

Wat Tenniscommentatoren Niet Weten

- Een analyse van vier jaar Wimbledon* -

Jan R. Magnus

&

Franc Klaassen

Samenvatting

Dit artikel gaat in op de statistische analyse van tennisdata. Gebruikmakend van gedetailleerde gegevens van vier jaar Wimbledon, worden allereerst enige tennishypotheses getoetst. Vervolgens beschrijven we de effecten van veranderingen in de spelregels, met name de regels bij het serveren. Als laatste worden op basis van een model voorspellingen gedaan over de ontwikkeling van de winstkansen van de spelers gedurende drie concrete Wimbledon-finales.

J.R. Magnus: CentER for Economic Research, Katholieke Universiteit Brabant:
Postbus 90153, 5000 LE Tilburg, Tel. 013-4663092, E-mail: Magnus@KUB.NL
F.J.G.M. Klaassen: Katholieke Universiteit Brabant, Vakgroep Econometrie:
Postbus 90153, 5000 LE Tilburg, Tel. 013-4668229, E-mail: F.J.G.M.Klaassen@KUB.NL

* Dit artikel werd eerder gepresenteerd op de workshop "Sport en Statistiek" in de Jaarbeurs te Utrecht op 27 oktober 1995. Wij zijn IBM-UK buitengewoon erkentelijk voor het beschikbaarstellen van de Wimbledon data.

1. Inleiding

Tennis als spel heeft een lange geschiedenis en was al bekend bij de Grieken (onder de naam "σφαριστική") en bij de Romeinen ("pila"). Het duurde echter tot ongeveer 1870 voordat het technisch mogelijk was om rubberen ballen te maken die goed op gras stuitten. Het eerste Wimbledon toernooi vond plaats in 1877 en er werd gespeeld om de "Silver Challenge Cup". Bij het eerste toernooi (alleen mannen uiteraard) werd reeds gespeeld om 3 gewonnen sets. Degene die het eerst zes games had gewonnen, won de set. Een set duurde dus maximaal elf games. Zelfs in 1877 was het direct duidelijk dat degene die mocht serveren in de elfde game een voordeel had. Derhalve gold al vanaf 1878 dat een speler een set moest winnen met een verschil van tenminste twee games. De regels zijn in essentie onveranderd gebleven sinds 1878.

Mede door de televisie is tennis een sport geworden die over de gehele wereld wordt bekeken en meebeleeft. Een tenniswedstrijd op Wimbledon duurt bij de heren, zeg, tweeëneenhalf uur. Verder is de effectieve speeltijd per punt ongeveer vijf seconden. Met zes punten per game duurt een game dus dertig seconden. Een set van tien games heeft zo een effectieve speeltijd van vijf minuten. Kortom, een wedstrijd van vier sets levert slechts twintig minuten aan effectieve speeltijd. De rest van de tijd (ruim twee uur) wordt gevuld door de commentatoren. In feite is een tenniswedstrijd op de TV een one-man (of two-men) show van de commentator(en), zo nu en dan onderbroken door vijf seconden actie. In dit licht bezien mag men van de commentatoren wel wat verwachten; zij spelen in feite de hoofdrol. Maar weten zij ook iets?

2. Enkele tennishypotheses

Wij hebben gegevens over ongeveer 500 wedstrijden, gespeeld in het heren- en damesenkel-spel op Wimbledon gedurende de jaren 1992-1995. Van elk van deze wedstrijden is ons het exacte scoreverloop bekend. Tevens weten we bij elk punt of de eerste of tweede service in was en of het punt beslist werd door een ace of een dubbele fout. Deze data zijn gebruikt om een aantal hypothesen te toetsen.

Hypothese 1: Het is een voordeel om in een set te beginnen met serveren.

Dit voordeel, als het al bestaat, kan alleen een psychologisch voordeel zijn. De redenering is, dat de ontvanger in de eerste game "steeds" een game achterstaat en dit zou een nadeel zijn. In de praktijk op Wimbledon wint 48,3% van de mannen een set als hij begint met serveren. Ook bij de vrouwen is het percentage lager dan 50%, namelijk 47,5%. Deze uitkomsten verschillen niet significant van 50%, maar ze duiden toch eerder op een servicenadeel dan op een servicevoordeel.

Hypothese 2: Met nieuwe ballen serveren geeft een voordeel.

Dit lijkt redelijk en wordt ook altijd vermeld door de commentatoren. Vooral bij de mannen wordt zo hard en effectvol geslagen, dat de ballen veel te lijden zouden hebben. Maar geven nieuwe ballen ook werkelijk een voordeel?

