De geringe praktische toepasbaarheid van de beschrijving van een tenniswedstrijd als een stochastisch proces : een reactie

F.J.G. Toevank en M.J.J. Wolters

Samenvatting

Dit artikel is een commentaar op het artikel 'Wat tenniscommentatoren niet weten' van J. Magnus en F. Klaassen. Zij beschrijven in hun artikel onder andere een tenniswedstrijd als een stochastisch proces, waarbij gebruik wordt gemaakt van slechts twee parameters. De praktische toepassing van deze beschrijving is volgens ons echter zeer gering, aangezien de parameters niet nauwkeurig kunnen worden geschat. Bovendien is de uitkomst van het proces, de kans om een wedstrijd te winnen, zeer gevoelig voor veranderingen in de parameters, door de getrapte telling bij tennis. Een kleine aanpassing van het model kan de robuustheid vergroten, maar dat gaat wel ten koste van de nauwkeurigheid.

1 Inleiding

Het is mogelijk om een tenniswedstrijd te beschrijven als een stochastisch proces. Hiervoor hoeven slechts enkele parameters te worden gedefinieerd. In onderstaand artikel betogen wij, in reaktie op Magnus en Klaassen, dat dit echter geen praktische toepassing heeft, aangezien de werkelijke waarden van de parameters niet nauwkeurig bekend zijn en de uitkomst van het proces zeer gevoelig is voor deze waarden. Een kleine aanpassing van het model kan de robuustheid vergroten.

2 Beschrijving van het proces

In Stewart [1989] wordt een tenniswedstrijd beschreven met behulp van één parameter: de kans dat één van beide spelers een punt wint (p). Vanzelfsprekend is die kans voor de andere speler dan 1-p. Door deze aanname is een wedstrijd te beschrijven als een negatief binomiale verdeling met afwijkende binomiaalcoëfficiënten. Stewart concludeert dat: '[...] you can, in principle, write down an explicit expression for the probability of winning a tennis-match. [...] I haven't actually carried this procedure out, because the result would be *enormous*.'

Omdat de opslag in het huidige tennis een zeer belangrijke plaats inneemt, hebben Magnus en Klaassen het proces van Stewart uitgebreid. In plaats van de kans dat een punt wordt gewonnen, nemen ze de kans dat een speler scoort op zijn eigen opslag. Elke speler heeft dus zijn eigen parameter, die we voor het gemak p_1 en p_2 zullen noemen. Aangezien dit model een uitbreiding is op Stewart, wordt analytisch uitschrijven van de kans op het winnen van een wedstrijd helemaal Sisyphus-arbeid. Daarom hebben we besloten een simulatieprogramma te schrijven dat onder deze aannames een tenniswedstrijd "naspeelt".

3 Resultaten

Om deze beschrijving nu te gebruiken voor winstkansschattingen met behulp van empirische data, zullen we schattingen moeten hebben voor p_1 en p_2. Deze schattingen zouden gemaakt

kunnen worden met behulp van historische data. Echter, om een nauwkeurige schatting van p_1 te krijgen, met een maximale fout van 0,005, zijn minimaal 10000 gespeelde punten nodig. Bij de exacte schattingen die Magnus en Klaassen presenteren neemt dit aantal toe tot maar liefst honderd miljoen. Wij vermoeden zeer sterk dat de hoeveelheid beschikbare gegevens niet toereikend is om deze nauwkeurigheid te bereiken.

Bovendien treden enkele andere moeilijkheden op, want er zijn vele factoren die deze parameters beïnvloeden. In de eerste plaats zijn p_1 en p_2 zeer afhankelijk van elkaar: het maakt een groot verschil tegen wie iemand speelt. Bovendien ontwikkelen spelers zich in een verschillend tempo, waardoor de parameters niet constant zullen zijn in de tijd. Ook de ondergrond en de weersomstandigheden spelen een rol. Omdat spelers niet vaak op dezelfde ondergrond met dezelfde weersomstandigheden spelen, neemt de hoeveelheid bruikbare data verder af. Het vaak aangehaalde begrip "vorm van de dag" is bovendien zo ongrijpbaar dat kwantificeren ervan onmogelijk lijkt.

De genoemde moeilijkheden zijn niet zo ernstig als de uitkomst van het proces, de kans om een wedstrijd te winnen, robuust zou zijn voor veranderingen van de waarden van deze parameters. Met het onderstaande voorbeeld willen we illustreren dat het proces juist èrg gevoelig is.

