GRONDSLAG VAN HET ELO SYSTEEM

Cees Timmer

HOOFDSTUK 1: INLEIDING

Iedere rechtgeaarde schaker heeft wel iets met de rating. Belangrijk vindt hij de rating van zijn tegenstander want dat bepaalt of hij bij winst of verlies elo punten kan vergaren of verliezen. Ook speelt de rating een belangrijke rol bij het rangschikken van spelers op speelsterkte. Bij toernooien worden spelers vaak ingedeeld in groepen op speelsterkte gebaseerd op rating. Sommige sceptici beweren dat de rating niets zegt. Vooral als zij een keer verliezen van een tegenstander met een veel lagere rating dan zijzelf hebben. Hun scepsis is slechts ten dele waar. Een feit is dat je de uitslag van een partij nooit met 100% zekerheid kunt voorspellen. Zoals we zullen zien in dit artikel kun je met behulp van de rating wel iets zeggen over de *kans* dat een partij eindigt in winst, verlies of remise voor een speler.

Het meest wijdverbreide ratingsysteem is bekend als het *Elo systeem*. Arpad E. Elo, geboren in 1903 in Hongarije, emigreerde naar de VS op 10-jarige leeftijd. Van 1935 tot 1965 was hij professor in de natuurkunde en astronomie op de Marquette University, Wisconsin. Het Elo systeem werd geadopteerd door de Amerikaanse schaakbond (USCF) in 1960 en door de FIDE in 1970. Al in 1959 had de USCF een eigen ratingsysteem waarbij willekeurig 2000 als het hoogste niveau voor sterke clubspelers werd gebruikt en klasses van 200 punten groot om de spelers naar speelsterkte in te delen. Elo behield deze "maatvoering" omdat men er al vertrouwd mee was.

Er is op internet veel informatie over het elo systeem te vinden. Probleem hierbij is dat deze informatie of vrij oppervlakkig is of zeer diepgaand statistisch. Met dit artikel heb ik geprobeerd een middenweg te vinden: sommige delen vragen wel enige wiskundige kennis maar nooit op hoog specialistisch niveau. Op sommige plaatsen in het artikel is iets dieper ingegaan op de mathematische statistiek voor een beter begrip van sommige statistische concepten.

HOOFDSTUK 2: VERWACHTE SCORE

Op basis van één partij kun je natuurlijk niets zeggen over de speelsterkte van de spelers. Dat wordt anders als de spelers meerdere keren tegen elkaar hebben gespeeld.

Beschouw een schaakpartij tussen speler A en speler B. Wint Speler A krijgt hij een 1 en speler B een 0, verliest Speler A krijgt hij een 0 en speler B een 1 en bij remise krijgen beide spelers een 0.5. De punten die een speler behaalt in één of meerdere partijen noemen we de *score*. De score kan worden uitgedrukt als absoluut (bijvoorbeeld 6.5 punt uit 10 partijen), procentueel (65%) of fractioneel (0.65). In de schaakstatistiek gebruiken we meestal de fractionele score.

Een score kan je beschouwen als een *toevalsvariabele* omdat je nooit met 100% zekerheid vooraf kunt bepalen wat de score gaat worden. Natuurlijk is het aannemelijk dat een sterke speler een hogere score zal behalen dan een zwakkere speler maar 100% zekerheid heb je niet. Wel kan je spreken over de *kans* op een bepaalde score. Een kans is wiskundig een getal tussen 0 en 1 en wordt aangeduid met de letter P van het Engelse *probability* (kans of waarschijnlijkheid).

Wat zijn de kansen voor speler A als hij tegen speler B speelt? Stel P_w de kans op een gewonnen partij (w = wins), P_l de kans op een verloren partij (l = losses) en P_d de kans op een remise (d = draws). Er geldt natuurlijk dat $P_w + P_l + P_d = 1$ omdat de kans voor speler A dat zijn partij eindigt in winst of verlies of remise altijd 100% is. Zijn dus twee kansen bekend dan kan de derde worden berekend. Voor de statistische fijnproevers onder ons: je mag de kansen hier eenvoudig optellen omdat winst, verlies en remise zogenaamde elkaar uitsluitende gebeurtenissen zijn, d.w.z. zij kunnen niet gelijktijdig optreden. Bijvoorbeeld: een partij kan niet eindigen in winst én verlies. Een voorbeeld van twee gebeurtenissen waar dit wel zo is: het trekken van een rode kaart of een koning uit een kaartspel. Deze twee gebeurtenissen kunnen wel gelijktijdig optreden: de getrokken kaart kan een rode koning zijn. In dat geval mag je de kansen niet zomaar optellen zonder een correctie toe te passen.

