THE QUEEN'S GAMBIT

In haar column *Ionica Smeets zag een getal* in *de Volkskrant* van 15 januari 2020 ging het over het getal 86. Ionica maakt aannemelijk dat er iets niet klopt aan een 'bewijs' van de stelling dat mannen beter kunnen schaken dan vrouwen. Dat bewijs zou zijn dat de beste vrouwelijke schaakster, Hou Yifan, (toen) pas op plaats 86 van de totaalranglijst van schakers stond.

door
Jaap Klouwen

Toevallig was in 2020 de populaire Netflix-serie *The Queen's Gambit* bezig aan een opmars in Nederland. Daarin wordt, fictief en in de jaren '50 en '60 spelend, de Amerikaanse Beth Harmon wereldkampioen schaken door – hoe kan het ook anders – een Russische grootmeester te verslaan. Zij zou dan dus rangnummer 1 hebben gehad, berekend met de zogeheten *Elo-rating* van de schaakwereld, ontwikkeld door de Hongaars-Amerikaanse natuurkundige en schaker Árpád Élő.

Met behulp van een voorbeeld van de Nederlandse schaker en hoogleraar neurale wetenschap Wei Ji Ma – dat ik hier heb aangepast, met een iets beter model – laat Ionica zien dat er een andere reden kan zijn voor die lage klassering 86 voor de beste vrouwelijke schaker. Het begint met de vaststelling dat er veel minder vrouwelijke dan mannelijke schakers zijn. Stel dat er *twee* vrouwelijke schakers in de top 100 staan. Aangenomen dat vrouwen en mannen even goed kunnen schaken, wat is dan het rangnummer van de beste van die twee vrouwen? Stel je vervolgens een tweede situatie voor: er zijn *tien* vrouwelijke schakers in de top 100. Waar staat dan (gemiddeld) de beste van hen? Merk op dat het gemiddelde rangnummer van deze beide groepen 50,5 (niet 50!) zal zijn. Maar gevoelsmatig moet de beste in de groep van tien een beter (lager) rangnummer hebben dan de beste in de groep van twee. Dat komt natuurlijk omdat er in de grotere groep meer mogelijkheden zijn voor een hogere klassering. Het leuke is: die gemiddelde beste klassering kun je uitrekenen!

Voor de groep van twee dames is de beste klassering 33⅔, zoals we hierna zullen aantonen. (Uiteraard kan zo'n gemiddelde beste klassering een decimaal getal zijn, want het is, deftig statistisch gezegd, een *verwachtingswaarde*. Zo'n verwachtingswaarde hoeft geen geheel getal te zijn.) Maar hoe zit het dan in de groep van tien vrouwelijk top 100-schakers? Wat is daar de verwachte beste klassering?

We zullen eerst voorrekenen hoe je aan die beste gemiddelde klassering van 33⅔ komt bij een *tweetal* vrouwelijke schakers in de top 100 van alle schakers. Wat is de kans dat het beste rangnummer van die twee bijvoorbeeld 40 is? Die kans is 2 maal de kans op precies rangnummer 40 vermenigvuldigd met de kans op een rangnummer hoger dan 40, ofwel 2 · (1/100) · (60/99) = 0,01212 (afgerond op 5 decimalen). Die factor 2 is nodig omdat beide genoemde gebeurtenissen ook in omgekeerde volgorde kunnen optreden. Noem $\underline{k}$ de kansvariabele die de rang van de beste vrouwelijke schaker aangeeft. Wat is de kans op het minimum k bij aselecte trekking van twee (ongelijke) getallen uit 1 tot en met 100. De kans is:

$2 \cdot P(\underline{k} = k) \cdot P(\underline{k} > k) = 2 \cdot (1/100) \cdot (100 - k)/99 = (200 - 2k)/9900$.

Het berekenen van het gemiddeld beste rangnummer gaat via de verwachtingswaarde van de kansverdeling. De verwachtingswaarde is de som van alle mogelijke uitkomsten die eerst vermenigvuldigd zijn met hun bijbehorende kans. Een van die bijdragen in de verwachtingswaarde E (*expected value*) is 40 · 0,01212 = 0,4848 (afgerond op 4 decimalen) van het eerdere voorbeeld. Merk op dat de kans op een minimum van 100 gelijk is aan 0, want twee keer 100 kiezen is onmogelijk.
Alle waarden optellend volgt voor de verwachtingswaarde:

$$E = \sum_{k=1}^{99} k \cdot \frac{200-2k}{9900} = \frac{2}{99}\sum_{k=1}^{99} k - \frac{2}{9900}\sum_{k=1}^{99} k^2.$$

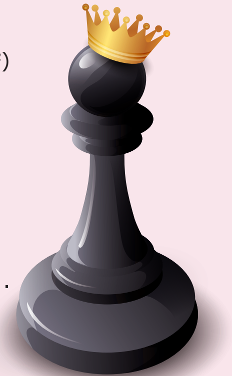
De beide sommen $\sum_{k=1}^{99} k$ en $\sum_{k=1}^{99} k^2$ (ofwel $1 + 2 + 3 + \cdots + 99$ en $1^2 + 2^2 + 3^2 + \cdots + 99^2$) zijn bekend en hebben als uitkomst (n in plaats van 99) respectievelijk
$\tfrac{1}{2} \cdot n \cdot (n + 1)$ en $n \cdot (n + 1) \cdot (2n + 1)/6$.
Invullen van $n = 99$ geeft:

$$\frac{2}{99}\sum_{k=1}^{99} k = \frac{2}{99} \cdot \frac{1}{2} \cdot 99 \cdot 100 = 100.$$

$$\frac{2}{9900}\sum_{k=1}^{99} k^2 = \frac{2}{9900} \cdot 99 \cdot 100 \cdot 199/6 = 66\tfrac{1}{3}.$$

Dus

$$E = 100 - 66\tfrac{1}{3} = 33\tfrac{2}{3}.$$



En nu dezelfde vraag bij een groep van 10 vrouwen in de top 100. Dan zijn in ieder geval de kansen op een minimaal rangnummer van 91 tot en met 100 gelijk aan 0. De andere kansen zijn nu aanzienlijk ingewikkelder. Die op bijvoorbeeld een minimum 40 is $10 \cdot (1/100) \cdot (60/99) \cdot (59/98) \cdot \cdots \cdot (52/91)$.
Al deze kansen in Excel zettend, vermenigvuldigend met die waarde zelf en optellend, volgt daarmee een verwachtingswaarde $9\tfrac{2}{11}$, afgerond dus 9,2. In het artikel van Ionica Smeets is die waarde (met een iets ander model en omgerekend, want zij berekent het *maximum*) 9,6.
Dus kunnen we concluderen dat we de beste vrouw wiskundig gezien een stuk hoger zouden mogen verwachten dan plaats 86? <