Stel, je hebt de parameterwaarden geschat en je vindt voor p_1 de waarde 0,64 en voor p_2 0,66, dan kunnen we de computer 10.000 maal een wedstrijd tussen deze twee spelers laten spelen. We kunnen nu een schatting maken van de winstkansen, dat is het percentage van de wedstrijden dat door een bepaalde speler wordt gewonnen. De 10.000 wedstrijden zijn voldoende om gegeven een betrouwbaarheidsinterval van 95% een maximale absolute fout te hebben van 0,005. Dit levert de volgende tabel op:

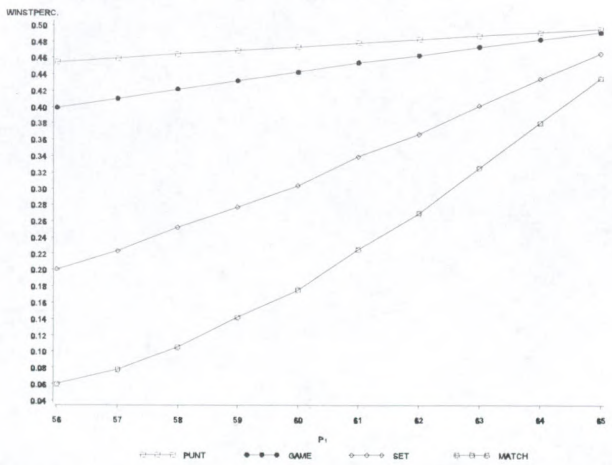
	Speler 1 $p_1 = 0{,}64$		Speler 2 $p_2 = 0{,}66$	
Totaal aantal gewonnen wedstrijden	3800	38%	6200	62%
Totaal aantal gewonnen sets	17557	43%	22865	57%
Totaal aantal gewonnen games	203633	48%	218490	52%
Totaal aantal gewonnen punten	1333483	49%	1380744	51%

We zien dat het totaal aantal gewonnen punten in 10.000 wedstrijden per speler nauwelijks verschilt: 49 procent van het totale aantal wordt gewonnen door de zwakkere speler. Kijken we echter naar het totaal aantal gewonnen wedstrijden, dan is het verschil tussen beide spelers al beduidend groter. Slechts 38 procent van de wedstrijden wordt gewonnen door de slechtste speler. Dit is het gevolg van de bij tennis gebruikte getrapte telling: er moeten punten gescoord worden om een game te winnen, games gewonnen om een set te winnen en sets om een wedstrijd te winnen.

Vervolgens houden we p_2 constant op 0,66 en verlagen we p_1 naar 0,63. Een afname die hoogstwaarschijnlijk binnen de foutenmarge van de schatting valt, gegeven de eerder genoemde moeilijkheden bij het schatten van de parameters. Dit levert de volgende resultaten op:

	Speler 1 $p_1 = 0{,}63$		Speler 2 $p_2 = 0{,}66$	
Totaal aantal gewonnen wedstrijden	3251	33%	6749	67%
Totaal aantal gewonnen sets	16101	40%	23993	60%
Totaal aantal gewonnen games	196316	47%	218516	53%
Totaal aantal gewonnen punten	1302067	49%	1370847	51%

Ondanks dat het percentage gewonnen punten nagenoeg constant blijft, constateren we een opmerkelijke terugval in het aantal gewonnen wedstrijden voor de zwakste speler. Een afname van 0,01 in p_1 zorgt voor een daling van ongeveer 5,5 ($\pm$ 1) procent in het percentage gewonnen wedstrijden. Dit onderschrijft de eerder genoemde gevoeligheid. Dit wordt nogmaals geïllustreerd aan de hand van de volgende figuur. Hierin hebben we voor p_2 een vaste waarde van 0,66 genomen en wordt p_1 gevarieerd.



Naar onze mening kan gezegd worden dat de geschatte winstkansen, die Magnus en Klaassen presenteren, waarschijnlijk erg onnauwkeurig zijn: de parameters p_1 en p_2 zijn niet zo precies te schatten als zij hebben gedaan door de te geringe hoeveelheid beschikbare en bruikbare data.

De vraag is nu of we het model niet robuuster kunnen maken door bijvoorbeeld niet de kans op puntwinst, maar de kans op gamewinst te schatten. Een dergelijke aanpassing levert een robuuster model. Dit gaat echter ten koste van de nauwkeurigheid van de schatting van de gebruikte parameters; er worden immers minder games dan punten per wedstrijd gespeeld. De afweging die dus moet worden gemaakt, is die tussen robuustheid en nauwkeurigheid. Magnus en Klaassen geven de voorkeur aan het laatste. Echter, omdat de nauwkeurigheid van de parameters toch al een zwak punt is, lijkt het ons beter om de robuustheid zwaarder te laten wegen.

4 Conclusie

Het is inderdaad mogelijk om een tenniswedstrijd te beschrijven als een stochastisch proces. De gevoeligheid voor een verandering in de parameters is door de getrapte telling echter zodanig groot, dat praktische toepassing alleen mogelijk is als men zeer nauwkeurige ex ante schattingen kan maken voor deze parameters. Wij zijn van mening dat er zoveel factoren van invloed zijn op de parameters, dat de gewenste nauwkeurigheid niet kan worden bereikt. Daarom hebben wij een lichte voorkeur voor een iets robuuster model, door in plaats van puntwinst uit te gaan van bijvoorbeeld gamewinst.

