft echter niet altijd het geval te zijn. Zo zou de kans om een punt te maken bijvoorbeeld af kunnen hangen van het momentum in een wedstrijd. Wanneer er al een aantal punten achter elkaar verloren zijn, is het vertrouwen vaak wat minder, waardoor de kans om het volgende punt te maken ook lager zou kunnen zijn.

In de literatuur zijn er een aantal onderzoeken te vinden over dit fenomeen. Zo hebben Jackson en Mourski [2] geconcludeerd dat de punten in een tenniswedstrijd niet onafhankelijk verdeeld zijn. Het psychologische aspect speelt wel degelijk een rol. Zij zagen bijvoorbeeld dat de kans om de tweede set te winnen groter wordt wanneer de eerste set al gewonnen is. Daarnaast concludeerde zij ook dat de vorm van de dag een rol speelt. De kans om van een bepaalde speler te winnen, is iedere dag weer anders.
Magnus en Klaassen [3] kwamen tot dezelfde conclusie in hun artikel: wanneer het vorige punt gewonnen is, is de kans om het volgende punt te winnen ook groter. Zij verwierpen dan ook de hypothese dat de punten in een tenniswedstrijd oanfhankelijk gelijk verdeeld zijn. Zij zagen echter wel dat dit effect kleiner wordt naarmate de speler sterker wordt. Hoe beter de speler, hoe beter de mentale gesteldheid en hoe minder zij zich laten beïnvloeden door punten uit het verleden. In hun onderzoek hebben ze 90000 punten onderzocht en hoewel de punten dus niet onafhankelijk gelijk verdeeld zijn, is de afwijking hiervan erg klein. Zelfs zo klein dat het Volgens Klaassen en Magnus s wel goed mogelijk is om met de methode uit dit document de winnaar van de wedstrijd te voorspellen.

8 Conclusie

In dit document is een methode gegeven om uit te kunnen rekenen wat de kans is om een tenniswedstrijd, en zelfs een tennistoernooi, te winnen. Deze kansen zijn gebaseerd op p_A^R, de kans voor speler A om een punt te pakken op zijn eigen service, en p_B^R, de kans voor speler B om een punt te pakken op zijn eigen service.
Voor speler A blijkt de kans om een game te winnen,

$$p_A^G = \sum_{j=0}^{2} p_A^G(4,j) + p_A^G(3,3)\frac{(p_A^R)^2}{(p_A^R)^2 + (q_A^R)^2}$$

te zijn, zoals berekend in hoofdstuk 2.
De kans voor speler A om de set te winnen is berekend in hoofdstuk 3 en wordt gegeven door

$$p_A^S = \sum_{j=0}^{4} p_A^S(6,j) + p_A^S(7,5) + p_A^S(6,6)p_A^T$$

Aan de hand van deze kans kon de kans op het winnen van de wedstrijd worden berekend, in hoofdstuk 4. Omdat er twee soorten wedstrijdvormen zijn, zijn er ook twee mogelijke kansen. In het geval van 'best of 3', wat op de grandslams door de vrouwen wordt toegepast, geldt

$$p_A^{WV} = (p_A^S)^2 + 2(p_A^S)^2 p_B^S$$

.
Bij de mannen op een grandslam geldt 'best of 5' en dit leidt tot de volgende kans:

$$p_A^{WM} = 3(p_A^S)^3 + 3(p_A^S)^3 p_B^S + 6(p_A^S)^3(p_B^S)^2$$

.
In hoofstuk 5 is de kans berekend om een toernooi te winnen. Dit is gedaan met behulp van matrices. Voor iedere ronde is er een aparte matrix waarin steeds alle mogelijke wedstrijden voor die ronde in worden gegeven. Door de recursieve formule, gegeven in (39) kan de kans worden berekend om de $n$de ronde te winnen, gegeven dat de $n$de ronde is behaald. Door deze conditionele verwachtingen te vermenigvuldigen met elkaar, kan voor iedere speler de kans worden berekend om het toernooi te winnen. Dit leidt uiteindelijk tot de kolomvector (44).
In hoofdstuk 6 is uiteindelijk bekeken of deze methode ook overeenkomt met de werkelijkheid. Hiervoor is aan de hand van de toernooien die voorafgaand aan de grandslams voor iedere speler i, p_i^R berekend. Deze kans geeft een beeld van de vorm van de speler en hoe goed deze speler is op een bepaalde ondergrond. Hiermee kan dan uiteindelijk berekend worden wat de kans is om van iedere andere speler te winnen, met (36), dan wel (37). Waarnaar er met de recursieve formule (39) berekend kan worden wat de kans is om het toernooi te winnen.