De kansen voor speler B zijn eenvoudig af te leiden uit die van speler A. Immers de kans op winst voor speler A is gelijk aan de kans op verlies voor speler B en omgekeerd; de kans op remise is voor beide spelers gelijk. In het vervolg van dit artikel zullen we alleen de kansen voor speler A beschouwen.

Hoe berekenen we P_w, P_l en P_d? Deze kansen kunnen we schatten uit een aantal gespeelde partijen tussen speler A en B. Stel dat zij N partijen tegen elkaar hebben gespeeld en dat N_w het aantal gewonnen partijen van speler A is, N_l het aantal verloren partijen van speler A en N_d het aantal remises. Dan is een schatting van de kansen voor speler A:

$$P_w = \frac{N_w}{N} \qquad P_l = \frac{N_l}{N} \qquad P_d = \frac{N_d}{N} \tag{1}$$

Hoe groter N wordt hoe nauwkeuriger de schatting van de kansen wordt. In de praktijk zal het niet meevallen om deze kansen te berekenen. Ten eerste moet je beschikken over voldoende partijen (bij voorkeur tenminste 10) tussen speler A en B en ten tweede zouden deze partijen binnen een niet al te lang tijdsbestek gespeeld moeten zijn. Bij de N partijen tussen speler A en B die we gebruiken voor de berekening van de kansen nemen we namelijk aan dat hun speelsterkte niet met de tijd sterk verandert.

Stel dat speler A en B een match van 10 partijen tegen elkaar spelen. In deze match wint speler A 5 partijen, verliest er 2 en speelt er 3 remise. Dat betekent dat speler A een score van 6.5 uit 10 heeft behaald ofwel een fractionele score van 0.65. Voor speler A kun je dan berekenen met behulp van Vgl. (1) dat $P_w = 0.5$, $P_l = 0.2$ en $P_d = 0.3$. Voor een toekomstige partij van A tegen B kun je op basis van deze gegevens voorspellen dat er 50% kans is dat A wint, 20% kans dat hij verliest en 30% kans dat het remise wordt.

Gewapend met de kennis van P_w, P_l en P_d kunnen we nu de kans voor iedere willekeurige score van speler A berekenen indien hij een aantal partijen tegen speler B zou spelen. In het algemeen geldt dat de kans $P(W, L, D)$ op W gewonnen partijen, L verloren partijen en D remises uit totaal N partijen voor een speler A tegen speler B gelijk is aan:

$$P(W,L,D) = \frac{N!}{W!L!D!} \cdot P_w^W \cdot P_l^L \cdot P_d^D \tag{2}$$

Laat je niet afschrikken door deze formule. Het uitroepteken staat wiskundig voor faculteit: $N!$ is een verkorte schrijfwijze voor 1 x 2 x 3 ... x N. Op dezelfde wijze voor $W!$, $L!$ en $D!$. De kansen P_w, P_l en P_d moeten respectievelijk tot de macht W, L en P verheven worden in de formule. Vgl. (2) definiëert een zogenaamde *trinominale kansverdeling*. Een schaakpartij wordt statistisch beschouwd als een *toevalsexperiment* met drie mogelijke uitkomsten (0, 1, 0.5) met ieder zijn eigen kans dat zo'n uitkomst optreedt. Deze kansen zijn natuurlijk afhankelijk van de speelsterkte van de spelers.

Wat is de kans op 5x winst, 3x verlies en 2x remise van speler A als hij 10 partijen tegen speler B speelt? Deze uitkomst geeft een score van 5x1 + 3x0 + 2x0.5 = 6 punten uit 10 of 60% als percentage of 0.6 als fractie. Met behulp van Vgl. (2) kunnen we deze kans berekenen:

$$P(5,3,2) = \frac{10!}{5!\,3!\,2!} \cdot 0.5^5 \cdot 0.2^3 \cdot 0.3^2 = 0.0567 \tag{3}$$

Uiteraard zijn er meer combinaties die 6 punten uit 10 opleveren:

$W = 6$ $L = 4$ $D = 0$: $P(6, 4, 0) = 0.00525$
$W = 4$ $L = 2$ $D = 4$: $P(4, 2, 4) = 0.0638$
$W = 3$ $L = 1$ $D = 6$: $P(3, 1, 6) = 0.0153$
$W = 2$ $L = 0$ $D = 8$: $P(2, 0, 8) = 0.000738$

Telt men deze kansen op dan krijgt men een kans van 0.14 voor een score van 6 uit 10.