Referenties

Stewart, Ian, (1989), *Game, Set & Math, - Enigmas and Conundrums* -, Basil Blackwell, Oxford and Cambridge, MA.

Wat tenniscommentatoren niet weten: Antwoord aan Toevank en Wolters

Jan R. Magnus en Franc Klaassen

Toevank en Wolters reageren op één aspect van ons artikel, namelijk de aanname in sectie 4 (over voorspellen) dat een wedstrijd beschreven kan worden door twee parameters die gedurende de wedstrijd niet veranderen: de kans dat speler 1 een punt wint op zijn/haar service (p_1) en analoog voor speler 2 (p_2). Hun kritiek is dat een kleine verandering in deze kansen grote gevolgen heeft voor de kans op wedstrijdwinst en zij stellen een "verbetering" voor.

In tegenstelling tot Toevank en Wolters (hierna TW) hoeven wij de kansen niet te simuleren, omdat ons programma het mogelijk maakt de kansen exact te berekenen. Net als TW stellen we p_2=0,66 en bekijken we het effect van een stijging van p_1=0,63 tot 0,64. We vinden:

	p_1=0,63	p_2=0,66	p_1=0,64	p_2=0,66
Kans om wedstrijd te winnen (max. 5 sets)	31,8%	68,2%	37,7%	62,3%
Kans om set te winnen	40,2%	59,8%	43,5%	56,5%
Kans om game te winnen	47,5%	52,5%	48,4%	51,6%

hetgeen redelijk correspondeert met de simulaties van TW.

TW stellen dat het moeilijk is om p_1 en p_2 nauwkeurig te schatten. In zekere zin is dat waar. Maar wij hebben beschikking over 88.883 punten gespeeld op Wimbledon. Daaruit moet toch een redelijke schatting te maken zijn. Terecht merken TW op dat de moeilijkheden om p_1 en p_2 te schatten worden vergroot doordat p_1 en p_2 afhankelijk van elkaar zijn (we houden er echter rekening mee tegen wie iemand speelt), doordat spelers zich in verschillend tempo ontwikkelen (dit komt tot uitdrukking in de plaatsing en ook hier houden wij rekening mee), doordat de ondergrond verschillend is (wij kijken alleen naar gras, dus dit argument speelt niet), en doordat de "vorm van de dag" van invloed is (hiervan abstraheren wij inderdaad).

Natuurlijk is het zo dat de uitkomsten gevoelig zijn voor kleine veranderingen in p_1 en p_2. Maar dit heeft weinig te maken met, zoals TW stellen, de getrapte manier van tellen bij tennis. Bij elke (redelijke) manier van tellen zou dit effect optreden. De "oplossing" van TW is om niet de puntwinst maar de gamewinst te schatten, zeg q_1 (de kans dat speler 1 zijn/haar service game wint) en q_2 (analoog voor speler 2). Hierdoor zouden de resultaten robuuster worden: onnauwkeurigheden in de schattingen $\hat{q}_1$ en $\hat{q}_2$ leiden tot kleinere onnauwkeurigheden in de kans op wedstrijdwinst dan onnauwkeurigheden in $\hat{p}_1$ en $\hat{p}_2$. Dit gaat echter alleen op als de onnauwkeurigheden in $\tilde{q}_1$ en $\tilde{q}_2$ die via het model volgen uit $\hat{p}_1$ en $\hat{p}_2$ groter zijn dan die in de direct geschatte $\hat{q}_1$ en $\hat{q}_2$. Dit blijkt niet het geval te zijn, omdat p_1 en p_2 veel nauwkeuriger geschat worden dan q_1 en q_2. Immers, bij het herenenkelspel hebben we waarnemingen over 59.466 punten en 9.367 games (gemiddeld 6,3 punten per game). Uit deze dataset schatten we $\hat{p}$=0,65 en $\hat{q}$=0,82. De standaardfouten zijn respectievelijk 0,0020 en 0,0040. Ons weer baserend op het model, zien we in bovenstaande tabel dat een verandering van 1%-punt in p (van 63% naar 64%) leidt tot een verandering van 1,8%-punt in q (van 79,5% naar 81,3%). Een eerste-orde benadering toont dan aan dat de standaardfout van $\tilde{q}$ bij benadering gelijk is aan 1,8 maal de standaardfout van $\hat{p}$, dus 1,8×0,0020=0,0036. Dit is niet groter dan de gevonden standaardfout van $\hat{q}$ van 0,0040, dus van een grotere robuustheid is geen sprake.