Dit document is van toegevoegde waarde op de bestaande literatuur. Niet alleen omdat hier alles duidelijk wordt uitgelegd, maar ook omdat hier alle theoretische kansen vergeleken worden met de data van 2018. Daarnaast wordt in de al bestaande literatuur vaak maar een of twee grandslams meegenomen, of wordt er alleen gekeken naar de mannen. In dit document worden alle grandslams meegenomen en wordt alles steeds gedaan voor zowel de mannen als de vrouwen. Dit geeft natuurlijk een veel beter beeld van de werkelijkheid dan wanneer maar een deel van de werkelijkheid wordt bekeken.

A Appendix

Notatie	Betreft	Wie serveert (of begint met serveren)	wie wint
p_A^R	rally	A	A
q_A^R	rally	A	B
p_B^R	rally	B	B
q_B^R	rally	B	A
p_A^G	game	A	A
q_A^G	game	A	B
p_B^G	game	B	B
q_B^G	game	B	A
p_A^{DC}	deuce	A	A
q_A^{DC}	deuce	A	B
p_B^{DC}	deuce	B	B
q_B^{DC}	deuce	B	A
p_A^S	set	A	A
q_A^S	set	A	B
p_B^S	set	B	B
q_B^S	set	B	A
p_A^T	tiebreak	A	A
q_A^T	tiebreak	A	B
p_B^T	tiebreak	B	B
q_B^T	tiebreak	B	A
$p_A^W V$	wedstrijd vrouwen	A	A
$q_A^W V$	wedstrijd vrouwen	A	B
$p_B^W V$	wedstrijd vrouwen	B	B
$q_B^W V$	wedstrijd vrouwen	B	A
$p_A^W M$	wedstrijd mannen	A	A
$q_A^W M$	wedstrijd mannen	A	B
$p_B^W M$	wedstrijd mannen	B	B
$q_B^W M$	wedstrijd mannen	B	A

De oplossing voor (8)-(11) is gegeven door:

$$
\begin{aligned}
p_A^S(6,0) &= (p_A^G q_B^G)^3 && \text{(A.1)}\\
p_A^S(6,1) &= 3(p_A^G)^3 q_A^G (q_B^G)^3 + 3(p_A^G)^4 p_B^G (q_B^G)^2 && \text{(A.2)}\\
p_A^S(6,2) &= 12(p_A^G)^3 q_A^G p_B^G (q_B^G)^3 + 6(p_A^G)^2 (q_A^G)^2 (q_B^G)^4\\
&\quad + 3(p_A^G)^4 (p_B^G)^2 (q_B^G)^2 && \text{(A.3)}\\
p_A^S(6,3) &= 24(p_A^G)^3 (q_A^G)^2 p_B^G (q_B^G)^2 + 24(p_A^G)^4 q_A^G (p_B^G)^2 (q_B^G)^2\\
&\quad + 4(p_A^G)^2 (q_A^G)^3 (q_B^G)^4 + 4(p_A^G)^5 (p_B^G)^3 q_B^G && \text{(A.4)}\\
p_A^S(6,4) &= 60(p_A^G)^3 (q_A^G)^2 (p_B^G)^2 (q_B^G)^3 + 40(p_A^G)^2 (q_A^G)^3 p_B^G (q_B^G)^4\\
&\quad + 20(p_A^G)^4 q_A^G (p_B^G)^3 (q_B^G)^2 + 5p_A^G (q_A^G)^4 (q_B^G)^5\\
&\quad + (p_A^G)^5 (p_B^G)^4 q_B^G && \text{(A.5)}\\
p_A^S(7,5) &= 100(p_A^G)^3 (q_A^G)^3 (p_B^G)^2 (q_B^G)^4 + 100(p_A^G)^4 (q_A^G)^2 (p_B^G)^3 (q_B^G)^3\\
&\quad + 25(p_A^G)^2 (q_A^G)^4 p_B^G (q_B^G)^5 + 25(p_A^G)^5 q_A^G (p_B^G)^4 (q_B^G)^2\\
&\quad + p_A^G (q_A^G)^5 (q_B^G)^6 + (p_A^G)^6 (p_B^G)^5 q_B^G && \text{(A.6)}
\end{aligned}
$$

