L. M. Simons

Competitiviteit van een sporttoernooi

Bachelorscriptie

Scriptiebegeleider: Dr. F. M. Spieksma

Datum Bachelorexamen: 27 juni 2014

Mathematisch Instituut, Universiteit Leiden

Abstract

In deze bachelorscriptie worden verschillende indices besproken die de competitiviteit, de mate van onvoorspelbaarheid, van een sporttoernooi kwantificeren. Hierbij wordt gebruik gemaakt van een vooraf bepaalde ranking van de deelnemers en de uitslagen van het gespeelde toernooi, waarbij gelijkspel buiten beschouwing wordt gelaten. De indices die behandeld zullen worden zijn de (optimale) toernooi-index, Slater's i, Kendall's τ, het competitief evenwicht en de gewogen toernooi-index. Daarnaast zullen relaties tussen de (optimale) toernooi-index, Slater's i en Kendall's τ gegeven worden, waarna een mooi resultaat besproken wordt. Tot slot wordt aan de hand van de dynamische ranking, gebaseerd op de win-verlies score, besproken hoe de vooraf vastgestelde ranking bepaald kan worden.

Trefwoorden: toernooi, ranking, competitiviteit, toernooi-index, Slater's i, Kendall's τ, dynamische ranking.

Inhoudsopgave

1 Inleiding **1**

2 Indices algemeen **3**

3 Toernooi-index **4**

3.1 Optimale toernooi-index 6

4 Slater's i **10**

5 Relatie optimale toernooi-index en Slater's i **12**

6 Kendall's τ **15**

6.1 Bepaling Σ 15

6.2 Verdeling Σ 19

7 Relatie Kendall's τ met toernooi-index en Slater's i **22**

7.1 De aannames 22

7.2 Relatie Kendall's τ met de toernooi-index en $\mathcal{S}(\rho)$ onder aannames 25

8 Gevolgen relatie indices **30**

9 Overige indices **32**

9.1 Competitief evenwicht 32

9.2 Gewogen toernooi-index 33

10 Dynamische ranking **34**

10.1 Originele win-verlies score, Park & Newman 34

10.2 Dynamische win-verlies score 35

10.3 Resultaten 37

11 Conclusie en vervolgonderzoek **39**

Index **41**

1 Inleiding

> *Over de doelstelling tijdens het komende WK is Van Gaal volstrekt duidelijk: de bondscoach wil met het Nederlands elftal wereldkampioen worden. (...) "Misschien verwachten we wel te veel met onze doelstelling. Ik denk echter dat deze ploeg elke tegenstander kan verslaan, al kunnen we ook verliezen", weet Van Gaal. "Dat hoort nu eenmaal bij sport en dat accepteren we."*

Het vorige citaat geeft weer hoe onvoorspelbaar een toernooi kan zijn [1]. In dit geval gaat het om het WK voetbal van 2014 waar op dit moment volop over wordt gespeculeerd, zo ook door Van Gaal. Hoe vaak zal een "zwakker" team winnen van een "sterker" team en welk team kunnen we dan als "zwak" beschouwen en welk team als "sterk"? De onvoorspelbaarheid van het antwoord op deze vraag is precies wat het kijken naar wedstrijden van een toernooi interessant maakt. Wanneer het namelijk niet te voorspellen is welk team het toernooi zal winnen, dan zullen we dit toernooi als competitief beschouwen.

In deze bachelorscriptie worden bepaalde indices behandeld, die in de onderzochte literatuur geïntroduceerd zijn. Deze indices geven informatie over de mate van onvoorspelbaarheid van een toernooi. Het tweede hoofdstuk geeft meer informatie over welke indices worden behandeld, waarnaast de meest gebruikte parameters geïntroduceerd worden. Vervolgens wordt in hoofdstuk 3 de toernooi-index geïntroduceerd, waarbij er ook gekeken wordt naar een optimalisatie, namelijk de optimale toernooi-index. In dit hoofdstuk zullen ook enkele lemma's afgeleid en bewezen worden, die meer kennis en inzicht geven in de twee behandelde indices. In hoofdstuk 4 volgt een derde index, Slater's i, waarvoor een relatie bestaat met de optimale toernooi-index. Deze relatie wordt behandeld en bewezen in hoofdstuk 5. De laatste index die een relatie heeft met de eerder genoemde indices, is Kendall's τ, die geïntroduceerd wordt in hoofdstuk 6. De relatie tussen Kendall's τ en de eerder genoemde indices bestaat slechts onder bepaalde aannames, die besproken worden in hoofdstuk 7. De relatie zelf zal ook in dit hoofdstuk afgeleid en bewezen worden. Vervolgens worden de gevolgen van de relaties beschreven in hoofdstuk 8. In het voorlaatste hoofdstuk wordt beschreven op welke manier men een realistische pre-ranking kan bepalen, namelijk de dynamische pre-ranking. Deze methode is toegepast op data verkregen van een tennistoernooi, wat samengevat wordt in hoofdstuk 10.3. Als laatste volgt een conclusie en worden ideeën gegeven voor een vervolgonderzoek.

In de literatuur worden vrijwel geen relaties gegeven tussen de indices en indien deze wel gegeven worden volgt er geen bewijs. Daarom is het doel van deze bachelorscriptie de indices duidelijk te beschrijven en uit te leggen aan de hand van voorbeelden, waarnaast de relaties tussen de indices worden afgeleid en bewezen. De lemma's en stellingen beschreven in deze bachelorscriptie zijn dus niet te vinden in de onderzochte literatuur, behalve stelling 5.2, waarin de relatie tussen de optimale toernooi-index en Slater's i gegeven wordt. Het bewijs van deze stelling is echter ook niet te vinden in de onderzochte literatuur.

Verder wordt in deze bachelorscriptie geen rekening gehouden met gelijkspel, omdat gelijkspel geen uitslag geeft over welk van de twee beschouwde teams "sterker" is. Dit is een belangrijk aspect in de indices, waardoor gelijkspel voor moeilijkheden kan zorgen. Verder is het belangrijk om op te merken dat vooral gesproken zal worden over teams en niet over individuen. Echter, nergens worden aannames gedaan waardoor de gevonden resultaten niet gelden voor individuen, waardoor overal het woord 'team' kan worden vervangen door 'individu'.

2 Indices algemeen

Om te bepalen welke sporttoernooi het meest competitief is, zijn de volgende indices bekend in de literatuur: de toernooi-index [2], Slater's i [2][3], Kendall's τ [2][4][5], het competitief evenwicht [2] en de gewogen toernooi-index [2]. Deze indices geven elk op hun eigen manier aan in hoeverre een toernooi competitief is. Hierbij wordt gebruik gemaakt van meerdere *pre-rankings* ρ, een gegeven ranking die bepaald is voordat het toernooi plaatsvindt, en een *toernooiranking* ρ_t, een ranking als gevolg van een gespeeld toernooi. In het definiëren van de indices houden we rekening met alle pre-rankings, aangezien we de conclusies zo algemeen mogelijk willen houden. In de praktijk is de pre-ranking ofwel bekend ofwel er is er één te construeren, wat beschreven wordt in het hoofdstuk genaamd Dynamische ranking. Als definitie van het *competitief* zijn van een toernooi gebruiken we de mate van onvoorspelbaarheid, die wordt berekend door de pre-ranking te vergelijken met het gespeelde toernooi of de toernooiranking. Bijvoorbeeld, als volgt dat een "sterker" team verloren heeft van een "zwakker" team, kunnen we dit zien als onvoorspelbaar en dus competitief. Hierbij wordt een team als "*sterker*"dan een ander team beschouwd, indien dit team eerder op de pre-ranking staat dan het andere team. Daarnaast wordt gebruik gemaakt van de volgende parameters:

n: het aantal teams,
m: het aantal keer dat elk team tegen elk ander team speelt,
N: het aantal wedstrijden.

In de literatuur wordt beschreven hoe de toernooiranking bepaald wordt, maar een wiskundige definitie ontbreekt [2, p. 2,3]. We definiëren de toernooiranking aan de hand van de *toernooimatrix* A. Hierbij staat a_{ij} voor het aantal gewonnen wedstrijden van team i spelend tegen team j. Voor team i, met $1 \leq i \leq n$, geldt

$$w_i = \frac{\sum_{j=1}^{n} a_{ij}}{m(n-1)}. \tag{1}$$

Hierbij noemen we w_i de *winfrequentie* van team i, dus het aantal overwinningen van team i ten opzichte van het aantal wedstrijden dat i gespeeld heeft [2, p. 8]. Stel nu dat geldt $w_i < w_j$ met $1 \leq i, j \leq n$, definieer dan $\rho_t(j) < \rho_t(i)$ waardoor i dus later op de ranking komt te staan dan j. Als voor team i en j, met $1 \leq i < j \leq n$, geldt $w_i = w_j$, neem dan $\rho_t(i) < \rho_t(j)$. Op deze manier kunnen we dus de toernooiranking construeren. Indien geldt $w_i = w_j$, wordt gebruik gemaakt van de nummering van de teams. In dit geval is de toernooiranking niet uniek, aangezien de teams ook anders genummerd kunnen worden. We spreken dus van een *unieke toernooiranking*, als deze ranking niet afhankelijk is van de nummering van de teams.

3 Toernooi-index

De eerste index die we zullen behandelen is de *toernooi-index*, $\mathcal{T}(\rho)$, die als functie gebruikmaakt van de pre-ranking ρ. Deze index is relatief nieuw in de literatuur en is afkomstig uit een artikel geschreven door Lundh (2006), waar vanaf nu naar wordt gerefereerd met hoofdartikel [2]. In dit artikel wordt helaas de interpretatie van de toernooi-index minimaal weergegeven, wat in dit hoofdstuk meer naar voren zal komen. Voordat we de toernooi-index kunnen definiëren, moeten we weten hoe we een gespeelde wedstrijd waarderen. Zij t_1 en t_2 twee teams die tegen elkaar spelen in wedstrijd i van het toernooi, $1 \leq i \leq N$. Neem aan dat t_2 onder t_1 staat in een gegeven, dus vooraf vastgestelde, pre-ranking, wat we noteren als $\rho(t_1) < \rho(t_2)$. Een voorbeeld van een waardering van wedstrijd i tussen team t_1 en team t_2, waarbij $1 \leq i \leq N$, is

$$v_i = \begin{cases} 1, & \text{als } t_1 \text{ wint,} \\ -1, & \text{als } t_2 \text{ wint,} \\ 0, & \text{als gelijkspel.} \end{cases} \tag{2}$$

Deze manier van het waarderen van een wedstrijd wordt ook wel *Just win baby*, oftewel JWB, genoemd [2, p. 1]. De ranking wordt als het ware gestraft met één strafpunt, als een team wint van een team dat hoger staat in de ranking. Gelijkspel wordt dus gewaardeerd met de waarde 0. Hiervoor is gekozen, omdat rekening houden met gelijkspel voor enkele moeilijkheden zou kunnen zorgen. Denk hierbij aan het feit dat bij gelijkspel niet wordt besloten welk team als beter beschouwd kan worden. Hierdoor is dus niet direct te bepalen welk team hoger geranked moet worden, waar we wel naar op zoek zijn.

Opmerking 3.1. De aanname die gemaakt wordt, is dat gelijkspel niet voorkomt. Vervolgens geldt dat het aantal wedstrijden gelijk is aan

$$N = \frac{n(n-1)m}{2}.$$

We kunnen nu de toernooi-index als volgt definiëren

$$\mathcal{T}(\rho) = \frac{\sum_{i=1}^{N} v_i}{N}, \tag{3}$$

waar we indirecte afhankelijkheid zien van een pre-ranking via de definitie van v_i [2]. Aan de toernooi-index is te zien in hoeverre de uitslag van een gespeeld toernooi overeenkomt met een gegeven pre-ranking, door te kijken naar het aantal onterecht gewonnen wedstrijden. Hierbij wordt met een *onterecht gewonnen wedstrijd* bedoeld, dat een team dat lager op de pre-ranking staat dan een ander team toch wint in de wedstrijd tussen deze twee teams. Gebruikmakend van de wedstrijd waardering (2) geldt

$$-1 \leq \mathcal{T}(\rho) \leq 1.$$

We zien dat de pre-ranking minpunten krijgt als een "zwakker" team wint van een "sterker" team, omdat dit een *inconsistentie* betekent in de pre-ranking ten opzichte van het gespeelde toernooi. In het geval $\mathcal{T}(\rho) = 1$ kunnen we concluderen dat het niet is voorgekomen dat een team dat lager op de pre-ranking stond, heeft gewonnen van een team dat hoger op de pre-ranking stond. Daarom zal een toernooi met $\mathcal{T}(\rho) = 1$ niet als competitief worden beschouwd. Een toernooi is wel competitief als $\mathcal{T}(\rho)$ verder van 1 ligt, omdat dit betekent dat er inconsistenties aanwezig zijn in de uitslag van het gespeelde toernooi ten opzichte van de pre-ranking. Echter, omdat we geen echte maatstaf voor het competitief zijn hebben, kunnen we niet zeggen in welke mate het toernooi competitief is als de toernooi-index verder van 1 ligt.