Bij de heren is de (empirische) kans om een servicegame te winnen op Wimbledon 81,7%. En inderdaad stijgt die kans iets als we alleen naar de games kijken waarin met nieuwe ballen wordt geserveerd, namelijk tot 83,3%. Dit verschil van 1,6% is echter niet significant. Bij de dames is de situatie omgekeerd. Bij hen is de kans op winst in een servicegame 64,2% en deze kans daalt tot 63,3% als met nieuwe ballen wordt geserveerd. Er is dus geen aanleiding om te geloven dat nieuwe ballen een voordeel zouden opleveren.

Hypothese 3: Na het breken van de servicegame van de tegenstander is er een grotere kans op verlies van de eigen servicegame.

Vermoedelijk komt deze hypothese voort uit het idee dat, na het breken van de service, de andere speler extra geprikkeld is om terug te breken, terwijl degene die de service net gebroken heeft, tevreden is met zijn of haar resultaat, zodat de concentratie (onbewust) iets daalt.

We hebben eerder al gezien dat bij de heren de kans om een servicegame te winnen 81,7% is. De kans op een break is dus 18,3%. Het blijkt dat de kans op een break na een break in de vorige game 17,2% is, dus lager in plaats van hoger. Het verschil is klein, maar de data ondersteunen deze hypothese dus geenszins. Ook bij de dames is het verschil onaanzienlijk.

De analyse van het Wimbledon materiaal is nog lang niet compleet. Wij hebben een collectie van 24 hypothesen die alle toetsbaar zijn. Onze voorlopige conclusie is dat de one-man show van de tenniscommentator sterk verbeterd kan worden door kennis te nemen van statistische analyses.

3. Regels

Zoals in de inleiding reeds vermeld, zijn de spelregels in tennis in essentie niet veranderd sinds 1878. De enige kleine wijziging was de tiebreak, ingevoerd om zeer lange sets te vermijden. Op dit moment hoort men steeds vaker dat de service te overheersend is, vooral bij de mannen en met name op het gras van Wimbledon. De kans om een servicegame te winnen is dus te groot en dit zou het spel minder aantrekkelijk maken.

Er zijn vele manieren om de kans op winst in een servicegame te verkleinen. Men kan het net verhogen of het servicevak verkleinen. Maar de meest gehoorde suggestie is om de tweede service af te schaffen. Wat zou hiervan het effect zijn?

Stel dat maar één service is toegestaan. Wellicht denkt men, dat die service dan iets makkelijker dan de huidige eerste service, maar iets moeilijker dan de huidige tweede service geslagen zal worden. Een ogenblik reflectie (of één blik in een boek over speltheorie) leert ons echter, dat dit onjuist is als we afzien van neveneffecten, die bijvoorbeeld ontstaan door een andere trainingsopzet. In de nieuwe situatie is de (enige) service gelijk aan de huidige tweede service.

De data van vier jaar Wimbledon leren ons, dat de kans dat de serveerder een punt wint als de tweede service in is bij de heren 59,7% bedraagt (bij de dames 53,5%). Bovendien weten we, dat de kans dat de tweede service in is bij de heren 86,7% bedraagt (86,8% bij de dames). Merk op dat we vooralsnog afzien van selectie-effecten, welke worden veroorzaakt doordat personen met een risicovolle eerste service zijn oververtegenwoordigd in beide percentages.

Als dus maar één service is toegestaan, dan wordt bij de heren de kans dat de serveerder een punt wint ($59,7\% \times 86,7\% =$) 51,8%. Aangezien deze kans volgens de data nu 65,1% is, betekent dit een daling van ruim 13%.

Bij de dames wordt de kans om op eigen service een punt te winnen ($53,5\% \times 86,8\% =$) 46,4%, een daling van ruim 10% (de kans is nu 56,5%). Merk op dat, net als bij de heren, de nieuwe kans dichter bij de 50% ligt, zodat de winst van een punt minder bepaald wordt door het feit wie er serveert. We zien bovendien dat het servicevoordeel door de nieuwe regel bij de dames is omgeslagen in een servicenadeel: de kans om een game te winnen is bij de dames nu groter voor de ontvangster dan voor de serveerster! Uit dit empirische resultaat van vier jaar Wimbledon volgt echter niet dat het afschaffen van de tweede service (of wellicht beter, van de eerste service) een slechte zaak voor het damestennis zou zijn. Het gevolg is wel een verandering in het karakter van het damestennis.

4. Voorspellen

Er is een lange traditie om uitkomsten van sportwedstrijden te voorspellen. Vóór het begin van een tenniswedstrijd op Wimbledon kan men gokken op de uitkomst van deze wedstrijd. Onze belangstelling gaat echter vooral uit naar het voorspellen van de uitkomst gedurende de wedstrijd. Onze computerprogramma's stellen ons in staat om de kans op wedstrijdwinst op elk punt van de wedstrijd te berekenen.