De volgende stap is dat we de kans willen weten voor iedere mogelijke score van speler A tegen speler B in 10 partijen. Alle mogelijke scores zijn 0, 0.5, 1.5, 2, ..., 10 uit 10, totaal 21 mogelijke scores. Voor ieder van deze scores moet men nagaan welke combinatie van W, L en D deze score kan opleveren. Er blijken totaal 66 combinaties van W, L en D te bestaan die alle mogelijke scores opleveren. Voor ieder van deze combinaties berekent men $P(W, L, D)$ met behulp van Vgl. (2) en sommeert de kansen van de combinaties met dezelfde score. Vervolgens maken we een grafiek van de fractionele scores op de x-as tegen de kansen op de y-as. Dat levert de volgende grafiek op:

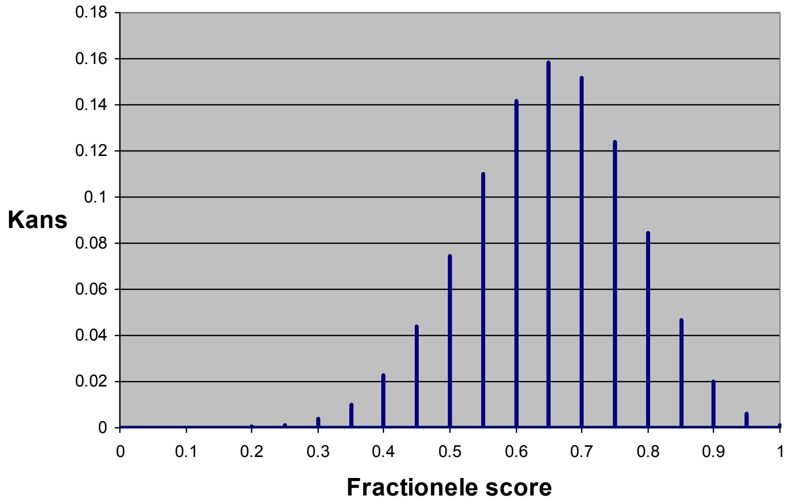

Fig. 1: Kansverdeling van de score uit 10 partijen met $P_w = 0.5$, $P_l = 0.2$ en $P_d = 0.3$.

Bovenstaande grafiek stelt een *kansverdeling* voor. De verdeling in Fig. 1 zal naderen tot een normale verdeling (klokvormige of Gausse curve) als N voldoende groot is. Bij $N = 10$ hebben we al een redelijk normale verdeling. Dit is overigens geen verrassing: de hoeksteen van de mathematische statistiek, het zogenaamde *centrale limiettheorema* zegt dat iedere niet-normale verdeling tendeert naar een normale verdeling bij toenemende N. Merk op dat alle kansen in Fig. 1 opgeteld 1 moeten opleveren.

Uit Fig. 1 is duidelijk dat de score 6.5 uit 10 de grootste kans heeft op te treden. We noemen deze score uitgedrukt als fractie de *verwachte score* die wordt aangeduid met het symbool w_e (subscript e komt van het Engelse "expected"). De kans op een score van 0.65 is 0.16 zoals uit Fig. 1 is af te lezen of te berekenen met Vgl. (2). Zouden we een grotere N nemen (bijvoorbeeld 30) dan blijft de verwachte score 0.65. Echter de kans op zo'n score wordt dan veel kleiner (0.0926) omdat er veel meer mogelijke scores zijn dan bij $N = 10$. Het is duidelijk dat naarmate er meer mogelijke scores zijn, de kans op één specifieke score steeds kleiner wordt. Uiteindelijk moeten wel alle kansen van de verdeling opgeteld 1 opleveren. Het is daarom relevanter de kans te berekenen dat een score groter of kleiner is dan een bepaalde score of dat de score tussen twee bepaalde scores in ligt. Bijvoorbeeld: wat is de kans dat speler A de tweekamp van 10 partijen van speler B wint? Daartoe tellen we alle kansen voor de scores ≥ 5.5 uit 10 (fractionele score 0.55) op dat 0.84 oplevert.

Overigens is het niet noodzakelijk de gehele kansverdeling uit te rekenen om de verwachte score te bepalen. De statistische theorie over de multinominale verdeling zegt dat het verwachte aantal keren dat voor speler A in 10 partijen tegen speler B winst voorkomt gelijk is aan 10 x $P_w = 5$, verlies 10 x $P_l = 2$ en remise 10 x $P_d = 3$, hetgeen resulteert in een score van 6.5 uit 10. Wil je echter ook de kans op deze score weten dan zal je met Vgl. (2) de berekening moeten uitvoeren zoals is toegelicht voor een score van 0.60.