Door gebruik te maken van deze definitie voor de toernooi-index, is het noodzakelijk te kijken naar alle gespeelde wedstrijden. In een groot toernooi zal dit veel werk zijn, waardoor het handig is de toernooi-index op een andere manier te definiëren. De toernooi-index kan namelijk herschreven worden door gebruik te maken van de toernooimatrix A, waarin de wedstrijd waardering wordt verwerkt. Daarnaast worden de strafpunten voor de ranking anders bepaald, namelijk met behulp van de *sign functie*

$$sgn(x) = \begin{cases} -1, & \text{als } x < 0, \\ 0, & \text{als } x = 0, \\ 1, & \text{als } x > 0. \end{cases}$$

We kunnen nu de toernooi-index (3) herschrijven tot [2]

$$\mathcal{T}(\rho) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n} sgn(\rho(j) - \rho(i)) a_{ij}. \tag{4}$$

Met deze definitie is te zien dat het aantal onterecht gewonnen wedstrijden anders berekend wordt. Per team beschouwen we alle wedstrijden die tegen de andere teams gespeeld zijn. Het aantal gewonnen wedstrijden van team i tegen team j geven we een negatieve waarde in de som als team i lager op de pre-ranking staat dan team j, zoals in vergelijking (3). De betekenis van de toernooi-index is dus niet veranderd, alleen worden de wedstrijden niet apart beschouwd maar in groepen wedstrijden waarin team i speelt tegen team j. Hierdoor is de toernooi-index eenvoudig te bepalen.

In het volgende voorbeeld is te zien hoe de toernooi-index wordt bepaald voor een klein toernooi.

Voorbeeld 3.1. Zij $n = 3$ en $m = 2$, dan spelen drie teams twee keer tegen elkaar. Beschouw de volgende mogelijke toernooimatrix

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

We zien dat geldt $\rho_t = [2, 3, 1]$. Zij $\rho = [1, 2, 3]$, dan geldt met behulp van definitie (4)

$$\begin{aligned}\mathcal{T}(\rho) &= \frac{-2+1-2-1}{6} \\ &= -\frac{2}{3}.\end{aligned}$$

We kunnen concluderen dat het gespeelde toernooi met deze pre-ranking vrij competitief is, doordat de waarde dichtbij -1 ligt. Dit betekent dat er vaak een team onterecht gewonnen heeft van een ander team, waarbij het winnende team van tevoren als minder "sterk" werd beschouwd. Neem nu $\rho = [3, 2, 1]$, dan geldt $\mathcal{T}(\rho) = \frac{2}{3}$. We zien dat met deze pre-ranking het toernooi minder competitief is, doordat team 3 duidelijk de winnaar is en ook bovenaan op de pre-ranking staat.

Uit dit voorbeeld is duidelijk te concluderen dat verschillende pre-rankings voor een ander resultaat zorgen. De eerst gekozen pre-ranking geeft aan dat het toernooi competitief is, waar de tweede pre-ranking juist aangeeft dat het toernooi niet competitief is. Het is dus zeer belangrijk hoe de gegeven pre-ranking eruit ziet, welke in dit geval niet bekend is. In de praktijk bestaat er overigens een idee over hoe de pre-ranking eruit hoort te zien, door bijvoorbeeld te kijken naar eerder gespeelde toernooien en wedstrijden.

3.1 Optimale toernooi-index

In het hoofdartikel wordt de optimale toernooi-index beschreven. Deze is te bepalen door gebruik te maken van alle pre-rankings, dus alle mogelijke permutaties van de teams. We berekenen voor elke pre-ranking de toernooi-index en zullen hiervan het maximum nemen, waardoor we de pre-ranking verkrijgen die het gespeelde toernooi het best weergeeft. Definieer daarom de optimale toernooi-index als volgt

$$\mathcal{T}_0 = \max_{\rho} \mathcal{T}(\rho) = \mathcal{T}(\rho_0), \tag{5}$$

waarbij ρ_0 de *optimale pre-ranking* wordt genoemd. In het geval dat de pre-ranking bekend is voor een toernooi, kunnen we de optimale toernooi-index berekenen en controleren of de gegeven pre-ranking de toernooi-index optimaliseert. Is dit niet het geval, dan komt een andere pre-ranking beter overeen met de uitslag van het toernooi. Hieruit kunnen we concluderen dat met de gegeven pre-ranking onvoorspelbaarheden zijn opgetreden met betrekking tot de uitslag van het toernooi, waardoor het toernooi als competitief beschouwd kan worden.

Omdat $-1 \leq \mathcal{T}(\rho) \leq 1$ voor alle ρ, volgt dat geldt $-1 \leq \mathcal{T}_0 \leq 1$. Maar kan de optimale toernooi-index wel alle genoemde waardes aannemen? Als geldt $\mathcal{T}_0 = -1$, dan moet voor alle pre-rankings gelden $\mathcal{T}(\rho) = -1$ wegens de definitie van de optimale toernooi-index.

Hierdoor moet dus gelden, voor alle pre-rankings, dat bij elke gespeelde wedstrijd het team dat lager op de pre-ranking staat wint van een team dat hoger op de pre-ranking staat. Dit kan echter niet gelden voor alle teams, omdat er geen toernooimatrix bestaat die hiervoor zorgt. Deze gedachte is nuttig om meer inzicht te verkrijgen in de werking en betekenis van de (optimale) toernooi-index, waardoor het volgende lemma is ontstaan.

Lemma 3.2. *Voor* $\mathcal{T}_0 = \max_{\rho} \mathcal{T}(\rho)$ *geldt*

$$\mathcal{T}_0 \neq -1.$$

Bewijs. Stel $\mathcal{T}_0 = -1$, dan geldt voor alle ρ

$$\mathcal{T}(\rho) = -1,$$

ofwel

$$\sum_{i=1}^{n} \sum_{j=1}^{n} sgn(\rho(j) - \rho(i)) a_{ij} = -N. \tag{6}$$

Dus geldt voor alle ρ en $1 \leq i, j \leq n$,

$$sgn(\rho(j) - \rho(i)) = \begin{cases} - & , \quad a_{ij} \neq 0, \\ + \text{ of } - & , \quad a_{ij} = 0. \end{cases} \tag{7}$$

Kies, zonder verlies van algemeenheid, een team α en β zodanig dat geldt

$$a_{\beta\alpha} \neq 0. \tag{8}$$

Dan bestaat er een pre-ranking ρ waarvoor geldt $\rho(\alpha) > \rho(\beta)$, omdat we alle mogelijke pre-rankings beschouwen. Hieruit volgt

$$\begin{aligned} -N &= \sum_{i=1}^{n} \sum_{j=1}^{n} sgn(\rho(j) - \rho(i)) a_{ij} \\ &= \sum_{\substack{(i,j),(i,j) \neq (\beta,\alpha) \\ 1 \leq i \leq n, 1 \leq j \leq n}} sgn(\rho(j) - \rho(i)) a_{ij} + a_{\beta\alpha}. \end{aligned}$$

Deze uitkomst zorgt voor een tegenspraak met een combinatie van de vergelijkingen (6), (7) en (8). Hieruit volgt $\mathcal{T}(\rho) \neq -1$, waardoor geldt

$$\mathcal{T}_0 \neq -1.$$

$\square$

Bovenstaand lemma is ontstaan door naar de definities van de toernooi-index en de optimale toernooi-index te kijken en goed na te denken over de betekenis hiervan. Op dezelfde manier is lemma 3.3 ontstaan. Per definitie van de optimale toernooi-index hebben we namelijk een bovengrens voor de toernooi-index, afhankelijk van de optimale pre-ranking. Bestaat er dan ook een ondergrens voor de toernooi-index, afhankelijk van de optimale pre-ranking, afgezien van de waarde -1? Na het uitschrijven van enkele voorbeelden is te zien dat er voor elke pre-ranking ρ een pre-ranking ρ' bestaat waarvoor geldt $-\mathcal{T}(\rho) = \mathcal{T}(\rho')$. Hierbij zijn de teams in ρ' precies in omgekeerde volgorde geranked, ten opzichte van ρ. Deze ρ' bestaat dus ook voor de optimale pre-ranking ρ_0, waarvoor geldt $\mathcal{T}(\rho_0) = \mathcal{T}_0$.

Lemma 3.3. *Voor een willekeurige pre-ranking ρ geldt*

$$-\mathcal{T}_0 \leq \mathcal{T}(\rho) \leq \mathcal{T}_0.$$

Bewijs. We zullen beide ongelijkheden apart bewijzen.

- Per definitie van de optimale toernooi-index, zie vergelijking (3), geldt

$$\mathcal{T}_0 = \max_{\rho} \mathcal{T}(\rho) \geq \mathcal{T}(\rho), \quad \forall \rho, \tag{9}$$

omdat er maar eindig veel pre-rankings bestaan.

- Zij $\mathcal{T}_0 = \max\limits_{\rho} \mathcal{T}(\rho) = \mathcal{T}(\rho_0)$. Beschouw een willekeurige ranking ρ^*. Dan geldt

$$\mathcal{T}(\rho^*) \leq \mathcal{T}(\rho_0).$$

Omdat we alle mogelijke rankings beschouwen, bestaat er een ρ' zodanig dat

$$\text{als } \rho^*(i) < \rho^*(j), \text{ dan } \rho'(i) > \rho'(j) \quad \forall 1 \leq i, j, \leq n.$$

Nu volgt

$$\begin{aligned} \mathcal{T}(\rho^*) &= \sum_{i=1}^{n} \sum_{j=1}^{n} sgn(\rho^*(j) - \rho^*(i)) a_{ij} \\ &= -\sum_{i=1}^{n} \sum_{j=1}^{n} sgn(\rho'(j) - \rho'(i)) a_{ij} \\ &= -\mathcal{T}(\rho'). \end{aligned} \tag{10}$$

Neem aan dat geldt $\mathcal{T}(\rho^*) < -\mathcal{T}(\rho_0)$. Dan volgt met vergelijking (10)

$$-\mathcal{T}(\rho') < -\mathcal{T}(\rho_0),$$

waaruit volgt

$$\mathcal{T}(\rho') > \mathcal{T}(\rho_0).$$

Dit is in tegenspraak met vergelijking (9), waaruit volgt dat de gemaakte aanname fout is. Daarom geldt $\mathcal{T}(\rho^*) \geq -\mathcal{T}(\rho_0)$, waarbij ρ^* willekeurig gekozen is. Hieruit volgt dat geldt

$$-\mathcal{T}_0 \leq \mathcal{T}(\rho), \quad \forall \rho.$$

□

Voorbeeld 3.1 (Vervolg)**.** In Tabel 1 zijn de toernooi indices voor alle mogelijke pre-rankings te vinden.

Pre-ranking	Toernooi-index
$[1,2,3]$	$-\frac{2}{3}$
$[1,3,2]$	$-\frac{2}{3}$
$[2,1,3]$	0
$[2,3,1]$	$\frac{2}{3}$
$[3,1,2]$	0
$[3,2,1]$	$\frac{2}{3}$

Tabel 1: Toernooi indices bij toernooimatrix A.

Er geldt $\mathcal{T}_0 = \max\limits_{\rho} \mathcal{T}(\rho) = \mathcal{T}(\rho_0) = \frac{2}{3}$ met als optimale pre-ranking

$$\rho_0 \in \{[2,3,1],[3,2,1]\}.$$

Hieruit kunnen we concluderen dat zelfs in het beste geval, dus met een pre-ranking die de uitslag het beste omschrijft, inconsistenties bestaan, omdat de optimale toernooi-index niet gelijk is aan 1. Hieruit kunnen we concluderen dat het toernooi enigszins competitief is. We gebruiken het woord "enigszins", aangezien we geen maat voor competitiviteit hebben gedefinieerd.

Omdat de pre-ranking niet bekend is, kunnen we niets zeggen over het al dan niet competitief zijn van het toernooi. Dan volgt namelijk een optimalisatie over alle mogelijke pre-rankings. We zien in bovenstaand voorbeeld, dat er twee mogelijke toernooirankings bestaan, indien men de toernooiranking niet zou baseren op de nummering van de teams. Deze twee mogelijke toernooirankings zijn precies de ranking die de collectie van ρ_0 vormen. Intuïtief is duidelijk dat de toernooiranking als optimale pre-ranking zorgt voor het minste aantal inconsistenties ten opzichte van de uitslag van het toernooi, wat bewezen wordt in stelling 8.1. In het geval dat voor ρ_0 meerdere mogelijke pre-rankings bestaan, is er minstens één inconsistentie in de optimale pre-ranking ten opzichte van het gespeelde toernooi.