De drie hieronder gepresenteerde figuren zijn gebaseerd op ons eenvoudigste model dat

verwant is aan de negatief-binomiale verdeling (zie bijvoorbeeld Riddle, 1988, hoewel daarin alleen de kans aan het begin van de wedstrijd wordt bepaald). Er wordt slechts gebruik gemaakt van twee kansen, namelijk P_A (de kans dat speler A een punt wint als A serveert) en P_B (de kans dat B, de tegenstander van A, een punt wint als B serveert). We veronderstellen dat beide kansen constant zijn over de wedstrijd.

Hoewel deze veronderstelling van onafhankelijkheid in alle (ons bekende) artikelen over tennis gemaakt wordt, is het de vraag of deze realistisch is en, zo niet, in hoeverre de voorspellingen door het opleggen ervan beïnvloed worden. De Wimbledon gegevens zullen in de nabije toekomst gebruikt worden om deze onafhankelijkheid te onderzoeken. Indien nodig, kunnen we dan bijvoorbeeld de kansen af laten hangen van de score, het scoreverloop, het feit of het punt een "kritiek" punt (bv. breakpoint) is, etc. Dit laatst zou men kunnen modelleren door elke speler over twee in plaats van één kans te laten beschikken, namelijk (voor speler A) P_A en P_A' , waarbij P_A nu de kans weergeeft om op eigen service een "gewoon" punt te winnen en P_A' de kans is om op eigen service een "kritiek" punt te winnen. De programmatuur hiervoor is reeds ontwikkeld, maar de moeilijke vraag is nog, wat precies een "kritiek" punt is.

Uitgaande van het zojuist beschreven model is natuurlijk de keuze van de a priori kansen van belang. P_A en P_B hangen af van de sterkte van de spelers. Wij hebben alle spelers op grond van hun plaatsing ingedeeld in 5 categorieën, te weten 1-2 (bevat spelers die als eerste of tweede zijn geplaatst), 3-4, 5-8, 9-16 en ongeplaatst. Deze indeling wordt ook gehanteerd bij professionele toernooien. P_A wordt nu geschat door de relatieve frequentie van de gebeurtenis dat spelers in dezelfde categorie als A op eigen service het punt winnen tegen spelers in dezelfde categorie als B. P_B wordt analoog geschat. Vanzelfsprekend wordt hierbij afgezien van verschillen tussen spelers uit eenzelfde categorie. Bovendien is de van de wereldranglijst afgeleide plaatsingslijst ook gebaseerd op resultaten uit niet-gras toernooien en bevat zij geen informatie over de "vorm" en onderlinge resultaten van spelers. Om dit soort tekortkomingen te verminderen kan men bijvoorbeeld gebruik maken van de weddenschappen die voor elke wedstrijd op Wimbledon worden afgesloten; dit zal in de toekomst wellicht gebeuren.

De eerste figuur hieronder betreft de memorabele finale Graf - Novotna in het damesenkelspel van Wimbledon 1993. Graf won met 7-6, 1-6, 6-4. Op basis van de data en de plaatsing van beide speelsters hebben we de volgende kansen P_A respectievelijk P_B afgeleid:

$$P_{\text{Graf}} = 0,6130 \text{ en } P_{\text{Novotna}} = 0,5184.$$

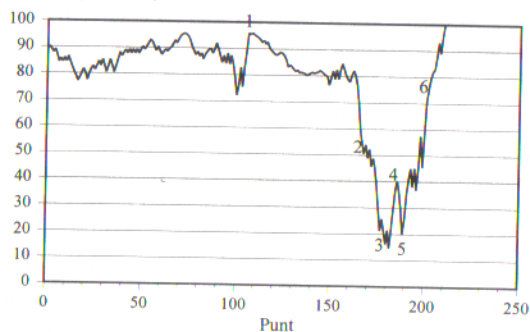
Het gevolg is dat Graf veruit favoriet is: bij het begin van de partij is haar kans om de wedstrijd te winnen 89,4%.

In de eerste set gaat de strijd gelijk op. Bij het begin van de tiebreak is de winstkans van Graf iets gedaald (84,5%). Na winst in de tiebreak (cijfer 1 in onderstaande figuur) is haar kans gestegen tot 95,9%. De tweede set is voor Novotna. Bij het begin van de derde set is de

Finale Dames Enkelspel Wimbledon 1993:

Graf verslaat Novotna na 210 punten met 7-6, 1-6, 6-4.

Kans (in procenten)



Figuur 1: De kans dat Graf de wedstrijd wint aan het eind van elk punt.

winstkans van Graf 79,8%. Bij 1-1 in de derde set breekt Novotna de service van Graf (cijfer 2) en bij 3-1 nogmaals (3). Als Novotna serveert op 4-1 en 40-30, is de winstkans van Graf nog slechts 14,6%. Dan breekt Graf terug (4) en na het wegwerken van twee breakpoints op eigen service (5), breekt ze wederom (6), om uiteindelijk de wedstrijd alsnog in haar voordeel te beslissen.