De oplossing voor (14)-(16) wordt gegeven door

$$
\begin{aligned}
p_A^T(7,0) &= (p_A^R)^3 (q_B^R)^4 && \text{(A.7)}\\
p_A^T(7,1) &= 3(p_A^R)^3 q_A^R (q_B^R)^4 + 4(p_A^R)^4 p_B^R (q_B^R)^2 && \text{(A.8)}\\
p_A^T(7,2) &= 16(p_A^R)^4 q_A^R p_B^R (q_B^R)^3 + 6(p_A^R)^5 (p_B^R)^2 (q_B^R)^2\\
&\quad + 6(p_A^R)^3 (q_A^R)^2 (q_B^R)^4 && \text{(A.9)}\\
p_A^T(7,3) &= 40(p_A^R)^3 (q_A^R)^2 p_B^R (q_B^R)^4 + 10(p_A^R)^2 (q_A^R)^3 (q_B^R)^5\\
&\quad + 4(p_A^R)^5 (p_B^R)^3 (q_B^R)^2 + 30(p_A^R)^4 q_A^R (p_B^R)^2 (q_B^R)^3 && \text{(A.10)}\\
p_A^T(7,4) &= 50(p_A^R)^4 q_A^R (p_B^R)^3 (q_B^R)^3 + 5(p_A^R)^5 (p_B^R)^4 (q_B^R)^2\\
&\quad + 50(p_A^R)^2 (q_A^R)^3 p_B^R (q_B^R)^5 + 5p_A^R (q_A^R)^4 (q_B^R)^6\\
&\quad + 100(p_A^R)^3 (q_A^R)^2 (p_B^R)^2 (q_B^R)^4 && \text{(A.11)}\\
p_A^T(7,5) &= 30(p_A^R)^2 (q_A^R)^4 p_B^R (q_B^R)^5 + p_A^R (q_A^R)^5 (q_B^R)^6\\
&\quad + 200(p_A^R)^4 (q_A^R)^2 (p_B^R)^3 (q_B^R)^3 + 75(p_A^R)^5 q_A^R (p_A^R)^4 (q_B^R)^2\\
&\quad + 150(p_A^R)^3 (q_A^R)^3 (p_B^R)^2 (q_B^R)^4 + 6(p_A^R)^6 (p_B^R)^5 q_B^R && \text{(A.12)}
\end{aligned}
$$

De coëfficënten ongelijk aan 0, $a_{ij}^S(n) = b_{ij}^S(n)$ van (27)-(28) voor $n = 1, 2$ worden gegeven door

$$
\begin{array}{lllll}
a_{20}^S(1) = 6, & a_{21}^S(1) = -24, & a_{22}^S(1) = 36, & a_{23}^S(1) = -24, & a_{24}^S(1) = 6,\\
a_{30}^S(1) = -9, & a_{31}^S(1) = 51, & a_{32}^S(1) = -99, & a_{33}^S(1) = 81, & a_{34}^S(1) = -24,\\
a_{40}^S(1) = 3, & a_{41}^S(1) = -24, & a_{42}^S(1) = 60, & a_{43}^S(1) = -60, & a_{44}^S(1) = 21
\end{array}
\qquad \text{(A.13)}
$$

$$
\begin{array}{llll}
a^S_{10}(2) = 5, & a^S_{11}(2) = -25, & a^S_{12}(2) = 50, & a^S_{13}(2) = -50, \\
a^S_{14}(2) = -16, & a^S_{15}(2) = 124, & & \\
a^S_{20}(2) = -336, & a^S_{21}(2) = 424, & a^S_{22}(2) = -256, & a^S_{23}(2) = 60, \\
a^S_{24}(2) = 2, & a^S_{25}(2) = -5, & & \\
a^S_{30}(2) = 18, & a^S_{31}(2) = -198, & a^S_{32}(2) = 696, & a^S_{33}(2) = -1080, \\
a^S_{34}(2) = 774, & a^S_{35}(2) = -210, & & \\
a^S_{40}(2) = -8, & a^S_{41}(2) = 124, & a^S_{42}(2) = -560, & a^S_{43}(2) = 1060, \\
a^S_{44}(2) = -896, & a^S_{45}(2) = 280, & & \\
a^S_{50}(2) = 1, & a^S_{51}(2) = -25, & a^S_{52}(2) = 150, & a^S_{53}(2) = -350, \\
a^S_{54}(2) = 350, & a^S_{55}(2) = -126. & &
\end{array} \tag{A.14}
$$