4 Slater's i

De derde index die we zullen beschouwen is *Slater's i*. Patrick Slater noemt het minimale aantal inconsistenties in een gegeven ordening i [3]. Slater geeft echter geen uitdrukking voor deze i, daarnaast wordt in de andere literatuur ook alleen een omschrijving gegeven. De ordening die wij willen beschouwen is de pre-ranking, waarvan we dus het aantal inconsistenties willen bepalen. Definieer daarom

$$\mathcal{S}(\rho) = \sum_{i=1}^{n} \sum_{\rho(j)<\rho(i)} a_{ij}, \tag{11}$$

waarbij we dus een inconsistentie tellen als een team i een wedstrijd gewonnen heeft van een team j, terwijl j hoger op de pre-ranking staat dan i. Omdat Slater's i het minimale aantal inconsistenties aangeeft in de pre-ranking, definiëren we Slater's i als volgt

$$\mathcal{S}_i = \min_{\rho} \{\mathcal{S}(\rho) \mid \rho \text{ is een ranking}\}. \tag{12}$$

Slater's i geeft dus met het aantal inconsistenties weer in welke mate het toernooi competitief is. Hierbij is een toernooi dus niet competitief indien er een minimum aantal inconsistenties plaatsvindt. In het volgende voorbeeld wordt duidelijk hoe Slater's i berekend wordt en wordt extra uitleg gegeven over het aantal inconsistenties in de pre-ranking ten opzichte van het gespeelde toernooi.

Voorbeeld 3.1 (Vervolg)**.** In de volgende tabel staan alle waardes van $\mathcal{S}(\rho)$, berekend met behulp van vergelijking (11).

Pre-ranking	S(ρ)
$[1,2,3]$	5
$[1,3,2]$	5
$[2,1,3]$	3
$[2,3,1]$	1
$[3,1,2]$	3
$[3,2,1]$	1

Tabel 2: $\mathcal{S}(\rho)$ voor alle ρ bij toernooimatrix A.

Uit de tabel kunnen we halen dat geldt $\mathcal{S}_i = 1$ horend bij de pre-ranking

$$\rho \in \{[2,3,1],[3,2,1]\}.$$

Omdat er geldt $\mathcal{S}_i = 1$, weten we dat we zelfs in het minimale geval nog een inconsistentie vinden in de pre-ranking ten opzichte van de toernooiuitslag. Deze inconsistentie wordt veroorzaakt doordat team 2 een keer van team 3 heeft gewonnen en andersom. Omdat de teams twee keer tegen elkaar spelen is er geen onderlinge winnaar aan te wijzen. Daarbij hebben beide teams team 1 verslagen, waardoor ook hieruit niet te concluderen is welk team "sterker" is. Hierdoor zal team 2 boven team 3 plaatsen een inconsistentie geven, maar andersom ook. Hieruit is te concluderen dat een hoge waarde van $\mathcal{S}_i$ betekent dat een toernooi competitief is, doordat zelfs de optimale pre-ranking niet aan kan geven welk team de meest waarschijnlijke winnaar is van het toernooi.

Merk op dat de gevonden collectie pre-rankings, die het aantal inconsistenties in Slater's i minimaliseert, gelijk is aan de collectie optimale pre-rankings. Dan rijst al snel de vraag of dit toeval is of dat er een relatie bestaat tussen de optimale toernooi-index en Slater's i. Dit laatste blijkt het geval te zijn, zoals volgt in het volgende hoofdstuk.

5 Relatie optimale toernooi-index en Slater's i

Dat er een relatie bestaat tussen de optimale toernooi-index en Slater's i is intuïtief duidelijk, aangezien beide indices gebruik maken van het aantal inconsistenties in een gegeven pre-ranking. De relatie die genoemd wordt in het bestudeerde hoofdartikel, maar welke niet wordt bewezen, is de volgende

$$\mathcal{S}_i = \frac{N(1-\mathcal{T}_0)}{2}. \tag{13}$$

Deze definitie van $\mathcal{S}_i$ is intuïtief ook duidelijk. Namelijk, als $\mathcal{T}_0 = -1$, dan komt de pre-ranking niet overeen met het gespeelde toernooi en geldt zelfs dat de pre-ranking het maximale aantal inconsistenties heeft, namelijk N. Dit is terug te zien in vergelijking (13). Voor het bewijs van de relatie, bewijzen we eerst de relatie tussen de toernooi-index en $\mathcal{S}(\rho)$, zoals beschreven in het volgende lemma.

Lemma 5.1. *Voor een willekeurige pre-ranking ρ geldt*

$$N\mathcal{T}(\rho) = N - 2\mathcal{S}(\rho).$$

Bewijs. Zij ρ een willekeurige ranking. Omdat een team nooit tegen zichzelf zal spelen, weten we dat in de toernooimatrix A altijd geldt

$$a_{ij} = 0, \text{ voor } i = j. \tag{14}$$

Verder geldt, omdat we geen rekening houden met gelijkspel,

$$a_{ij} + a_{ji} = m. \tag{15}$$

Zij ρ een willekeurige ranking. Met behulp van defenitie (4) vinden we nu

$$\begin{aligned}
N\mathcal{T}(\rho) &= \sum_{i=1}^{n}\sum_{j=1}^{n} sgn(\rho(j)-\rho(i))a_{ij} \\
&\overset{(14)}{=} \sum_{i=1}^{n}\left(-\sum_{\rho(j)<\rho(i)} a_{ij} + \sum_{\rho(i)<\rho(j)} a_{ij}\right) \\
&\overset{(15)}{=} -\sum_{i=1}^{n}\sum_{\rho(j)<\rho(i)} a_{ij} + \sum_{i=1}^{n}\sum_{\rho(i)<\rho(j)} (m - a_{ji}) \\
&= -\sum_{i=1}^{n}\sum_{\rho(j)<\rho(i)} a_{ij} + \sum_{i=1}^{n}\sum_{\rho(i)<\rho(j)} m - \sum_{i=1}^{n}\sum_{\rho(i)<\rho(j)} a_{ji}
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{i=1}^{n} \sum_{\rho(j)<\rho(i)} a_{ij} + m[(n-1)+(n-2)+\ldots+1] - \sum_{i=1}^{n} \sum_{\rho(i)<\rho(j)} a_{ji} \\
&= -\sum_{i=1}^{n} \sum_{\rho(j)<\rho(i)} a_{ij} + \frac{n(n-1)m}{2} - \sum_{i=1}^{n} \sum_{\rho(i)<\rho(j)} a_{ji}.
\end{aligned}
$$

Zij $i = \alpha$ willekeurig en kies $j = \beta$ vóór α in de ranking willekeurig, dus $\rho(\beta) < \rho(\alpha)$. Dan komt $a_{\alpha\beta}$ voor in

$$\sum_{i=1}^{n} \sum_{\rho(j)<\rho(i)} a_{ij}.$$

Bovendien geldt dat α na β in de ranking staat, waardoor $a_{\alpha\beta}$ voorkomt in

$$\sum_{i=1}^{n} \sum_{\rho(i)<\rho(j)} a_{ji}.$$

Omdat $i = \alpha$ en $j = \beta$ willekeurig gekozen zijn, geldt dit voor alle i en alle j die vóór i in de ranking staan. Dus geldt

$$\sum_{i=1}^{n} \sum_{\rho(j)<\rho(i)} a_{ij} = \sum_{i=1}^{n} \sum_{\rho(i)<\rho(j)} a_{ji}.$$

Hierdoor geldt

$$
\begin{aligned}
N\mathcal{T}(\rho) &= N - 2\sum_{i=1}^{n} \sum_{\rho(j)<\rho(i)} a_{ij} \\
&\overset{(11)}{=} N - 2\mathcal{S}(\rho).
\end{aligned}
$$

Doordat ρ willekeurig gekozen is, geldt

$$N - 2\mathcal{S}(\rho) = N\mathcal{T}(\rho), \qquad \forall \rho.$$

$\square$

Vervolgens kunnen we lemma 5.1 gebruiken om vergelijking (13), oftewel de relatie tussen de optimale toernooi-index en Slater's i, te bewijzen.

Stelling 5.2. *Zij N het aantal beslissende wedstrijden, $\mathcal{T}_0$ de optimale toernooi-index en $\mathcal{S}_i$ Slater's index. Dan geldt*

$$N - 2\mathcal{S}_i = N\mathcal{T}_0.$$

Bewijs. Lemma 5.1 geeft

$$N - 2\mathcal{S}(\rho) = N\mathcal{T}(\rho), \qquad \forall\rho.$$

Zij ρ willekeurig. Als we $N\mathcal{T}(\rho)$ willen maximaliseren, zullen we $N - 2\mathcal{S}(\rho)$ ook moeten maximaliseren. Dit houdt in dat $\mathcal{S}(\rho)$ geminimaliseerd moet worden. Hieruit volgt dus

$$\begin{aligned} \max_{\rho} N\mathcal{T}(\rho) &= \max_{\rho} \left(N - 2\mathcal{S}(\rho)\right) \\ N\mathcal{T}_0 &\overset{(5)}{=} N - 2 \min_{\rho} \mathcal{S}(\rho) \\ N\mathcal{T}_0 &\overset{(12)}{=} N - 2\mathcal{S}_i. \end{aligned}$$

□

We kunnen met behulp van stelling 5.2 concluderen dat $\mathcal{S}_i$ naast het tellen van het minimale aantal inconsistenties, ook weergeeft in hoeverre een optimale pre-ranking te bepalen is. In het geval $\mathcal{S}_i = 0$ weten we dat er een optimale pre-ranking bestaat zonder inconsistenties. Hoe verder de waarde van 0 verwijderd raakt, hoe meer inconsistenties zich in de optimale pre-ranking bevinden.

6 Kendall's τ

Een vierde index die we zullen beschouwen is *Kendall's τ*, vernoemd naar Maurice Kendall [4]. Deze index gaat uit van een pre-ranking die wordt vergeleken met een toernooiranking. τ geeft een correlatie tussen de twee rankings, door de toernooiranking een bepaalde score, Σ, te geven. Σ geeft het verschil in het aantal consistenties en het aantal inconsistenties van de toernooiranking ten opzichte van de pre-ranking. Hoe Σ bepaald kan worden, wordt afgeleid in de volgende paragraaf. We definiëren

$$\begin{aligned} \tau &= \frac{\Sigma}{n(n-1)/2} \\ &= \frac{2\Sigma}{n(n-1)}, \end{aligned} \tag{16}$$

met $-1 \leq \tau \leq 1$ [4]. Te zien is dat er niet direct gebruik wordt gemaakt van het gespeelde toernooi. Echter, indirect is dit wel het geval aangezien de toernooiranking een gevolg is van het gespeelde toernooi.

6.1 Bepaling Σ

De score Σ kan op verschillende manieren berekend worden, zoals wordt uitgelegd in een artikel geschreven door Maurice Kendall [4]. We zullen een kleine aanpassing maken in één van de methodes die wordt beschreven [4, p. 83], aangezien de herordening die daar gebruikt wordt niet te gebruiken is in het geval van een toernooi. Als we de pre-ranking en toernooiranking op de beschreven manier herordenen, worden de inconsistenties en consistenties niet meer op de juiste manier geteld. Aan de hand van een voorbeeld wordt uitgelegd hoe Σ bepaald kan worden.