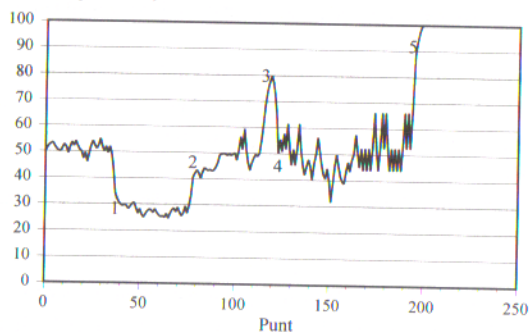
In figuur 2 analyseren we de damesfinale 1995 tussen Graf en Sanchez. Hiervoor is afgeleid:

$$P_{\text{Graf}} = 0,6186 \text{ en } P_{\text{Sanchez}} = 0,6186.$$

Finale Dames Enkelspel Wimbledon 1995:

Graf verslaat Sanchez na 199 punten met 4-6, 6-1, 7-5.

Kans (in procenten)



Figuur 2: De kans dat Graf de wedstrijd wint aan het eind van elk punt.

De kans dat Graf wint bij het begin van de partij is 50,0%. Sanchez breekt Graf in de zevende game (cijfer 1 in de figuur) en wint de eerste set. De winstkans van Graf is dan 25,0%. In de tweede set breekt Graf de service van Sanchez in de vierde game (2) en wint de tweede set. In de enerverende derde set breekt Graf op 1-1 (3) en breekt Sanchez direct terug (4). De strijd gaat gelijk op tot 5-5. Dan volgt de beroemde servicegame van Sanchez: dertien keer deuce! Graf wint deze game uiteindelijk (5) en haar winstkans is ineens gestegen tot 88,7%. De laatste game wint Graf "op love".

Tenslotte geeft figuur 3 de finale herenenkelspel 1995 tussen Sampras en Becker, waarvoor is afgeleid:

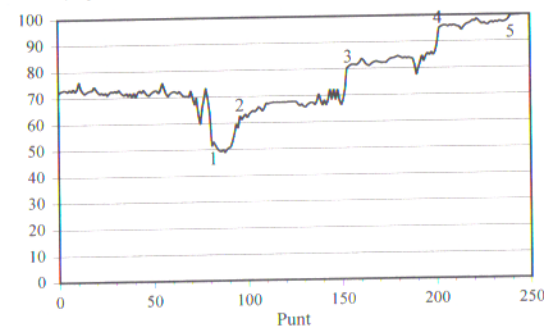
$$P_{\text{Sampras}} = 0,6961 \text{ en } P_{\text{Becker}} = 0,6574,$$

zodat bij het begin van de partij Sampras 71,9% kans heeft om zijn titel te prolongeren.

Finale Heren Enkelspel Wimbledon 1995:

Sampras verslaat Becker na 246 punten met 6-7, 6-2, 6-4, 6-2.

Kans (in procenten)



Figuur 3: De kans dat Sampras de wedstrijd wint aan het eind van elk punt.

In de eerste set zijn er geen breaks en wint Becker de tiebreak (1). In de tweede set breekt Sampras onder andere bij 1-1 (2) en wint de set. In de derde en vierde set breekt Sampras de service van Becker drie maal (3, 4 en 5), waarbij de laatste break de winstkans voor Sampras nog slechts marginaal verhoogt, aangezien hij al bijna zeker is van de overwinning; één game later is de wedstrijd inderdaad voorbij.

5. Conclusie

Tennis leent zich bijzonder goed voor statistische analyse. Niet alleen worden er veel punten gescoord (in tegenstelling tot bijvoorbeeld voetbal), maar ook is de invloed van de scheids-

rechter betrekkelijk klein (eveneens in tegenstelling tot voetbal). Onze analyse richt zich op vier deelgebieden. Allereerst het toetsen van gangbare tennishypotheses, ten tweede de analyse van de mogelijke effecten van veranderingen in de spelregels, ten derde de vorming van wedstrijdstrategieën en de coaching, vooral wat betreft de service, en ten vierde het ontwikkelen van een model dat de uitkomst van een wedstrijd optimaal voorspelt. Op het derde na, zijn in dit artikel alle deelgebieden kort aan bod gekomen. Op basis van de gedetailleerde gegevens van enige jaren Wimbledon hopen we in de nabije toekomst het tennisspel verder te kunnen karakteriseren.

5. Referentie

Riddle, L.H. (1988), "Probability Models for Tennis Scoring Systems", *Applied Statistics*, 37, 63-75.