De coëfficënten ongelijk aan 0, $a^T_{ij}(n) = b^T_{ij}(n)$ van (34)-(35) voor $n = 1, 2, 3$ worden gegeven door

$$
\begin{array}{lllll}
a^T_{30}(1) = 4, & a^T_{31}(1) = -16, & a^T_{32}(1) = 24, & a^T_{33}(1) = -16, & a^T_{34}(1) = 4, \\
a^T_{40}(1) = -3, & a^T_{41}(1) = 16, & a^T_{42}(1) = -30, & a^T_{43}(1) = 24, & a^T_{44}(1) = -7
\end{array} \tag{A.15}
$$

$$
\begin{array}{llll}
a^T_{20}(2) = 10, & a^T_{21}(2) = -50, & a^T_{22}(2) = 100, & a^T_{23}(2) = -100, \\
a^T_{24}(2) = 50, & a^T_{25}(2) = -10, & & \\
a^T_{30}(2) = -24, & a^T_{31}(2) = 166, & a^T_{32}(2) = -424, & a^T_{33}(2) = 516, \\
a^T_{34}(2) = -304, & a^T_{35}(2) = 70, & & \\
a^T_{40}(2) = 18, & a^T_{41}(2) = -166, & a^T_{42}(2) = 530, & a^T_{43}(2) = -774, \\
a^T_{44}(2) = 532, & a^T_{45}(2) = -140, & & \\
a^T_{50}(2) = -4, & a^T_{51}(2) = 50, & a^T_{52}(2) = -200, & a^T_{53}(2) = 350, \\
a^T_{54}(2) = -280, & a^T_{55}(2) = 84. & &
\end{array} \tag{A.16}
$$

$$
\begin{array}{llll}
a^T_{10}(3) = 6, & a^T_{11}(3) = -36, & a^T_{12}(3) = 90, & a^T_{13}(3) = -120, \\
a^T_{14}(3) = 90, & a^T_{15}(3) = -36, & a^T_{16}(3) = 6, & \\
a^T_{20}(3) = -25, & a^T_{21}(3) = 230, & a^T_{22}(2) = -775, & a^T_{23}(3) = 1300, \\
a^T_{24}(3) = -1175, & a^T_{25}(3) = 550, & a^T_{26}(3) = -105, & \\
a^T_{30}(3) = 40, & a^T_{31}(3) = -510, & a^T_{32}(3) = 2200, & a^T_{33}(3) = -4500, \\
a^T_{34}(3) = 4800, & a^T_{35}(3) = -2590, & a^T_{36}(3) = 560, & \\
a^T_{40}(3) = -30, & a^T_{41}(3) = 510, & a^T_{42}(3) = -2750, & a^T_{43}(3) = 6750, \\
a^T_{44}(3) = -8400, & a^T_{45}(3) = 5180, & a^T_{46}(3) = -1260, & \\
a^T_{50}(3) = 10, & a^T_{51}(3) = -230, & a^T_{52}(3) = 1550, & a^T_{53}(3) = -4550, \\
a^T_{54}(3) = 6580, & a^T_{55}(3) = -5620, & a^T_{56}(3) = 1260, & \\
a^T_{60}(3) = -1, & a^T_{61}(3) = 36, & a^T_{62}(3) = -315, & a^T_{63}(3) = 1120, \\
a^T_{64}(3) = -1890, & a^T_{65}(3) = 1512, & a^T_{66}(3) = -462. &
\end{array}
\tag{A.17}
$$

Referenties

[1] Newton, K., Keller, J., *Probability of Winning at Tennis 1. Theory and Data.* Blackwell Publishing, 2005.

[2] Jackson, D., Mosurki, K., *Heavy Defeats in Tennis: Psychological Momentum or Random Effect?*, CHANCE, 1997

[3] Magnus, J.R., Klaassen F.J.G.M., *Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel model*, Journal of the American Statistical Association, 96, 2001.

[4] *https://github.com/JeffSackmann/tennis_atp/blob/master/atp_matches_2018.csv.*

[5] *https://github.com/JeffSackmann/tennis_wta/blob/master/wta_matches_2018.csv*