Je kunt nu voorspellen wat de score van *A* wordt in een partij tegen *B*, namelijk 0.65. Dat klinkt vreemd omdat *A* slechts een 1, 0 of 0.5 kan scoren in een partij. Echter moet de verwachte score worden opgevat als de score die *A* gehad zou hebben indien hij meerdere partijen tegen *B* zou hebben gespeeld. Het begrip “verwachte score” speelt een belangrijke rol in het ratingconcept van Elo zoals we hierna zullen zien. De kans dat speler *A* een partij tegen *B* wint is $P_w = 0.5$ zoals experimenteel bepaald uit 10 gespeelde partijen tussen *A* en *B*. Merk op dat de kans dat speler *A* een tienkamp van speler *B* wint veel groter is (0.84) zoals in bovenstaand voorbeeld was uitgerekend. Speler *A* heeft over meerdere partijen meer kans zijn grotere speelsterkte waar te maken.

De verwachte score zou je kunnen opvatten als een maat voor de speelsterkte. In bovenstaand voorbeeld is de speelsterkte van speler *A* duidelijk groter dan die van speler *B*. Alleen deze maat is niet erg praktisch omdat 1) het in de praktijk vaak onmogelijk is de waarden van P_w, P_l en P_d te bepalen voor verschillende tegenstanders en 2) de maat niet algemeen geldig is, d.w.z. hij heeft altijd betrekking op één bepaalde tegenstander. Zoals we experimenteel zullen laten zien in HOOFDSTUK 6 kan de speelsterkte van een speler bepaald ten opzichte van één bepaalde tegenstander afwijken van zijn speelsterkte bepaald tegen verschillende tegenstanders.

Bovenstaande beschouwingen bewijzen vooral hun nut als je een voorspelling wilt maken van de score over één of meerdere partijen tussen twee bepaalde spelers.

HOOFDSTUK 3: ELO SYSTEEM

Elo noemde de speelsterkte “rating”. Hij nam aan dat de rating van een speler een normale verdeling volgt. De gedachte erachter is dat de speelsterkte van een speler niet constant is: hij speelt in de ene partij sterker dan in de andere partij. Met andere woorden zijn speelsterkte fluctueert van partij tot partij. Deze fluctuaties volgen een normale verdeling. In formule uitgedrukt als:

$$R_0 \sim N(\mu_0, \sigma^2)$$

Dit is een wiskundige manier om uit te drukken dat de rating van speler *A* aangeduid als R_0 een normale verdeling volgt met gemiddelde μ_0 en variantie σ^2. Voor de rating van zijn tegenstander *B* aangeduid als R_c geldt op vergelijkbare wijze:

$$R_c \sim N(\mu_c, \sigma^2)$$

Elo nam aan dat de spreiding van beide verdelingen gelijk is. Deze aanname is wel enigzins discutabel omdat de ene speler meer variatie in speelsterkte kan vertonen dan een andere speler. De grootheid μ_0 stelt de gemiddelde rating van speler *A* voor en μ_c die van tegenstander *B*.

Statistisch intermezzo:
Formeel gesproken zijn μ_0 en μ_c *populatiegemiddelden* van de speelsterktes. Hoe bepaal je nu μ_0 en μ_c? Kortweg gezegd: deze kun je niet exact bepalen maar wel benaderen met

iedere gewenste nauwkeurigheid. Voor het schatten van μ_0 moet je de speelsterkte van speler *A* een groot aantal keren "meten" en het gemiddelde uitrekenen. Dit heet het *steekproefgemiddelde*. Statistisch kan worden aangetoond dat dit steekproefgemiddelde een zuivere schatting van μ_0 is, d.w.z. zou je een oneindig aantal keren de speelsterkte meten dan is het steekproefgemiddelde exact gelijk aan het populatiegemiddelde μ_0. Uiteraard is dit praktisch onmogelijk en zijn μ_0 en μ_c theoretische waarden.

Nu is het probleem dat je de rating niet absoluut kunt meten. Men kan niet aan de hand van een reeks zetten zeggen: deze speelsterkte is 2039. Er is wel eens gesuggeerd de partij te laten evalueren door een sterk schaakprogramma die dan hieruit de speelsterkte berekent. Wel kan je zeggen: als een speler wint wordt er verondersteld dat zijn rating groter is dan die van zijn tegenstander. Omgekeerd als hij verliest neemt men aan dat zijn rating lager is dan die van zijn tegenstander. Als de partij remise eindigt, neemt men aan dat beide spelers dezelfde rating hebben. Omdat een rating niet absoluut bepaald kan worden, heeft een rating daarom alleen betekenis met betrekking tot andere ratings. Daarom gaan we nu kijken naar het verschil in rating.