Voorbeeld 6.1. Beschouw de volgende pre-ranking en toernooiranking voor drie teams

$$\begin{aligned} \rho &= [2, 3, 1], \\ \rho_t &= [3, 1, 2]. \end{aligned}$$

We zullen de teams anders nummeren, zodat de pre-ranking in oplopende volgorde komt te staan en we een simpele manier creëeren om het aantal inconsistenties en consistenties te tellen. Te zien is dat team 2 in de pre-ranking op de eerste plaats staat, maar in de toernooiranking op de derde plaats. Daarom zullen we een 1 zetten in de toernooiranking op de plaats van team 2. Als we dit doen voor alle teams vinden we

$$\begin{aligned} \rho^* &= [1, 2, 3], \\ \rho_t^* &= [2, 3, 1]. \end{aligned}$$

Beschouw vervolgens team 1. In de pre-ranking staat team 1 bovenaan, maar in de toernooiranking onderaan. We zien dus dat de ranking van team 1 in de toernooiranking voor twee inconsistenties zorgt en geen consistenties, doordat team 1 twee teams links van zich heeft staan en geen teams rechts. Streep nu team 1 weg uit de pre-ranking en toernooiranking. Beschouw vervolgens team 2 en herhaal de procedure. Alleen team 3 is over in beide rankings, en zal geen waardes toevoegen aan Σ. We vinden nu

$$\Sigma = -2 + 0 + 0 + 1 = -1.$$

We zien in het bovenstaande voorbeeld dat Σ het aantal inconsistenties en consistenties op een handige manier telt. Daarbij zien we ook een formule voor Σ ontstaan, namelijk de volgende

$$\begin{aligned} \Sigma &= \#\text{ consistenties} - \#\text{ inconsistenties} \\ &\coloneqq n_c - n_d, \end{aligned} \tag{17}$$

waar we gebruik maken van de definitie van τ die beschreven staat in een boek geschreven door Langville en Meyer [5, p. 204]. Omdat we in totaal $\frac{n(n-1)}{2}$ paren vergelijken en elk paar in de verzameling consistenties of inconsistenties plaatsen, geldt

$$\frac{n(n-1)}{2} = n_c + n_d. \tag{18}$$

Hieruit volgt een nieuwe uitdrukking voor Σ als gevolg van een combinatie van de definitie beschreven door Kendall en de definitie beschreven door Langville en Meyer

$$\begin{aligned} \Sigma &= n_c - n_d \\ &= \frac{n(n-1)}{2} - 2 \cdot n_d \\ &= \frac{n(n-1)}{2} - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathbb{1}_{\{\rho_t^*(i) > \rho_t^*(j)\}}. \end{aligned}$$

Dankzij de oplopende pre-ranking kunnen we voor elk team het aantal inconsistenties tellen door in de toernooiranking te tellen hoeveel teams met een hoger nummer voor het beschouwde team staan. Op deze manier vergelijken we de aangepaste toernooiranking met de aangepaste, en bovendien oplopende, pre-ranking.

Voor het bepalen van τ is bovenstaande definitie van Σ vrij omslachtig. We kunnen het omzetten van de pre-ranking en toernooiranking achterwege laten en de paren teams vergelijken in de oorspronkelijke vorm. Doordat ρ^* namelijk in oplopende volgorde staat, kunnen we in de indicatorfunctie in Σ de volgende voorwaarde toevoegen

$$\rho^*(i) < \rho^*(j),$$

aangezien dit altijd het geval is. Daarnaast kunnen we een extra indicatorfunctie toevoegen aan de som, met als voorwaarde

$$\rho^*(i) > \rho^*(j),$$

aangezien dit nooit het geval is. De genoemde voorwaardes worden toegevoegd, zodat we eenvoudiger een relatie kunnen vinden tussen Kendall's τ, $\mathcal{T}(\rho)$ en $\mathcal{S}(\rho)$. We vinden de volgende nieuwe, meer eenvoudige, definitie van Σ

$$\begin{aligned}\Sigma &= \frac{n(n-1)}{2} - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathbb{1}_{\{\rho_t^*(i)>\rho_t^*(j)\}} \\ &= \frac{n(n-1)}{2} - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \left[\mathbb{1}_{\{\rho^*(i)<\rho^*(j),\rho_t^*(i)>\rho_t^*(j)\}} + \mathbb{1}_{\{\rho^*(i)>\rho^*(j),\rho_t^*(i)<\rho_t^*(j)\}}\right].\end{aligned}$$

We onderzoeken dus per twee teams of ze in de toernooiranking anders zijn geranked dan op de pre-ranking. Omdat ρ^* en ρ_t^* slechts een andere nummering geeft aan de teams, waardoor de ranking van de teams niet verandert, geldt

$$\Sigma = \frac{n(n-1)}{2} - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \left[\mathbb{1}_{\{\rho(i)<\rho(j),\rho_t(i)>\rho_t(j)\}} + \mathbb{1}_{\{\rho(i)>\rho(j),\rho_t(i)<\rho_t(j)\}}\right]. \tag{19}$$

Eerst zullen we aan de hand van het volgende voorbeeld laten zien dat deze Σ dezelfde waardes oplevert als de waardes die eerder gevonden zijn.

Voorbeeld 6.1 (Vervolg)**.** We vinden

$$\begin{aligned}n_d &= \sum_{i=1}^{n}\sum_{j=i+1}^{n} \left[\mathbb{1}_{\{\rho^*(i)<\rho^*(j),\rho_t^*(i)>\rho_t^*(j)\}} + \mathbb{1}_{\{\rho^*(i)>\rho^*(j),\rho_t^*(i)<\rho_t^*(j)\}}\right] \\ &= 1+1+0 \\ &= 2,\end{aligned}$$

en

$$\begin{aligned}n_d &= \sum_{i=1}^{n}\sum_{j=i+1}^{n} \left[\mathbb{1}_{\{\rho(i)<\rho(j),\rho_t(i)>\rho_t(j)\}} + \mathbb{1}_{\{\rho(i)>\rho(j),\rho_t(i)<\rho_t(j)\}}\right] \\ &= 1+0+1 \\ &= 2.\end{aligned}$$

Er geldt dus inderdaad dat we de pre-ranking en toernooiranking niet hoeven om te schrijven in dit voorbeeld. Substitueren we de waarde $n_d = 2$ in (19), dan vinden we

$$\Sigma = -1,$$

wat we eerder ook als waarde hebben gevonden.

Omdat Σ het aantal inconsistenties aftrekt van het aantal consistenties, kunnen we concluderen dat Σ een maximale waarde aanneemt als alle teams op de juiste plaats staan ten opzichte van de pre-ranking. In dit geval geldt

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbb{1}_{\{\rho_t^*(j)<\rho_t^*(i)\}} = 0,$$

waardoor Σ als maximale waarde $\frac{n(n-1)}{2}$ aanneemt. Doordat we een normalisatie willen voor τ, ontstaat definitie (16) voor Kendall's τ. Merk op dat τ direct afhankelijk is van de gekozen pre-ranking ρ en de toernooiranking ρ_t. Daarom noteren we vanaf nu voor Kendall's τ

$$\begin{aligned}
\tau &= \tau(\rho, \rho_t) \\
&\overset{(16)}{=} \frac{2\Sigma}{n(n-1)} \\
&\overset{(19)}{=} \frac{2\Big(\frac{n(n-1)}{2} - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\Big[\mathbb{1}_{\{\rho(i)<\rho(j),\rho_t(i)>\rho_t(j)\}} + \mathbb{1}_{\{\rho(i)>\rho(j),\rho_t(i)<\rho_t(j)\}}\Big]\Big)}{n(n-1)} \\
&= 1 - \frac{4\sum_{i=1}^{n}\sum_{j=i+1}^{n}\Big[\mathbb{1}_{\{\rho(i)<\rho(j),\rho_t(i)>\rho_t(j)\}} + \mathbb{1}_{\{\rho(i)>\rho(j),\rho_t(i)<\rho_t(j)\}}\Big]}{n(n-1)}.
\end{aligned} \tag{20}$$

In het volgende voorbeeld is te zien hoe $\tau(\rho, \rho_t)$ berekend wordt.

Voorbeeld 3.1 (Vervolg)**.** Beschouw de volgende rankings

$$\begin{aligned}
\rho &= [1, 2, 3], \\
\rho_t &= [3, 2, 1].
\end{aligned}$$

Er geldt, gebruikmakend van vergelijking (20),

$$\begin{aligned}
\tau(\rho, \rho_t) &= 1 - \frac{4(1+1+1)}{3 \cdot 2} \\
&= -1.
\end{aligned}$$

We beschouwen namelijk team 1 en team 2, waarbij 1 boven 2 in de pre-ranking staat. Echter, deze twee teams worden anders geranked in de toernooiranking, wat zorgt voor een 1 in de som van Σ. Hetzelfde geldt voor team 1 en team 3, evenals voor team 2 en team 3. Zo ontstaat de waarde -1 voor $\tau(\rho, \rho_t)$, omdat de toernooiranking precies de omgedraaide ranking is van de pre-ranking. Deze twee rankings hebben dus een negatieve correlatie van -1.

Als we de vergelijking van lemma 5.1 omschrijven, vinden we het volgende

$$\mathcal{T}(\rho) = 1 - \frac{4\mathcal{S}(\rho)}{n(n-1)m}.$$

Vervolgens kunnen we deze uitdrukking voor $\mathcal{T}(\rho)$ vergelijken met de uitdrukking voor $\tau(\rho, \rho_t)$ in vergelijking (20). We vermoeden dat er een relatie bestaat tussen $\mathcal{T}(\rho)$ en $\tau(\rho,\rho_t)$, aangezien de uitdrukkingen vrijwel overeenkomen. Deze relatie wordt gegeven in stelling 7.5.

6.2 Verdeling Σ

Het is belangrijk wat de significantie is van de gevonden $\tau(\rho, \rho_t)$, en daarmee dus de gevonden Σ. Doordat we alle mogelijke pre-rankings kunnen beschouwen, is het nodig te onderzoeken of de waarde van $\tau(\rho, \rho_t)$ met toeval is aangenomen. Daarom is het nodig om de verdeling van Σ te onderzoeken, waarmee de verdeling van $\tau(\rho, \rho_t)$ te vinden is. Dit beschrijft Kendall in het artikel [4], maar zijn uitleg is echter nogal kort en niet volledig. Daarom is deze paragraaf geschreven ter verduidelijking.

We zullen een vaste pre-ranking beschouwen, om conclusies te kunnen trekken over de berekening van Σ. Als we teams $1, \ldots, n$ beschouwen, bekijken we de *normale pre-ranking* $\rho = [1, 2, \ldots, n]$. Zij $\Sigma(n+1)$ de verzameling mogelijke waardes van Σ, in het geval van $n+1$ teams. Deze $\Sigma(n+1)$ is afhankelijk van $\Sigma(n)$. Stel namelijk dat we n teams beschouwen, de normale pre-ranking en een extra team, team $n+1$. Dan zal het extra team vooraan plaatsen een waarde $-n$ toevoegen aan Σ, omdat dit zorgt voor n inconsistenties, zie vergelijking (17). Het extra team tussen het eerste en het tweede team zal om dezelfde reden een waarde $-n+2$ toevoegen aan Σ. We vinden dus voor de verzameling toevoegingen

$$\mathcal{O}(n) = \{-n, -n+2, \ldots, n-2, n\}.$$

Deze verzameling $\mathcal{O}$ moet bij elk element uit $\Sigma(n)$ opgeteld worden om $\Sigma(n+1)$ te verkrijgen. We vinden $\Sigma(1), \ldots, \Sigma(4)$ met behulp van $\mathcal{O}(1), \mathcal{O}(2)$ en $\mathcal{O}(3)$, zoals te zien is in Tabel 3.

$\Sigma(1) = \{0\}$	$\mathcal{O}(1) = \{-1, 1\}$
$\Sigma(2) = \{-1, 1\}$	$\mathcal{O}(2) = \{-2, 0, 2\}$
$\Sigma(3) = \{-3, -1, 1, 3\}$	$\mathcal{O}(3) = \{-3, -1, 1, 3\}$
$\Sigma(4) = \{-6, -4, -2, 0, 2, 4, 6\}$	

Tabel 3: Waardes van Σ, bepaald met behulp van $\mathcal{O}$.

Hieruit kunnen we concluderen dat geldt

$$\Sigma(n) = \Big\{ -\frac{1}{2}n(n-1), -\frac{1}{2}n(n-1)+2, -\frac{1}{2}n(n-1)+4, \dots, \\ \frac{1}{2}n(n-1)-4, \frac{1}{2}n(n-1)-2, \frac{1}{2}n(n-1) \Big\}. \tag{21}$$

In Tabel 4 is schematisch te zien hoe de elementen van $\Sigma(n+1)$ ontstaan uit $\Sigma(n)$ door elk element uit $\mathcal{O}(n)$ aan elk element uit $\Sigma(n)$ toe te voegen.

$\mathcal{O}(n)$ \ $\Sigma(n)$	s_1	s_2	s_3	$\cdots$
$-n$	$s_1 - n$	$s_1 + 2 - n$	$s_1 + 4 - n$	
$-n+2$	$s_1 - n + 2$	$s_1 + 2 - n + 2$	$s_1 + 4 - n + 2$	
$-n+4$	$s_1 - n + 4$	$s_1 + 2 - n + 4$	$s_1 + 4 - n + 4$	
$\vdots$				

Tabel 4: Waardes van $\Sigma(n+1)$, verkregen uit $\Sigma(n)$, waarbij s_i het i-de element uit $\Sigma(n)$ representeert.

