

Testen van series handjes

Hans van Staveren

Inleiding

Recentelijk kwam de vraag of een bepaalde grote serie spellen wel voldeed aan de normale kansverdeling. Deze serie was geschud door BigDeal, en gespeeld op een club die daarvoor nog nooit computerhanden had gespeeld. Sommige leden werden helemaal hyper van de spellen. Deze serie leek vrij normaal, maar dit leidde tot testen van andere series(handgeschud) en uiteindelijk tot de ontwikkeling van software die kan bepalen of grote series spellen *goed* of *niet goed* geschud zijn.

Deze software is zeker **niet** bedoeld om spellen te testen voor het spelen, maar om grote hoeveelheden spellen te testen achteraf, om te controleren of de methode van schudden goed genoeg is. Spellens controleren voor het spelen moet **nooit** gebeuren.

Theorie verdelingen

Je kunt van alles en nog wat testen op spellen, maar in dit document beperken we ons tot de distributies. Bridge handjes variëren van zeer gebalanceerd (4-3-3-3) tot extreem ongebalanceerd (13-0-0-0) en alles daar tussenin. Wat is nu eigenlijk normaal?

Nu, de theorie zegt het volgende: elk handje heeft evenveel kans te hebben om voor te komen. Als dat zo is dan is de fractie van een bepaald soort handje, zeg een 4-4-3-2, gelijk aan het aantal 4-4-3-2 handjes, gedeeld door het totaal aantal handjes. Nu is het aantal 4-4-3-2 handjes gelijk aan:

$$\binom{13}{4} \cdot \binom{13}{4} \cdot \binom{13}{3} \cdot \binom{13}{2} \cdot 12$$

waarbij het eerste haakje representeert het aantal manieren om 4 kaarten te kiezen uit 13. Het tweede, derde en vierde haakje kiest de rest, en de 12 komt omdat er 12 manieren (4x3) zijn om de kleuren te kiezen waar de 3-kaart en 2-kaart zit.

Als je dat getal dan deelt door het aantal manieren om 13 kaarten te kiezen uit 52, dan krijg je de theoretisch verwachte fractie van de 4-4-3-2 verdelingen. Dit is 0,2155117565 en is de grootste fractie van alle verdelingen.

Bij iedere correcte manier van schudden, met de hand of met de computer, hoort de fractie 4-4-3-2 verdelingen dus ongeveer zo groot te zijn. Let op het woord ongeveer. Hoe bepaal je nu of in een bepaalde serie aan die *ongeveer* voorwaarde is voldaan? Met het blote oog kijken en met de hand wuiven is mogelijk, maar geeft wetenschappelijk niet echt voldoening.

Theorie kansdichtheden

De kansverdeling die hier van toepassing is is de zogenaamde *binomiale* verdeling. Dit is een kansverdeling waarbij je een aantal van n experimenten doet, die *onafhankelijk* van elkaar slagen met kans p . Denk bijvoorbeeld aan het 100 keer gooien met een dobbelsteen en het aantal zessen tellen. Die kansverdeling is binomiaal met parameters 100 en $1/6$, het aantal experimenten en de

kans op succes. Als je 100 bridgehandjes schudt zal dus het aantal 4-4-3-2 verdelingen binomiaal verdeeld zijn met parameters 400 en 0,2155117565 want er zijn 400 handen op 100 spellen.

Als je op een te testen serie het aantal 4-4-3-2 handjes telt, kun je met een standaard stuk wiskunde uitrekenen wat de kans is op dat aantal of minder, dit heet de cumulatieve kansverdeling(CK). De CK van 0 van die handjes is natuurlijk zeer klein, de CK van het aantal handen(4x het aantal spellen) is natuurlijk 100%. Als het aantal 4-4-3-2 handen zodanig is dat de CK van dat aantal ver verwijderd is van de 50% is er reden om te twijfelen of die serie wel goed geschud is. In de toetsingstheorie van de statistiek is het gebruikelijk om van te voren een percentage te noemen waarbinnen je het als correct accepteert. Gebruikelijk is iets als 95%, wat betekent dat als de CK tussen de 2.5% en 97.5% ligt je het aantal als correct accepteert, en als het daarbuiten ligt als vermoedelijk fout.

Praktijk van de test

Ik heb software gemaakt die alle theoretische kansen van de verdelingen kent. Die software leest alle handen en telt alle distributies. Aan het eind wordt voor alle verdelingen berekend of het aantal in de praktijk voorkomende binnen die 95% valt. Afhankelijk van het aantal spellen worden omstreeks 30 verdelingen getest. Van die 30 vallen er dus tussen de 0 en 30 buiten de 95% limieten. Dat er een klein aantal buiten valt is dus volstrekt normaal, want 1 op de 20 is wat er buiten de 95% valt. Maar als er heel veel buiten vallen wordt het verdacht.

Theorie van de test

Er vanuit gaande dat de handen correct geschud zijn zal ieder van de tests op een verdeling dus falen met kans 5%. Iedere test is onafhankelijk, dus de hele serie tests is dan weer binomiaal verdeeld met parameters 30 en 0.05.

Dat geeft de mogelijkheid om diezelfde wiskundige CK functie weer te gebruiken om de kans te berekenen dat deze hele serie tests overeenkomt met de aanname dat de spellen correct zijn geschud. We moeten dus nu een drempel kiezen waarboven we de spellen niet meer correct achten. Deze drempel kan bijvoorbeeld 99% zijn.

Als die drempel wordt overschreden zal de software dus van de set spellen zeggen dat hij niet goed geschud is. De kans dat de software dat fout heeft is dan 1%. De keuze van de drempel zorgt dus voor een tradeoff: meer sets afkeuren met lagere zekerheid, of minder afkeuren met grotere zekerheid. Dit is standaard in toetsingstheorie.

Uitkomsten

Met deze software zijn een aantal sets spellen getest. Het zal voor de kenners weinig verrassend zijn dat de spellen die met het de facto standaard programma BigDeal zijn geschud aan de eisen voldoen. Er zijn ook wat sets handmatig geschudde spellen getest, momenteel van twee verenigingen. Beide sets falen de test, ze zijn (veel) te vlak, waarbij de spellen van de ene vereniging nog erger zijn dan die van de ander. We praten hier over minstens iets van 20% te weinig renonces en 20% teveel 4-3-3-3 handen. Saaie spellen dus.

Om even van onze wetenschappelijke hoogte af te dalen naar de gewone bridger is het niet onbelangrijk om te vermelden dat de spellen die het vlakst zijn door die verenigingsleden juist vaak worden gezien als te wild.

Als de NBB ooit testen van spelsets als service gaat aanbieden zal het dus absoluut noodzakelijk zijn om enige uitleg te verschaffen; het gevoel van de gemiddelde nederlandse bridger en de wiskunde staan zeer ver van elkaar vandaan.