Voor de verdeling van het verschil in rating R_v geldt:

$$R_v = R_0 - R_c \sim N(\mu_0 - \mu_c, 2\sigma^2) \tag{4}$$

De statistische theorie zegt dat het verschil van twee normaal verdeelde toevalsgrootheden met ieder een variantie van σ^2 weer een normale verdeling volgt met variantie $2\sigma^2$. Beschouw de distributiefunctie van R_v voor $\mu_0 - \mu_c = 0$ zoals is weergegeven in Fig. 2.

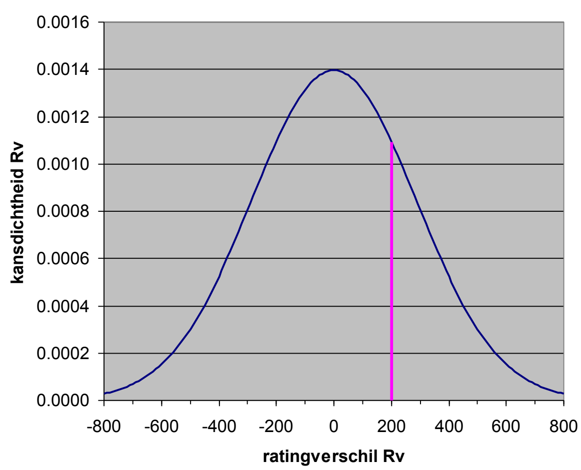

Fig. 2: Distributie van R_v voor $\mu_A - \mu_B = 0$

Wat zegt deze grafiek? Het geeft de kansverdeling van het verschil in rating tussen twee spelers die dezelfde rating hebben. Omdat de rating van beide spelers kan variëren van partij tot partij, kan in een bepaalde partij de ene speler sterker of zwakker spelen dan de andere speler op dat moment. Stel dat speler *A* in een bepaalde partij even sterk speelt als speler *B* dan is het ratingverschil $R_v = 0$ en zal de uitslag remise (0.5) zijn. In Fig. 2

betekent dit dat het oppervlak onder de curve van $R_v = -\infty$ tot $R_v = 0$ gelijk is aan 0.5 (in de statistiek heet zo'n stuk oppervlak "kansmassa"). Stel nu dat speler *A* 200 elopunten sterker speelt dan speler *B* in een bepaalde partij (zie de rode vertikale lijn in Fig. 2), dan beschouwen we het oppervlak onder de curve van $R_v = -\infty$ tot $R_v = 200$. Dit oppervlak is door Elo *per definitie* vastgesteld op 0.75. Om dit oppervlak te verkrijgen moet de standaard afwijking ($\sigma\sqrt{2}$) van de distributie gelijk aan 2000/7 gekozen worden. Men kan voor ieder willekeurige ratingverschil r de kansmassa bepalen, welke gelijk is aan het oppervlak onder de distributiecurve van $R_v = -\infty$ tot $R_v =$ r. Dit geeft de verwachte score welke wordt aangeduid met w_e waarvoor geldt $0 < w_e < 1$.

HOOFDSTUK 4: RATINGBEREKENING

Om de verwachte score te berekenen kan men eenvoudiger gebruik maken van de cumulatieve distributiefunctie van R_v, te weten $\Phi_{\mu,\sigma^2}(R_v)$ gedefiniëerd als

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{(u-\mu)^2}{2\sigma^2}\right)} du \qquad (5)$$

Deze functie (een zogenaamde *integraalfunctie*) ziet er wel wat ingewikkeld uit maar laat zich als volgt eenvoudig uitleggen. De functie geeft voor iedere waarde van x het oppervlak onder de curve zoals in Fig. 2 weergegeven van $R_v = -\infty$ tot $R_v = x$. Algemeen wordt een oppervlak onder een continue curve berekend met een integraal zoals te zien is in Vgl. (5).
De functiewaarden voor verschillende waarden van x kunnen berekend worden in het MS programma Excel met de functie NORMDIST(R_v;0;2000/7;TRUE) en is als grafiek in Fig. 3 is weergegeven. Er bestaan ook tabellen voor Vgl. (5).