We zien dat elk element in Tabel 4 overeenkomt met het element dat daar rechts boven staat. Als we de rijen trapsgewijs noteren, verkrijgen we kolommen met dezelfde waardes, zoals te zien is in Tabel 5. Dit wordt aangegeven met $\searrow$. Alle elementen verkregen met behulp van s_1 staan bijvoorbeeld op de hoofddiagonaal.

$\mathcal{O}(n)$ \ $\Sigma(n)$	$s_1 \searrow$	$s_2 \searrow$	$s_3 \searrow$		$\cdots$
$-n$	$s_1 - n$	$s_1 + 2 - n$	$s_1 + 4 - n$		
$-n+2$		$s_1 + 2 - n$	$s_1 + 4 - n$	$s_1 + 6 - n$	
$-n+4$			$s_1 + 4 - n$	$s_1 + 6 - n$	$s_1 + 8 - n$
$\vdots$			$\ddots$	$\ddots$	$\ddots$

Tabel 5: Waardetabel van $\Sigma(n+1)$, verkregen uit $\Sigma(n)$.

We weten $\Sigma(1) = \{0\}$, omdat één team plaatsen geen ophef veroorzaakt in een bepaalde ranking. Verder geldt $\mathcal{O}(1) = \{-1, 1\}$, per definitie van de verzameling toevoegingen. Dus vinden we $\Sigma(2) = \{-1, 1\}$, waarbij beide waardes één keer voorkomen. We kunnen dit in een tabel zetten net zoals in Tabel 5, maar nu houden we ook de frequentie bij zoals te zien is in Tabel 6. Zo zien we in de tabel dat een diagonaal ontstaat, waarbij boven de diagonaal de waarde van het element in $\Sigma(n+1)$ staat en onder de lijn de frequentie van deze waarde. In de vijfde rij van de tabel staan dus de waardes van $\Sigma(3)$ met hun frequenties en in de negende rij de waardes van $\Sigma(4)$ met hun frequenties.

$\Sigma(2) \rightarrow$	$\mathcal{O} \downarrow$	-1	1					
	-2	-1 1	1 1					
	0		-1 1	1 1				
	2			-1 1	1 1			
$\Sigma(3) \rightarrow$	-3	-3 1	-1 2	1 2	3 1			
	-1		-3 1	3 2	3 2	3 1		
	1			-3 1	-1 2	1 2	3 1	
	3				-3 1	-1 2	1 2	3 1
$\Sigma(4) \rightarrow$		-6 1	-4 3	2 5	0 6	2 5	4 3	6 1

Tabel 6: Frequentie- en waardetabel van $\Sigma(n)$, $n \in \{3,4\}$.

Zij $x \in \mathcal{O}(n)$, $y \in \Sigma(n+1)$ en definieer $f(y)$ als frequentie van y. Met behulp van Tabel 6 kunnen we concluderen dat geldt

$$f(y) = \sum_x f(y-x).$$

Vervolgens kan een frequentietabel voor Σ worden gemaakt [4, p. 88]. Met behulp van deze tabel kunnen we de frequenties in een grafiek zetten welke we vervolgens verbinden met lijnen, een zogenaamd *frequentie polygoon*. Het blijkt dat dit frequentie polygoon sterk overeenkomt met de normale verdeling. Kendall geeft een bewijs voor de normaliteit van $\tau(\rho, \rho_t)$ voor n groot [4, p. 90-93], waarbij hij ook de standaard fout van $\tau(\rho, \rho_t)$ berekent. Uiteindelijk concludeert hij dat $\tau(\rho, \rho_t)$ goed te gebruiken is voor het berekenen van de correlatie tussen twee rankings.

7 Relatie Kendall's τ met toernooi-index en Slater's i

Een relatie tussen de eerder genoemde indices is moeilijk te vinden. Echter, er zijn een aantal voorwaarden of aannames waaronder bepaalde relaties bestaan tussen de indices. In het hoofdartikel wordt een relatie gegeven tussen Kendall's τ en de optimale toernooi-index $\mathcal{T}_0$ voor $m \geq 2$ en $n = 2$ [2, p. 9], namelijk

$$\tau(\rho, \rho_t) = \Big(1 + \frac{1}{m-1}\Big)\mathcal{T}_0^2 - \frac{1}{m-1}.$$

Echter, er wordt niet verwezen naar een bron waar deze relatie vandaan komt en de relatie wordt in het hoofdartikel niet bewezen. Omdat deze relatie alleen geldt in een paar specifieke gevallen, zijn we op zoek gegaan naar een meer algemene relatie. Helaas is deze niet gevonden, maar tijdens het onderzoek is er wel een relatie naar boven gekomen die geldt onder bepaalde voorwaarden. Deze relatie wordt met behulp van lemma 7.2 en lemma 7.4 bewezen in stelling 7.5.

7.1 De aannames

Voordat we de lemma's bespreken, moeten twee aannames gemaakt worden. Ten eerste moet gelden dat, indien twee teams beschouwd worden, één team *verslagen* moet worden door het andere team. Hierbij definiëren we dat team α team β verslaat door alle gespeelde wedstrijden tegen team β te winnen. Indien de teams m keer tegen elkaar spelen, geldt dus $a_{\alpha\beta} = m$. De eerste aanname is dat voor de elementen van de toernooimatrix A moet gelden

$$a_{ij} = 0 \Leftrightarrow a_{ji} \neq 0, \quad \forall i, j. \tag{22}$$

De tweede, en laatste, aanname die gemaakt wordt, is

$$\sum_{j=1}^{n} a_{ij} \neq \sum_{j=1}^{n} a_{kj}, \quad \forall i \neq k. \tag{23}$$

Deze aanname zorgt ervoor dat er een unieke toernooiranking bestaat bij het gespeelde toernooi, dus onafhankelijk van de nummering van de teams. De toernooiranking wordt namelijk gebaseerd op het aantal gewonnen wedstrijden per team, wat met deze aanname nooit gelijk is voor meerdere teams.

Opmerking 7.1. De aannames van vergelijking (22) en (23) gelden in heel hoofdstuk 7 en worden niet meer vermeld bij de lemma's en stellingen.

Naast de aannames hebben we een paar definities nodig. Beschouw daartoe teams $z_1, \ldots, z_n$ waarbij team z_k, met $1 \leq k \leq n$, staat voor het team dat $n-k$ keer een team verslagen heeft. Er geldt dus dat z_n geen één keer een ander team verslagen heeft, dus dit is de rij in de toernooimatrix die bestaat uit nullen. Dan ontstaat dus de volgende toernooiranking

$$\rho_t = [z_1, z_2, \ldots, z_n].$$

De laatste definitie die nodig is voor onderstaande lemma's, is die van een *permutatie op de toernooimatrix*. Bechouw hiervoor team z_k, $1 \leq l \leq n$, met z_k gerepresenteerd door rij l in de toernooimatrix. We willen rij l verplaatsen naar rij k, zodat de teams op volgorde van meest aantal gewonnen wedstrijden naar minst aantal gewonnen wedstrijden staan, in de rijen van de toernooimatrix. Definieer daarom de permutatie op de rijen als volgt,

$$\pi_{\text{rij}}(z_k) = (a_{lj} \;\; a_{kj}) \quad , 1 \leq j \leq n, \tag{24}$$

waarbij a_{ij} een element in de toernooimatrix A is [1]. Waar in rij k een 0 stond op a_{kk}, staat na de permutatie van de rijen een 0 in rij l op a_{lk}. Om de toernooimatrix zijn juiste vorm te laten houden, met nullen op de diagonaal, zullen vervolgens kolom l en k verwisseld moeten worden in de toernooimatrix. Definieer daarom de permutatie op de kolommen als volgt

$$\pi_{\text{kolom}}(z_k) = (a_{il} \;\; a_{ik}) \quad , 1 \leq i \leq n. \tag{25}$$

Lemma 7.2. *Beschouw toernooimatrix $A_{n\times n}$, een $n \times n$-matrix. Dan geldt dat $A^{\pi}_{n\times n}$ van de volgende vorm is*

$$A^{\pi}_{n\times n} = \begin{pmatrix} 0 & m & \cdots & m \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & m \\ 0 & \cdots & \cdots & 0 \end{pmatrix}, \tag{26}$$

waarbij voor permutatie π geldt

$$\pi = \pi_{kolom} \circ \pi_{rij}.$$

Bewijs. We bewijzen lemma 7.2 aan de hand van inductie naar het aantal teams van het toernooi, n. Zij $n_0 = 2$, de kleinste waarde die nodig is om te beschouwen. Beschouw vervolgens z_2, de rij met alleen maar nullen. Als geldt dat de rij met nullen in $A_{2\times 2}$, de

[1] Merk op dat $(a_{lj} \;\; a_{kj})$ in de definitie van π_{rij} staat voor een cykelnotatie, waarbij a_{lj} dus overgaat in a_{kj}. Hiermee worden a_{lj} en a_{kj} verwisseld.

representatie van z_2, op rij 2 staat, dan staat $A_{2\times 2}$ al in de vorm (26). Is dit niet het geval, dan is $A_{2\times 2}$ van de volgende vorm

$$A_{2\times 2} = \begin{pmatrix} 0 & 0 \\ m & 0 \end{pmatrix}.$$

Dan volgt per definitie van π_{rij}, in vergelijking (24),

$$A^{\pi_{\text{rij}}}_{2\times 2} = \begin{pmatrix} m & 0 \\ 0 & 0 \end{pmatrix},$$

en per definitie van π en π_{kolom}, in vergelijking (25),

$$A^{\pi}_{2\times 2} = \begin{pmatrix} 0 & m \\ 0 & 0 \end{pmatrix}.$$

Dus het lemma is waar voor een 2×2-toernooimatrix A.

Zij $n_0 < N < n$ willekeurig en neem aan dat het lemma geldt voor alle indices tot en met N (inductieveronderstelling). Beschouw nu $N+1$ teams. Stel dat z_{N+1}, de rij bestaande uit nullen, op de l-de rij van matrix $A_{(N+1)\times(N+1)}$ staat. Als $l = N+1$, dan staat de rij met nullen al op de juiste plaats, zoals in (26). Indien dit niet het geval is, passen we de permutatie π toe op de l-de rij in $A_{(N+1)\times(N+1)}$, waardoor volgt

$$A^{\pi}_{(N+1)\times(N+1)} = \begin{pmatrix} & & m \\ & A_{N\times N} & \vdots \\ & & m \\ 0 & \dots & 0 \end{pmatrix}.$$

We zien dat de laatste kolom bestaat uit elementen m, behalve op plaats $a_{(N+1)(N+1)}$. Doordat rij l team $z_{(N+1)}$ representeert, staat kolom l dus voor het aantal wedstrijden dat van team $z_{(N+1)}$ gewonnen is. Per definitie heeft team $z_{(N+1)}$ alle wedstrijden verloren, dus bestaat kolom l uit elementen m, behalve op plaats a_{ll}. Na de permutatie van de kolommen, komt de kolom met m-en in de $(N+1)$-de kolom te staan.

Dankzij de inductieveronderstelling weten we dat $A_{N\times N}$ om te schrijven is naar $A^{\pi}_{N\times N}$. Hieruit kunnen we concluderen dat $A^{\pi}_{(N+1)\times(N+1)}$ in de gevraagde vorm (26) staat. Omdat N willekeurig gekozen is, geldt dit dus voor alle $N \in \{3, \dots, n-1\}$. □

Opmerking 7.3. Het is belangrijk dat de structuur van het toernooi gelijk blijft. Hiermee wordt bedoeld dat als team i voor team j op de toernooiranking stond, dit ook het geval moet zijn na het toepassen van de permutatie op A. Stel dus

$$\rho_t(i) < \rho_t(j),$$

dan volgt per definitie van π_{rij} en π_{kolom} dat geldt

$$\pi_{\rho_t} = \begin{pmatrix} 1 & 2 & \dots & n \\ \rho_t(1) & \rho_t(2) & \dots & \rho_t(n) \end{pmatrix}.$$

We zien dat na het toepassen van de permutatie de toernooiranking niet verandert. Dit wordt veroorzaakt, doordat de nummering van de teams wordt aangepast met behulp van de toernooiranking.

7.2 Relatie Kendall's τ met de toernooi-index en $\mathcal{S}(\rho)$ onder aannames

De vorm A^π, zoals beschreven in bovenstaand lemma, is een speciaal geval in de toernooimatrices. Dit wordt veroorzaakt, doordat er een unieke toernooiranking bestaat, onafhankelijk van de nummering van de teams. Hierdoor kunnen we een relatie afleiden tussen $\mathcal{S}(\rho)$ en $\tau(\rho, \rho_t)$. Merk hierbij op dat n_d gebruikt wordt in $\tau(\rho, \rho_t)$.