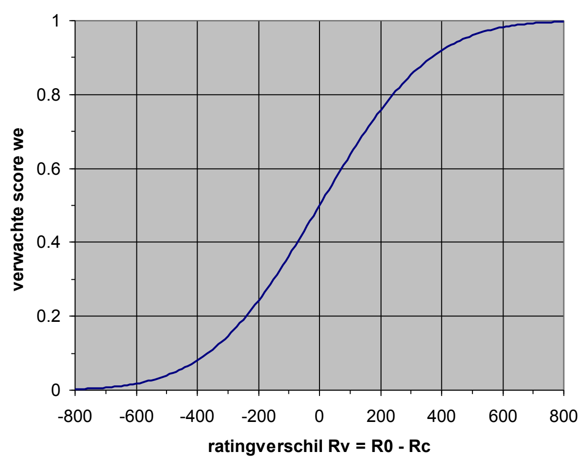

Fig. 3: Cumulatieve normale verdeling van R_v

Bovenstaande grafiek geeft het verband tussen het verschil in rating van twee spelers en de verwachte score. Zoals uit Fig. 3 is af te lezen is bij een ratingverschil van 0 de verwachte score 0.5 en bij een ratingverschil van 200 de verwachte score 0.75 zoals bepaald per definitie door Elo.

Bij het bepalen van de verwachte score is dus het uitgangspunt dat de werkelijke speelsterkte van beide spelers gelijk is. Aan de hand van het verschil in rating tussen beide spelers wordt met behulp van de boven beschreven statistische verdeling bepaald wat de verwachte score van speler A tegen speler B is. Is deze groter of kleiner dan 0.5 dan wordt de rating van A bijgesteld. Indien zijn verwachte score 0.5 is, blijft zijn rating ongewijzigd. We noemen de werkelijk behaalde score w. Van belang is dan de grootheid $w - w_e$. Is $w - w_e > 0$ dan is beter gepresteerd dan verwacht op basis van het ratingverschil en krijgt de speler er ratingpunten bij. Is $w - w_e < 0$ dan is slechter gepresteerd dan verwacht op basis van het ratingverschil en gaan er ratingpunten af. Is $w - w_e = 0$ dan is gepresteerd zoals verwacht op basis van het ratingverschil en blijft de rating van de speler ongewijzigd. Om $w - w_e$ te "vertalen" in ratingpunten die erbij of eraf gaan wordt $w - w_e$ vermenigvuldigt met een constante factor k, de zogenaamde *verversingsfactor*. De ratingverandering R_t kan worden weergegeven als $R_t = k\,(w - w_e)$. Als $w > w_e$ dan is $R_t > 0$ en als $w < w_e$ dan is $R_t < 0$. De nieuwe rating R_n kan dan verkregen worden als:

$$R_n = R_l + R_t = R_l + k\,(w - w_e) \tag{6}$$

waarin R_l de oude rating is. De subscript l komt van "lijst" en slaat op de laatst gepubliceerde ratinglijst van de KNSB.
Aan de hand van een voorbeeld laten we zien hoe de berekening uitgevoerd wordt. Het enige lastige van de berekening is het bepalen van w_e waarvoor MS Excel kan worden gebruikt.

Tabel 1: Voorbeeld van een ratingberekening met $k = 25$.

R_0	R_c	w	w_e	R_t
1492	1233	1.0	0.818	4.6
1492	1357	0.0	0.682	-17.0
1492	1480	1.0	0.517	12.1
1492	1497	0.0	0.493	-12.3
			Totaal	-13

In bovenstaand voorbeeld was de oude rating $R_l = 1492$, dus de nieuwe rating wordt $R_n = 1492 - 13 = 1479$. R_0 is de rating van de speler waarvoor de nieuwe rating bepaald moet worden en R_c is de rating van zijn tegenstander. Deze ratings komen van de eerste KNSB-ratinglijst die voorafgaat aan de rondedatum.

HOOFDSTUK 5: K-FACTOR

De verversingsfactor of k-factor is in de eerste plaats afhankelijk van het aantal gespeelde partijen N_v waarop R_0 is gebaseerd. Bij ratings tussen 2100 en 2400 speelt tevens de waarde van R_0 een rol. Zie Tabel 2 voor de k-factor zoals gebruikt door de KNSB.

Tabel 2: Waarden van k voor verschillende R_0 en N_v

Verversingsfactor *k*		
R_0	$N_v < 75$	$N_v \geq 75$
≤ 2100	$\frac{216}{\sqrt{N_v}}$	25
$2100 < R_0 < 2400$	$\frac{216}{\sqrt{N_v}}$	$25 - \frac{R_0 - 2100}{20}$
$R_0 \geq 2400$	$\frac{216}{\sqrt{N_v}}$	10

De k-factor bepaalt de maximale ratingaanpassing per partij. De k-factor is minimaal 10 voor spelers met een rating ≥ 2400 waarbij de oude rating gebaseerd is op meer dan 75 partijen. De k-factor is maximaal 88 voor spelers met $N_v = 6$ (het vereiste minimaal aantal partijen voor een ratingberekening). Hoe minder partijen waarop de oude rating is gebaseerd hoe groter de ratingverandering bij een zelfde verschil tussen de werkelijke en de verwachte score. Bij ratings boven de 2100 en 2400 zijn de veranderingen nog kleiner. Anders gezegd: een rating gebaseerd op weinig partijen fluctueert meer dan een rating gebaseerd op veel partijen.