Lemma 7.4. *Beschouw een toernooimatrix A. Dan geldt*

$$\frac{N(1 - \mathcal{T}(\rho))}{2} = \mathcal{S}(\rho) = m \cdot n_d.$$

Bewijs. Met lemma 5.1 geldt de linker gelijkheid. Zij A een toernooimatrix die voldoet aan (22) en (23). Met lemma 7.2 volgt dat we A om kunnen schrijven naar

$$A^\pi = \begin{pmatrix} 0 & m & \cdots & m \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & m \\ 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

Noem $A^\pi = A$. Dan volgt

$$a_{kl} = 0, \quad 1 \leq l \leq k \leq n,$$

dus geldt

$$\sum_{\rho(j)<\rho(k)} a_{kj} = 0, \quad 1 \leq j \leq k. \tag{27}$$

Dan kunnen we het volgende concluderen

$$\begin{aligned} \mathcal{S}(\rho) &\overset{(11)}{=} \sum_{i=1}^{n} \sum_{\rho(j)<\rho(i)} a_{ij} \\ &= \sum_{\rho(j)<\rho(1)} a_{1j} + \sum_{\rho(j)<\rho(2)} a_{2j} + \cdots + \sum_{\rho(j)<\rho(n)} a_{nj} \end{aligned}$$

$$
\begin{aligned}
&\overset{(27)}{=} \sum_{\substack{\rho(j)<\rho(1)\\ 1<j\leq n}} a_{1j} + \sum_{\substack{\rho(j)<\rho(2)\\ 2<j\leq n}} a_{2j} + \cdots + \sum_{\substack{\rho(j)<\rho(n)\\ n<j\leq n}} a_{nj} \\
&= \sum_{j=2}^{n} m \cdot \mathbb{1}_{\{\rho(j)<\rho(1)\}} + \sum_{j=3}^{n} m \cdot \mathbb{1}_{\{\rho(j)<\rho(2)\}} + \cdots + m \cdot \mathbb{1}_{\{\rho(n)<\rho(n-1)\}} \\
&= m \cdot \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbb{1}_{\{\rho(j)<\rho(i)\}}.
\end{aligned}
$$

Omdat we n_d willen vergelijken met $\mathcal{S}(\rho)$, zullen we deze omschrijven. Per definitie geldt

$$
n_d = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\mathbb{1}_{\{\rho(i)<\rho(j), \rho_t(i)>\rho_t(j)\}} + \mathbb{1}_{\{\rho(i)>\rho(j), \rho_t(i)<\rho_t(j)\}} \right].
$$

Dankzij de gekozen vorm van A, kunnen we de volgende normale toernooiranking gebruiken

$$
\rho_t = [1, 2, 3, \cdots, n].
$$

Omdat de toernooiranking vaststaat, geldt voor het aantal inconsistenties tussen de preranking en toernooiranking dus het volgende

$$
\begin{aligned}
n_d &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbb{1}_{\{\rho(i)>\rho(j), \rho_t(i)<\rho_t(j)\}} \\
&= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbb{1}_{\{\rho(i)>\rho(j)\}}.
\end{aligned}
$$

Hiermee is bewezen dat de rechter gelijkheid geldt, dus

$$
\mathcal{S}(\rho) = m \cdot n_d.
$$

□

Met behulp van onderstaand voorbeeld wordt duidelijk waarom de aanname in vergelijking (22) noodzakelijk is voor lemma 7.2 en dus ook voor de bewezen relatie.

Voorbeeld 7.1. Stel dat bij geen enkel paar van de teams één team het andere team heeft verslagen in het toernooi. Bij drie teams die drie keer tegen elk ander team spelen kan de toernooimatrix er als volgt uit zien

$$
A = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 2 & 0 \end{pmatrix}.
$$

Door naar het aantal gewonnen wedstrijden per team te kijken, komen we tot de volgende toernooiranking

$$\rho_t = [3, 1, 2].$$

Zij de pre-ranking als volgt

$$\rho = [1, 2, 3].$$

Het aantal inconsistenties in de pre-ranking kunnen we als volgt berekenen

$$\begin{aligned}\mathcal{S}(\rho) &= a_{21} + a_{31} + a_{32} \\ &= 1 + 2 + 2 \\ &= 5.\end{aligned}$$

Als we naar het aantal inconsistenties n_d kijken, zien we dat het paar (1,2) in beide rankings hetzelfde staat, maar het paar (1,3) staat in de toernooiranking andersom en (2,3) ook. Hierdoor kunnen we concluderen dat er twee inconsistenties zijn, dus

$$m \cdot n_d = 3 \cdot 2 = 6.$$

Lemma 7.2 geldt in dit geval dus niet, omdat n_d geen rekening houdt met hoe sterk de inconsistentie is. $\mathcal{S}(\rho)$ telt als het ware het aantal schendingen, dus onterecht gewonnen wedstrijden, waar n_d alleen de inconsistenties telt in de rankings zelf.

In het volgende voorbeeld worden de andere aannname van lemma (7.2) behandeld.

Voorbeeld 7.2. Stel dat er bij het gegeven toernooi geen unieke toernooiranking te maken is, hierbij is dus niet uniek te bepalen welk team "sterker" is dan de rest. Beschouw hiervoor vier teams die elk één wedstrijd tegen elkaar spelen en beschouw daarbij de volgende toernooimatrix

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

We zien dat team 1 en 2 beide twee wedstrijden gewonnen hebben en team 3 en 4 beide één wedstrijd. Hierdoor is een unieke toernooiranking niet te bepalen. De toernooiranking wordt dus afhankelijk van de nummering van de teams, dus

$$\rho_t = [1, 2, 3, 4].$$

Zij de pre-ranking als volgt

$$\rho = [1, 2, 3, 4].$$

Dan vinden we

$$\begin{aligned}\mathcal{S}(\rho) &= a_{21} + a_{31} + a_{32} + a_{41} + a_{42} + a_{43} \\ &= 0 + 0 + 0 + 1 + 0 + 0 + 0 \\ &= 1.\end{aligned}$$

Het is echter makkelijk in te zien dat geldt

$$n_d = 0,$$

aangezien de pre-ranking en toernooiranking gelijk gekozen zijn. Ook in dit voorbeeld zien we dus dat de aanname van een mogelijk unieke toernooiranking nodig is voor het geldig zijn van lemma 7.2. Dit heeft te maken met de manier waarop de inconsistenties geteld worden in het berekenen $\mathcal{S}(\rho)$ en n_d. In $\mathcal{S}(\rho)$ wordt meegerekend dat team 4 één keer gewonnen heeft van team 1, terwijl team 1 boven team 4 op de pre-ranking staat.

Dan kunnen we nu de relatie tussen de toernooi-index $\mathcal{T}(\rho)$, het aantal inconsistenties $\mathcal{S}(\rho)$ en Kendall's τ samenvatten in een stelling.

Stelling 7.5. *Beschouw een toernooimatrix A. Zij ρ een willekeurige pre-ranking. Dan geldt*

$$\tau(\rho, \rho_t) = \mathcal{T}(\rho) = 1 - \frac{2\mathcal{S}(\rho)}{N}.$$

Bewijs. Uit lemma 5.1 volgt direct de rechter gelijkheid. Daarnaast volgt uit lemma 7.4 dat geldt $\mathcal{S}(\rho) = m \cdot n_d$, waardoor het volgende geldt

$$\begin{aligned}\tau(\rho, \rho_t) &\overset{(17)}{=} \frac{n_c - n_d}{n(n-1)/2} \\ &\overset{(18)}{=} \frac{2\left(\frac{1}{2}n(n-1) - n_d\right)}{n(n-1)} - \frac{2n_d}{n(n-1)} \\ &= \frac{n(n-1)}{n(n-1)} - \frac{4n_d}{n(n-1)} \\ &= 1 - \frac{4n_d}{n(n-1)}.\end{aligned}$$

Wegens lemma 5.1 weten we dat geldt

$$\begin{aligned}\mathcal{T}(\rho) &= 1 - \frac{2\mathcal{S}(\rho)}{N} \\ &= 1 - \frac{4\mathcal{S}(\rho)}{n(n-1)m}.\end{aligned}$$

Vervolgens geldt met behulp van lemma 7.4

$$\mathcal{T}(\rho) = 1 - \frac{4n_d}{n(n-1)}.$$

We zien dat de linker gelijkheid ook geldt, dus

$$\tau(\rho, \rho_t) = \mathcal{T}(\rho).$$

$\square$

8 Gevolgen relatie indices

De indices zijn aan elkaar gerelateerd, waardoor we meer inzicht hebben verkregen in de betekenis van de indices. Eén van deze inzichten is de relatie van de toernooiranking en de optimale toernooi-index. Waar we eerst de toernooi-index zagen als een index die aangeeft in hoeverre de uitslag van een gespeeld toernooi overeenkomt met een gegeven pre-ranking, kunnen we met behulp van stelling 7.5 deze index anders beschouwen. De toernooi-index telt, bij nader inzien, het aantal consistenties waar het aantal inconsistenties vanaf wordt getrokken. Deze manier van benaderen is terug te vinden in definitie (3). We zien in definitie (2) dat v_i een negatieve waarde toekent aan een wedstrijd als een team wint van een team dat hoger op de pre-ranking staat. Bij nader inzien kunnen we dit tellen als een inconsistentie. Hieruit kunnen we concluderen dat de toernooi-index optimaal is als het aantal inconsistenties minimaal is. Dit is het geval als de pre-ranking gelijk is aan de toernooiranking, zoals beschreven en bewezen wordt in de volgende stelling.

Stelling 8.1. *Beschouw een toernooimatrix A die voldoet aan de aannames van vergelijkingen* (22) *en* (23)*. Dan weten we dat de toernooiranking de toernooi-index maximaliseert, dus*

$$1 - \frac{2\mathcal{S}_i}{N} = \mathcal{T}_0 = \mathcal{T}(\rho_t).$$

Bewijs. De linker gelijkheid volgt direct uit stelling 5.2. Zij ρ_t de toernooiranking bij matrix A, gegeven zoals in de stelling. Dan volgt, met behulp van stelling 5.2,

$$\begin{aligned}\mathcal{T}_0 &= 1 - \frac{2\mathcal{S}_i}{N} \\ &\overset{(12)}{=} 1 - \frac{2 \min\limits_{\rho} \mathcal{S}(\rho)}{n(n-1)m}.\end{aligned}$$

Vervolgens volgt met lemma 7.4

$$\mathcal{T}_0 = 1 - \frac{2m \min\limits_{\rho} n_d(\rho)}{n(n-1)m}.$$

Daarnaast weten we

$$\begin{aligned}n_d(\rho_t) &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\mathbb{1}_{\{\rho_t(i)<\rho_t(j), \rho_t(i)>\rho_t(j)\}} + \mathbb{1}_{\{\rho_t(i)>\rho_t(j), \rho_t(i)<\rho_t(j)\}}\right] \\ &= 0.\end{aligned}$$

We zien dus dat het aantal inconsistenties van de pre-ranking ten opzichte van de toernooiranking minimaal is als we de pre-ranking gelijk nemen aan de toernooiranking.

Intuïtief gezien is dit ook duidelijk, aangezien er dan geen inconsistenties zijn. Hierdoor geldt

$$\mathcal{T}_0 = 1 - \frac{2\mathcal{S}(\rho_t)}{n(n-1)m},$$

en met lemma 5.1 volgt

$$\mathcal{T}_0 = \mathcal{T}(\rho_t).$$

□

Intuïtief is het duidelijk dat de toernooiranking als pre-ranking de waarde 1 geeft aan de optimale toernooi-index. Dit volgt direct uit stelling 8.1, waar wordt opgemerkt dat geldt

$$n_d(\rho_t) = 0.$$

Gevolg 8.2. Onder de aannames van vergelijkingen (22) en (23) volgt, met behulp van lemma 7.4 en stelling 8.1, dat geldt

$$\begin{aligned}\mathcal{T}(\rho_t) &= \mathcal{T}_0 \\ &= 1 - \frac{2m \min\limits_{\rho} n_d(\rho)}{n(n-1)m} \\ &= 1.\end{aligned}$$

Daarnaast geldt voor een gegeven pre-ranking, dat $\mathcal{T}(\rho)$ en $\mathcal{S}(\rho)$ afhankelijk zijn van elkaar, zie lemma 5.1. Beide indices geven dus, achteraf gezien, dezelfde conclusies over het al dan niet competitief zijn van een gespeeld toernooi. Vervolgens hebben we in stelling 7.5 gezien dat onder de aannames (22) en (23) een relatie bestaat tussen $\mathcal{T}(\rho)$, $\mathcal{S}(\rho)$ en $\tau(\rho, \rho_t)$. Dit geeft aan dat, onder de aannames, de drie indices dezelfde conclusies geven over het toernooi. Hierdoor kan men de keuze maken om de index te gebruiken die de minste tijd kost om te berekenen.

9 Overige indices

Naast de indices die met elkaar samenhangen, bestaan er indices die onafhankelijk zijn van deze eerder genoemde indices. De onafhankelijke indices geven informatie over het competitief zijn van het gespeelde toernooi. Ze kunnen gebruikt worden ter controle van één van de eerder genoemde indices.

9.1 Competitief evenwicht

De eerste index die we beschouwen, onafhankelijk van $\mathcal{T}(\rho)$, $\mathcal{S}(\rho)$ en $\tau(\rho, \rho_t)$, is het *competitief evenwicht* σ_c [2]. Deze wordt als volgt gedefinieerd

$$\sigma_c = \sqrt{\frac{\sum_{i=1}^{n}(w_i - \frac{1}{2})^2}{n}}, \tag{28}$$

waarbij w_i gedefinieerd is zoals in vergelijking (1). Per team wordt de afwijking van het evenwicht berekend, waarbij er een evenwicht is als het aantal wedstrijden dat het team gewonnen heeft gelijk is aan het aantal wedstrijden dat het team verloren heeft. Vervolgens wordt voor het gehele toernooi de afwijking van het evenwicht bepaald door middel van de gemiddelde afwijking van alle teams. We kunnen een sporttoernooi als competitief beschouwen als het competitief evenwicht dicht bij 0 ligt. In dit geval zullen de teams, gemiddeld genomen, evenveel wedstrijden verliezen als winnen, waardoor alle teams als gelijk beschouwd kunnen worden. Hierdoor is het voor een wedstrijd niet te voorspellen welk team zal winnen.

Voorbeeld 3.1 (Vervolg)**.** Er geldt

$$w_1 = 0, \quad w_2 = \frac{3}{4}, \quad w_3 = \frac{3}{4}.$$

Dus vinden we

$$\begin{aligned} \sigma_c &= \sqrt{\frac{\frac{1}{4} + \frac{1}{16} + \frac{1}{16}}{3}} \\ &= \sqrt{\frac{1}{8}} \\ &\approx 0.3536. \end{aligned}$$

Herinner de waarde van de optimale toernooi-index $\mathcal{T}_0 = \frac{2}{3}$, waaruit te concluderen was dat het toernooi enigszins competitief is. We zien dat σ_c dichter bij 0 dan bij 1 ligt, waaruit te concluderen is dat het gespeelde toernooi eerder als competitief dan als niet competitief beschouwd kan worden. We kunnen ook concluderen dat σ_c geen strafpunten meegeeft in het geval dat een "zwakker" team wint van een "sterker" team, wat wel het geval is bij de optimale toernooi-index. Hierdoor geeft σ_c een waarde dichter bij 0 dan $\mathcal{T}_0$, waarmee σ_c het toernooi eerder als competitief zal aanduiden.

9.2 Gewogen toernooi-index

Naast de eerder genoemde indices, kan er ook gekeken worden naar een *gewogen toernooi-index* $\mathcal{W}(w)$ [2]. Hierbij is w een vector bestaande uit elementen w_i, het gewicht behorend bij team i, $1 \leq i \leq n$. Een voorbeeld van het geven van gewichten is door te kijken naar voorgaande toernooien, waarbij meer gewicht gegeven wordt aan een team dat in de al gespeelde toernooien vaak gewonnen heeft. Met behulp van deze gewichten kunnen we de toernooi-index (4) herschrijven tot

$$\mathcal{W}(w) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_i - w_j) a_{ij}, \tag{29}$$

waarbij

$$N = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}.$$

Er worden strafpunten meegegeven in $\mathcal{W}(w)$ in het geval dat een team met een lager gewicht w wint van een team met een hoger gewicht w, wat we weer kunnen zien als een inconsistentie. Deze inconsistentie krijgt echter niet de waarde 1, zoals in eerder genoemde indices, maar het verschil van de gewichten van de teams. Hierdoor wordt een hoger gewicht toegekend aan een inconsistentie als de gewichten van de teams verder uit elkaar liggen, waardoor het competitief zijn van een toernooi beter beoordeeld wordt.

10 Dynamische ranking

Zoals in de vorige hoofdstukken duidelijk is geworden, is het bepalen van een (juiste) pre-ranking van groot belang. Tot nu hebben we rekening gehouden met alle mogelijke pre-rankings om zo algemeen mogelijk conclusies te kunnen trekken. In dit hoofdstuk wordt een methode behandeld om een realistische pre-ranking te bepalen, gebaseerd op een artikel geschreven door Motegi en Masuda (2012) [6]. Hierbij wordt gebruik gemaakt van de win-lose score van Park en Newman.

10.1 Originele win-verlies score, Park & Newman

Park en Newman baseren de pre-ranking voor een toernooi op directe en indirecte winsten van de teams [6][7]. Hierbij wordt gebruik gemaakt van een matrix B, de getransponeerde matrix van de toernooimatrix,

$$B = A^\top. \tag{30}$$

Voor teams i en j, $1 \leq i, j \leq n$, geldt dus dat b_{ij} staat voor het aantal directe winsten van j tegen i. Daarnaast wordt een constante α gedefinieerd, met $0 \leq \alpha < 1$, als gewicht dat meegegeven wordt aan een indirecte winst. In het geval dat i wint van j en vervolgens wint j van k, dan ontvangt i naast 1 punt door te winnen van j ook een punt α, doordat j heeft gewonnen van k. Als vervolgens k wint van een ander team l, ontvangt i nog een score α^2 voor deze indirecte winst. Zo ontstaat de volgende *winmatrix* W

$$\begin{aligned} W &= B + \alpha B^2 + \alpha^2 B^3 + \cdots \\ &= B(I + \alpha B + \alpha^2 B^2 + \cdots) \\ &= B(I - \alpha B)^{-1}. \end{aligned} \tag{31}$$

Beschouw de term $\alpha^2 B^2$ en beschouw $\alpha^2 b_{ij}$. Dan geldt, per definitie van B^2,

$$b_{ij} = \sum_{k=1}^{n} b_{ik} b_{kj}.$$

Hieruit kunnen we concluderen dat b_{ij} staat voor het aantal keer dat team j heeft gewonnen van team i via team k. Als we op deze manier naar alle termen van W kijken, zien we dat w_{ij} staat voor de win score van team j, verkregen door directe en indirecte winsten, tegen team i. De *win score* van team j wordt gegeven door $\mathbf{w}(j)$, waarbij

$$\begin{aligned} \mathbf{w} &= W^\top 1 \\ &= (I - \alpha B^\top)^{-1} B^\top 1 \\ &= (I - \alpha A)^{-1} A 1, \end{aligned} \tag{32}$$

met $\mathbb{1} = (1, 1, \dots, 1)^\top$. Op dezelfde manier kunnen we de *verlies score* van team j bepalen, door $B = A^\top$ te vervangen door A. Hieruit volgt

$$\begin{aligned}\mathbf{l} &= W\mathbb{1} \\ &= (I - \alpha B)^{-1} B\mathbb{1} \\ &= (I - \alpha A^\top)^{-1} A^\top \mathbb{1}.\end{aligned} \tag{33}$$

Vervolgens wordt de *win-verlies score* per team gegeven in de volgende vector

$$\mathbf{s} = \mathbf{w} - \mathbf{l}, \tag{34}$$

waar de pre-ranking op gebaseerd wordt.

10.2 Dynamische win-verlies score

Het bepalen van de pre-ranking wordt gebaseerd op directe en indirecte winsten, zoals te zien is in de vorige paragraaf. Echter, de pre-ranking is dan enkel gebaseerd op winsten zonder rekening te houden met het tijdstip waarop deze winsten behaald zijn. Het is voor te stellen dat een team een kleiner gewicht aan punten hoort te krijgen van een indirecte winst, behaald door een wedstrijd van een paar maanden geleden of zelfs een jaar geleden. Daarom introduceren Motegi en Masuda de *dynamische pre-ranking*, die rekening houdt met het tijdstip waarop een winst behaald wordt en het aantal winsten [6]. We beschouwen de win score op tijdstip t_s, met $1 \leq s \leq n_{\max}$. De matrix B_{t_s} staat voor de directe winsten behaald op tijdstip t_s. Vervolgens worden de directe en indirecte winsten bepaald van een vorig tijdstip t_{s-1}. Deze directe en indirecte winsten worden vermenigvuldigd met een factor $e^{-\beta(t_s - t_{s-1})}$, omdat de winsten, behaald op tijdstip t_{s-1}, op tijdstip t_s minder waard zijn dan de winsten behaald op tijdstip t_s. De directe en indirecte winsten worden als volgt berekend

$$\sum_{m_s \in \{0,1\}} \alpha^{m_s} B_{t_{s-1}} B_{t_s}^{m_s} = B_{t_{s-1}} + \alpha B_{t_{s-1}} B_{t_s}.$$

De eerste term geeft de directe winsten op tijdstip t_{s-1} en de tweede term de indirecte winsten, door te kijken naar directe winsten op tijdstip t_s en directe winsten op t_{s-1}. Zo ontstaat de volgende *dynamische winmatrix* ter vervanging van de winmatrix van vergelijking (31)

$$
\begin{aligned}
W_{t_s} =& B_{t_s} + e^{-\beta(t_s - t_{s-1})} \sum_{m_s \in \{0,1\}} \alpha^{m_s} B_{t_{s-1}} B_{t_s}^{m_s} \\
&+ e^{-\beta(t_s - t_{s-2})} \sum_{m_{s-1}, m_s \in \{0,1\}} \alpha^{m_{s-1}+m_s} B_{t_{s-2}} B_{t_{s-1}}^{m_{s-1}} B_{t_s}^{m_s} \\
&+ \cdots + e^{-\beta(t_s - t_1)} \sum_{m_2, \ldots, m_s \in \{0,1\}} \alpha^{\sum_{i=1}^{s} m_i} B_{t_1} B_{t_2}^{m_2} \cdots B_{t_s}^{m_s}.
\end{aligned} \tag{35}
$$

Verder geldt

$$
\begin{aligned}
e^{-\beta(t_s - t_{s-2})} &= e^{-\beta(t_s - t_{s-1} + t_{s-1} - t_{s-2})} \\
&= e^{-\beta(t_s - t_{s-1})} e^{-\beta(t_{s-1} - t_{s-2})},
\end{aligned}
$$

waarmee vergelijking (36) als volgt wordt omgeschreven

$$
\begin{aligned}
W_{t_s} =& B_{t_s} + e^{-\beta(t_s - t_{s-1})} \sum_{m_s \in \{0,1\}} \alpha^{m_s} B_{t_s}^{m_s} \\
& \Big[B_{t_{s-1}} + e^{-\beta(t_{s-1} - t_{s-2})} \sum_{m_{s-1} \in \{0,1\}} \alpha^{m_{s-1}} B_{t_{s-2}} B_{t_{s-1}}^{m_{s-1}} \\
& + \cdots + e^{-\beta(t_{s-1} - t_1)} \sum_{m_2, \ldots, m_{s-1} \in \{0,1\}} \alpha^{\sum_{i=2}^{s-1} m_i} B_{t_1} B_{t_2}^{m_2} \cdots B_{t_{s-1}}^{m_{s-1}} \Big] \\
=& B_{t_s} + e^{-\beta(t_s - t_{s-1})} W_{t_{s-1}} (\alpha^0 B_{t_s} + \alpha B_{t_s}) \\
=& B_{t_s} + e^{-\beta(t_s - t_{s-1})} W_{t_{s-1}} (I + \alpha B_{t_s}).
\end{aligned} \tag{36}
$$

Te zien is dat de dynamische winmatrix op tijdstip t_s afhankelijk is van de dynamische winmatrix op het vorige tijdstip. Daardoor is een beginvoorwaarde van belang. Op tijdstip t_1 beschouwen we voor het eerst de uitslag van de gespeelde wedstrijden en hoeft er dus geen rekening gehouden te worden met indirecte winsten, omdat deze simpelweg nog niet bestaan. Daarom definiëren we de *dynamische win score*, ter vervanging van de win score van vergelijking (32), als volgt

$$
\begin{aligned}
\mathbf{w}_{t_s} &= \begin{cases} B_{t_s}^\top 1 & ,\text{als } s = 1, \\ B_{t_s}^\top 1 + e^{-\beta(t_s - t_{s-1})} \mathbf{w}_{t_{s-1}} (I + \alpha B_{t_s}^\top) & ,\text{als } s > 1, \end{cases} \\
&\overset{(30)}{=} \begin{cases} A_{t_s} 1 & ,\text{als } s = 1, \\ A_{t_s} 1 + e^{-\beta(t_s - t_{s-1})} \mathbf{w}_{t_{s-1}} (I + \alpha A_{t_s}) & ,\text{als } s > 1. \end{cases}
\end{aligned}
$$