Er is binnen de internationale schaakwereld veel discussie over de k-factor. Verschillende schaakorganisaties (zoals bijvoorbeeld FIDE, USCF en KNSB) hanteren vaak verschillende k-factoren. Het probleem hierbij is dat de k-factor geen objectieve statistische onderbouwing heeft. Het idee erachter is dat een rating gebaseerd op veel partijen (> 75) als betrouwbaarder wordt gezien dan een rating gebaseerd op weinig partijen. Daarom worden veranderingen in een betrouwbare rating minder zwaar meegeteld, met andere woorden wordt een kleinere k-factor gebruikt. Bij een rating gebaseerd op meer dan 75 partijen wordt de k-factor bovendien kleiner gemaakt voor sterke spelers (rating > 2100). Het idee hierachter is om het voor ervaren en sterke schakers moeilijker te maken ratingpunten te vergaren maar ook te verliezen per partij. Tal van statistici hebben voorstellen gedaan om het Elo systeem te verbeteren maar tot op de dag van vandaag blijft het Elo systeem stand houden als het ratingsysteem in de schaakwereld.

In Tabel 3 is weergegeven hoeveel elo punten (R_t) je rating stijgt of daalt bij verlies ($w = 0$), remise ($w = 0.5$) of winst ($w = 1$) als je tegenstander R_v elo punten *minder* heeft voor $k = 25$. Ten tijde van dit schrijven geldt voor de meeste OSV-leden deze k-waarde.

Tabel 3: R_t waarden voor $k = 25$.

R_v	w_e	R_t		
		$w = 0$	$w = 0.5$	$w = 1$
0	0.50	-12.5	0	12.5
25	0.53	-13.4	-0.9	11.6
50	0.57	-14.2	-1.7	10.8
75	0.60	-15.1	-2.6	9.9
100	0.64	-15.9	-3.4	9.1
125	0.67	-16.7	-4.2	8.3
150	0.70	-17.5	-5.0	7.5
175	0.73	-18.2	-5.7	6.8
200	0.76	-19.0	-6.5	6.0
225	0.78	-19.6	-7.1	5.4
250	0.81	-20.2	-7.7	4.8
275	0.83	-20.8	-8.3	4.2
300	0.85	-21.3	-8.8	3.7
325	0.87	-21.8	-9.3	3.2
350	0.89	-22.2	-9.7	2.8
375	0.91	-22.6	-10.1	2.4
400	0.92	-23.0	-10.5	2.0

Tabel 3 kun je ook gebruiken als je tegenstander R_v elo punten *meer* heeft. Alle min-tekens in kolom $w = 0.5$ worden dan plus-tekens en verwisselen de R_t-waarden in kolom $w = 1$ en kolom $w = 0$ van plaats met verwisseling van teken. Voorbeeld: je tegenstander heeft 200 elo punten meer. Bij verlies gaan er 6 elo punten af, bij remise komt er 6.5 elo punt bij en bij winst komen er 19 elo punten bij.

Uit de Tabel 3 blijkt dat bij een groot ratingverschil (bijv. 400) tussen twee spelers de sterkste speler veel elo punten (23) kan verliezen bij partij-verlies en maar weinig elo punten (2) kan verdienen bij partij-winst. Omgekeerd zal de zwakkere speler bij partij-verlies weinig elo punten (2) verliezen maar veel elo punten (23) erbij krijgen bij partij-winst.

HOOFDSTUK 6: RATING BINNEN OSV

In de archieven van OSV bevindt zich een groot aantal partij- en ratinggegevens. Ik was benieuwd in hoeverre het verschil in rating tussen twee spelers overeenkomt met de verwachte score bepaald over een aantal gespeelde partijen tussen deze spelers zoals besproken in HOOFDSTUK 2. Alleen partijen uit de huiscompetitie en het Open Osse Schaakkampioenschap zijn in de beschouwingen betrokken omdat deze partijen op basis van het speeltempo voor ratingverwerking in aanmerking komen. De meeste kans om twee spelers te vinden die een redelijk aantal keren ($N > 10$) tegen elkaar hebben gespeeld zijn natuurlijk de vaste “bewoners” van Groep 1. Gevolg van deze keuze is wel dat deze spelers geen groot verschil in rating zullen hebben anders zouden zij niet permanent in Groep 1 spelen.