Vervolgens ontstaat de definitie van de *dynamische verlies score*, ter vervanging van de verlies score van vergelijking (33), door $B = A^\top$ te vervangen door A

$$
\mathbf{l}_{t_s} = \begin{cases} A_{t_s}^\top 1 & ,\text{als } s = 1, \\ A_{t_s}^\top 1 + e^{-\beta(t_s - t_{s-1})} \mathbf{w}_{t_{s-1}} (I + \alpha A_{t_s}^\top) & ,\text{als } s > 1. \end{cases}
$$

Tot slot wordt de *dynamische win-verlies score*, als vervanging voor de win-verlies score uit vergelijking (34), gedefinieerd als

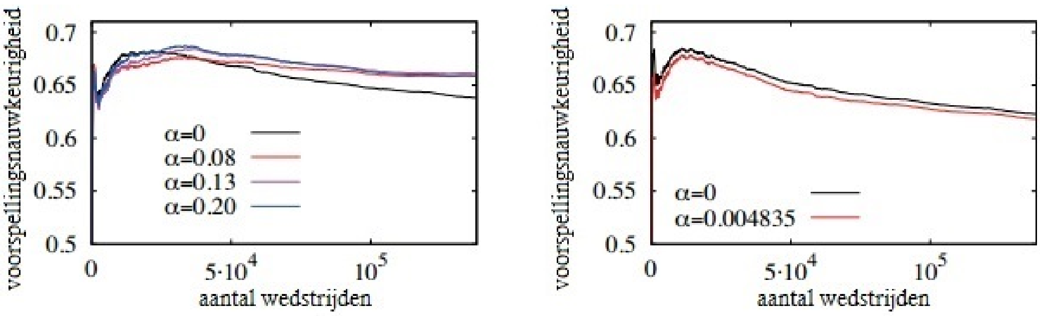
$$\mathbf{s}_{t_s} = \mathbf{w}_{t_s} - \mathbf{l}_{t_s}.$$

10.3 Resultaten

De win-verlies score op tijdstip t_{s+1} is, zoals eerder gezien, afhankelijk van de win-verlies scores van de voorgaande tijdstippen. Als op tijdstip t_s alle win-verlies scores van de teams bekend zijn, kan met behulp van de dynamische ranking de win-verlies score op het volgende tijdstip voorspeld worden [6]. In het geval dat twee teams beschouwd worden waarbij de teams een gelijke score hebben tot en met tijdstip t_s, wordt er vanuit gegaan dat de wedstrijd op t_{s+1} op een gelijkspel uitloopt. Indien de voorspelling op t_s niet blijkt te kloppen, wordt een schending waargenomen. Om vervolgens te bepalen of het toernooi competitief is, wordt gebruik gemaakt van een *voorspellingsnauwkeurigheid* Pr. Dit wordt gedefinieerd als de fractie goed voorspelde wedstrijden, dus

$$Pr = \frac{N_{Pr} - s - g}{N_{Pr} - g},$$

waarbij N_{Pr} staat voor het aantal voorspelde wedstrijden, s het aantal schendingen en g het aantal wedstrijden dat in gelijkspel eindigde. Deze Pr is berekend voor verschillende waardes van α en β, in vergelijkingen (31) en (36), met behulp van data verkregen van een tennistoernooi [6, p. 11]. Daarbij zijn de volgende grafieken gemaakt met behulp van numerieke simulaties [6, p. 5].



Figuur 1: De voorspellingsnauwkeurigheid berekend met behulp van de dynamische win-verlies score (links), met $\beta = 1/365$, en de originele win-verlies score (rechts).

Uit de linker grafiek, bepaald met behulp van de dynamische win-verlies score, is te concluderen dat de voorspellingsnauwkeurigheid voor $\alpha = 0.13$ het grootst is, met uitzondering voor een klein aantal wedstrijden. Ook is te zien dat de voorspelling enigszins ongevoelig is voor α, als $0.08 \leq \alpha \leq 0.20$. Daarnaast werd de voorspellingsnauwkeurigheid onderzocht voor verschillende waardes voor β, waarbij men tot de conclusie kwam dat de uitkomsten vrijwel volledig overeenkwamen voor $0 \leq \beta \leq 1/365$. Hierdoor werd een robuustheid geconstateerd ten opzichte van α en β. Deze robuustheid werd vervolgens onderzocht door de correlatie tussen de scores voor verschillende waardes van α en β te berekenen. Dit werd gedaan door Kendall's τ toe te passen op de top k lijst van de tennissers [6, p. 6-8].

In de rechter grafiek, bepaald met behulp van de originele win-verlies score, is te zien dat de voorspellingsnauwkeurigheid voor $\alpha = 0$ hoger is dan voor $\alpha = 0.004835$, hoewel dit weinig scheelt. De voorspellingsnauwkeurigheid is onderzocht voor verschillende waardes van α, waaruit geconcludeerd werd dat de nauwkeurigheid monotoon daalt in α. Daarnaast is te zien dat de nauwkeurigheid voor $\alpha = 0$, bepaald met behulp van de originele win-verlies score, minder goed is dan voor $\alpha = 0$ en $0.08 \leq \alpha \leq 0.2$ door gebruik te maken van de dynamische win-verlies score.

Uit dit onderzoek kan geconcludeerd worden dat de dynamische win-verlies score voor een betere voorspelling zorgt dan de originele win-verlies score, aangezien rekening wordt gehouden met het tijdstip waarop een winst van een wedstrijd plaatsvindt. Daarnaast kan de methode om de win-verlies score dynamisch te maken ook gebruikt worden als benadering voor het beoordelen van de sterkte van de bestudeerde teams, fluctuerend over de tijd.

11 Conclusie en vervolgonderzoek

We hebben verschillende indices behandeld die elk informatie geven over de mate van onvoorspelbaarheid van een toernooi, dus de competitiviteit van een toernooi. Zo is voor de optimale toernooi-index bewezen dat geldt

$$\mathcal{T}_0 \neq -1,$$

en voor de toernooi-index dat geldt

$$-\mathcal{T}_0 \leq \mathcal{T}(\rho) \leq \mathcal{T}_0.$$

Daarnaast is tussen de toernooi-index $\mathcal{T}(\rho)$ en $\mathcal{S}(\rho)$ een relatie gevonden en bewezen. Hieruit volgde direct de volgende relatie tussen de optimale toernooi-index, $\mathcal{T}_0$, en Slater's i, $\mathcal{S}_i$,

$$N - 2\mathcal{S}_i = N\mathcal{T}_0.$$

Vervolgens zijn we op zoek gegaan naar een relatie tussen Kendall's $\tau(\rho, \rho_t)$ en de eerder genoemde indices. Deze is helaas alleen bewezen onder bepaalde aannames, namelijk de aanname dat tussen twee teams moet gelden dat één team het andere team moet verslaan en de aanname dat de toernooiranking uniek te bepalen moet zijn. Een belangrijke relatie die werd gevonden, is

$$\frac{N(1 - \mathcal{T}(\rho))}{2} = \mathcal{S}(\rho) = m \cdot n_d.$$

Hiermee kon een relatie tussen Kendall's $\tau(\rho, \rho_t)$, de toernooi-index en $\mathcal{S}(\rho)$ bewezen worden, onder de gemaakte aannames, namelijk

$$\tau(\rho, \rho_t) = \mathcal{T}(\rho) = 1 - \frac{2\mathcal{S}(\rho)}{N}.$$

Met deze relatie werd een intuïtief duidelijke stelling bewezen, namelijk dat de toernooiranking als optimale pre-ranking de toernooi-index maximaliseert. Deze stelling is alleen bewezen onder de aannames, maar lijkt ook te gelden zonder de aannames. Dit zou bewezen kunnen worden in een vervolgonderzoek, door gebruik te maken van een eventueel meer algemene relatie tusssen $\tau(\rho, \rho_t)$, $\mathcal{T}(\rho)$ en $\mathcal{S}(\rho)$. Als gekeken wordt naar de definitie van deze indices, bestaat vermoedelijk een relatie in het geval dat de toernooiranking uniek te bepalen is, dus zonder gebruik te maken van de nummering van de teams. Hiervoor zou gekeken moeten worden naar een andere definitie van de toernooiranking, door bijvoorbeeld gebruik te maken van een andere wedstrijd waardering.

Met het competitief evenwicht, σ_c, de gewogen toernooi-index, $\mathcal{W}(x)$, en de andere indices hebben we verschillende parameters die allemaal de competitiviteit van een sporttoernooi kwantificeren. Om te bepalen welke index dit het beste doet, zou men een gespeeld toernooi kunnen beschouwen en voor dit toernooi de indices berekenen. Hierbij is het noodzakelijk een goede pre-ranking te bepalen, bijvoorbeeld aan de hand van de dynamische ranking. Deze dynamische ranking zou met behulp van numerieke methoden bepaald kunnen worden. We kunnen concluderen dat de optimale toernooi-index te algemeen is, aangezien geoptimaliseerd wordt over alle mogelijke pre-rankings. Het doel voor een vervolgonderzoek kan zijn deze collectie pre-rankings minimaal te krijgen, door op zoek te gaan naar realistische aannames. Deze aannames kunnen een bijdrage leveren aan het beter representeren van de kwaliteiten van de teams, waardoor de collectie mogelijke pre-rankings verkleind wordt. Hierbij wordt dus gemiddeld over de collectie pre-rankings, wat hetzelfde is als middelen over alle mogelijke waardes van Σ in Kendall's $\tau(\rho, \rho_t)$. Hierbij kan de frequentie worden meegerekend, door gebruik te maken van bepaalde gewichten.

Nadat de meest betrouwbare en meest bruikbare index is bepaald, kan men verschillende toernooien van verschillende sporten beschouwen en onderzoeken om te bepalen welke sport het meest competitief is. Hierbij zou rekening gehouden moeten worden met gelijkspel. In deze bachelorscriptie is het buiten beschouwing gelaten, maar met meer onderzoekstijd kunnen de moeilijkheden en problemen, die veroorzaakt worden door gelijkspel, opgelost worden. Dit zou men kunnen doen door gelijkspel te beoordelen met één punt en een winst met twee punten in de wedstrijd waarderingen, om onderscheid te houden tussen gelijkspel en winst.

Als laatste aanbeveling voor een vervolgonderzoek hebben we het gebruik van een maatstaf voor competitiviteit. In deze bachelorscriptie werd namelijk gezegd dat een toernooi niet competitief is, indien de toernooi-index dicht bij 1 ligt. Dit geeft echter niet aan hoe ver de waarde van 1 af moet liggen, om het toernooi wel als competitief te kunnen beschouwen. Dit is ook het geval bij de andere indices.

Index

Competitief, 3

Dynamische pre-ranking, 35
Dynamische verlies score, 36
Dynamische win score, 36
Dynamische win-verlies score, 37
Dynamische winmatrix, 35

Frequentie polygoon, 21

Gewogen toernooi-index, 33

Inconsistentie, 5

Kendall's τ, 15

Normale pre-ranking, 19

Onterecht gewonnen wedstrijd, 4
Optimale pre-ranking, 6

Permutatie toernooimatrix, 23
Pre-ranking, 3

Sign functie, 5
Slater's i, 10
Sterker, 3

Toernooi-index, 4
Toernooimatrix, 3
Toernooiranking, 3

Unieke toernooiranking, 3

Verlies score, 35
Verslaan, 22
Voorspellingsnauwkeurigheid, 37

Win score, 34
Win-verlies score, 35
Winfrequentie, 3
Winmatrix, 34

Referenties

[1] Voetbalzone (2012). *Van Gaal: 'Misschien verwachten we wel te veel met onze doelstelling'*. Verkregen van http://www.voetbalzone.nl/doc.asp?uid=216121. Geraadpleegd op 06-06-2014.

[2] Lundh, T. (2006). Which Ball is the Roundest? - A Suggested Tournament Stability Index. *Journal of Quantitative Analysis in Sports*, Volume 2, Issue 3.

[3] Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48, 303-312.

[4] Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30, 81–93.

[5] Langville, A. N. & Meyer, C. D. (2012). *The Science of Rating and Ranking- Who's # 1?*. Princeton University Press, 200-204.

[6] Motegi, S. & Masuda, N. (2012). A network-based dynamical ranking system for competitive sports. *Scientific Reports*, Volume 2.

[7] Park, J. & Newman, M. (2005). A network-based ranking system for US college football. *Journal of Statistical Mechanics: Theory and Experiment*, P10014.