De volgende werkwijze werd gebruikt: vanaf 2007 tot op heden werden van twee spelers alle partijen verzameld die zij tegen elkaar hadden gespeeld. Op basis van de partij-uitslagen werd de behaalde score w bepaald voor de spelers. Tevens werd de KNSB-rating van de spelers opgezocht die zij ten tijde van iedere partij hadden. Op basis van het ratingverschil per partij werd de verwachte score w_e per partij berekend volgens Vgl. (5). Omdat het ratingverschil per partij kan verschillen zal ook w_e per partij verschillen. Berekend is daarom ook de gemiddelde w_e met het bijbehorende 95% vertrouwens-interval. In Tabel 4 staan de resultaten van de bevindingen.

Tabel 4: Behaalde en verwachte scores tussen twee spelers.

	N	R_A	R_B	R_v	w	w_e	**95% CI voor w_e**
A = Klein Douwel B = Hartendorp	23	1897	1819	78	0.57	0.61	0.44-0.77
A = Snoek B = Broekkamp	26	1940	1833	107	0.69	0.65	0.47-0.81
A = Broekkamp B = Hartendorp	12	1829	1820	9	0.79	0.51	0.37-0.65
A = Scheepers B = de Kroo	10	1665	1661	4	0.70	0.51	0.37-0.64

Verklaring van de symbolen:
N : aantal gespeelde partijen
R_A : gemiddelde rating van speler A
R_B : gemiddelde rating van speler B
R_v : gemiddeld verschil in rating tussen speler A en B
w : behaalde score van speler A uit N partijen tegen speler B
w_e : gemiddelde verwachte score van speler A op basis van verschil in rating
CI : vertrouwensinterval (Eng. confidence interval)

Voor de eerste twee gevallen vind je een goede overeenkomst tussen de werkelijk behaalde score over meerdere partijen en de verwachte score op basis van het ratingverschil. Echter voor de laatste twee gevallen is de werkelijk behaalde score duidelijk hoger dan verwacht op basis van het ratingverschil: de behaalde score ligt buiten het 95% vertrouwensinterval van w_e. Speler B heeft in totaal ratingpunten verloren in zijn partijen tegen speler A maar blijkbaar dit weer gecompenseerd met goede resultaten tegen andere tegenstanders.

Rating als maat voor de speelsterkte zoals bepaald tegen verschillende tegenstanders geeft een goede voorspelling van behaalde scores in een toernooi zoals weergegeven in Fig. 4. Deze gegevens komen uit het Open Osse Schaakkampioenschap 2015. Alleen spelers die minimaal 8 van de 9 ronden hebben deelgenomen zijn opgenomen in de figuur.

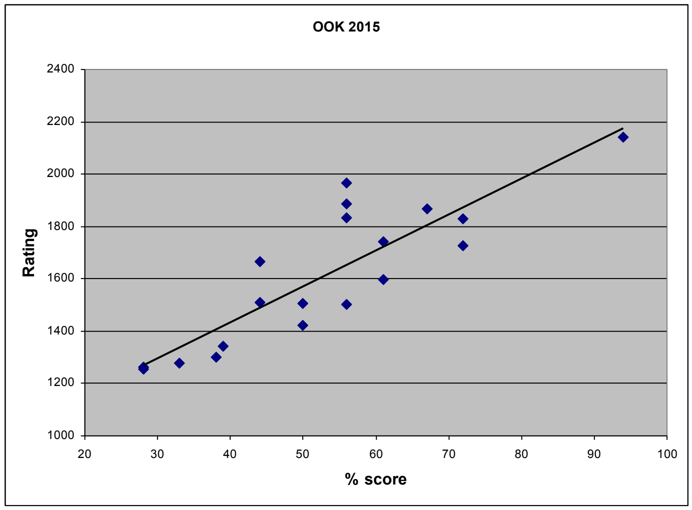

Fig. 4: Verband tussen rating en toernooi resultaat.

HOOFDSTUK 7: TOT SLOT

Niet alle details over de ratingberekening zijn aan het bod gekomen zoals de berekening van een startrating als je nog geen rating hebt, de Lijst-Prestatie-Rating (LPR) en de LPR limiteringen. Informatie hierover is te vinden in het handboek van de KNSB (link: http://www.schaakbond.nl/rating/knsb-ratinglijst/rekenregels-knsb-ratingsysteem).

Ik hoop met dit artikel duidelijk gemaakt te hebben dat rating terdege wel iets zegt mits op de juiste wijze geïnterpreteerd.

Oss, maart 2016