Durft u het risico aan?

Hoe het uitkeringspercentage van de vernieuwde Nederlandse Lotto te schatten?

TON DIEKER EN HENK TIJMS

De Lotto is in Nederland een grote speler op de kansspelmarkt. Met onderdelen als Lotto, Lucky 10 en Krasloten haalt zij veel geld binnen. De Lotto heeft onlangs de opzet en het prijzenschema van het spel Lotto totaal vernieuwd. Wat is ongeveer het percentage van de inzet van de deelnemers dat de vernieuwde Lotto uitkeert aan prijzengeld? De Lotto beweert dat dit ongeveer 50 procent is, maar geeft geen informatie over de inleg en slechts gedeeltelijke informatie over de aantallen winnaars van de mogelijke prijzen. Niettemin is het uitkeringspercentage te schatten met elementaire kansrekening. Een leuke toepassing van kansrekening die ook bruikbaar is voor onderwijsdoeleinden.

De eerste trekking van de nieuwe Lotto op 10 juni 2000 tijdens het tv-programma *Lotto Weekend Miljonairs* leverde de eerste 800.000 winnende loten op. Waar met het 'oude' Lotto-spel een winkans gold van 1 op 40, geldt volgens de Lotto nu een winkans van 1 op 5,2. Voordien werden zes getallen uit de getallen 1, ..., 45 getrokken. In de nieuwe opzet worden zes (reguliere) getallen plus één bonusgetal uit de getallen 1, ..., 45 getrokken en daarnaast één kleur uit zes mogelijke kleuren. Ook het prijzenschema is vernieuwd (zie tabel 1). De jackpot wordt nu gewonnen als de zes getallen goed zijn in combinatie met de goede kleur. De jackpot begint met 4 miljoen gulden en wordt elke week dat de jackpot blijft staan, verhoogd met een half miljoen gulden. Voor een ingevuld rijtje dat meedingt naar de jackpot is de prijs verhoogd van *ƒ* 1 naar *ƒ* 1,50. Het is nog mogelijk rijtjes zonder kleur in te vullen voor f 1, maar daar wordt in de praktijk vrijwel geen gebruik van gemaakt. In het vervolg nemen we aan dat elk meespelend rijtje een *ƒ* 1,50-rijtje is (op de gratis *ƒ* 1-loten na).

Nr	Combinatie	Prijs	Kans (p_i)
0	Jackpot: 6 + kleur goed	4 miljoen of meer*	2.046E-08
1	6 goed	1 miljoen*	1.023E-07
2	5 + Bonusgetal + kleur goed	150.000*	1.228E-07
3	5 + Bonusgetal goed	100.000*	6.139E-07
4	5 + kleur goed	1.500	4.665E-06
5	5 goed	1.000	2.333E-05
6	4 + Bonusgetal + kleur goed	375	1.166E-05
7	4 + Bonusgetal goed	250	5.832E-05
8	4 + kleur goed	37.50	0.000216
9	4 goed	25	0.001079
10	3 + Bonusgetal + kleur goed	15	0.000288
11	3 + Bonusgetal goed	10	0.001438
12	3 + kleur goed	7.50	0.003452
13	3 goed	5	0.017262
14	2 + Bonusgetal + kleur goed	2 gratis loten met jackpot	0.002589
15	2 + Bonusgetal goed	2 gratis loten	0.012946
16	2 + kleur goed	1 gratis lot met jackpot	0.022656
17	2 goed	1 gratis lot	0.113282

* *prijs wordt bij meerdere winnaars gedeeld*

Tabel 1. Prijzenschema vernieuwde Lotto.

Uitkeringspercentage 50 procent?

Een interessante vraag is: wat is het percentage van de inleg dat gemiddeld als prijzengeld wordt uitgekeerd? De Lotto beweert dat dit ongeveer 50 procent is (*de Volkskrant*, 30 december 2000). Hoe kun je dit als buitenstaander controleren?

De Lotto verstrekt geen gegevens over de inzet van de deelnemers. De Duitse Lotto geeft op haar website hierover wel informatie. De enige informatie die de Nederlandse Lotto geeft, is de wekelijkse publicatie van het aantal winnaars van de jackpot, de hoofdprijs en van de tweede tot en met de vijfde prijs (voor het gemak nummeren we de 18 prijscategorieën in tabel 1 als prijs 0, prijs 1, ..., prijs 17, waarbij prijs 0 de jackpot is en prijs 17 twee goed is).

In deze bijdrage zullen we laten zien hoe het uitkeringspercentage geschat kan worden met behulp van de beschikbare informatie over de aantallen prijswinnaars. Elementaire maar interessante kansrekening volstaat voor dit leuke probleem.[1]

Aantal ingevulde rijtjes

In tabel 2 hebben we de gegevens verzameld over de aantallen prijswinnaars van de eerste zes van de 18 prijzen in de laatste 20 weken van het jaar 2000 en de eerste 5 weken van het jaar 2001. Het totale aantal winnaars van de eerste zes prijzen in deze 25 weken is gelijk aan $w = 2307$. Laat n het *onbekende* aantal rijtjes zijn dat over deze 25 weken is ingevuld. Laat p_i de kans zijn om de i-de prijs te winnen met n ingevuld rijtje ($i = 0, 1, ..., 17$). In eerste instantie nemen we aan dat elk rijtje een *ƒ* 1,50-rijtje is. Dan wordt de onbekende n geschat uit de vergelijking:

$$n\,(p_0 + p_1 + p_2 + p_3 + p_4 + p_5) = w \qquad (1)$$

De kansen p_i (i = 0, 1, ..., 17) worden in tabel 1 gegeven. Deze kansen zijn eenvoudig uit te rekenen met behulp van het hypergeometrische kansmodel met 6 rode ballen (reguliere getallen), 1 blauwe bal (bonusgetal) en 38 zwarte ballen (andere getallen). Bijvoorbeeld:

$$p_{10} = \frac{1}{6} \times \binom{6}{3}\binom{1}{1}\binom{38}{2} \bigg/ \binom{45}{6}$$

De relatie (1) geeft de schatting

$$n = 79\,960\,227$$

voor het totale aantal ingevulde rijtjes over de eerste 25 weken. Het is redelijk om

$$R = \frac{79\,960\,227}{25} = 3\,198\,409$$

als schatting te nemen voor het aantal rijtjes dat in een willekeurig gekozen week wordt ingevuld. Onder de voorlopige aanname dat elk rijtje een *f* 1,50-rijtje is, vinden we voor de wekelijkse inleg als eerste schatting:

$$f\,1{,}50 \times R = f\,4\,797\,614 \qquad (2)$$

De formules (1) en (2) zullen we later iets aanpassen om rekening te houden met het feit dat een aantal van de ingevulde rijtjes van gratis loten afkomstig is en het feit dat sommige van deze gratis loten niet voor alle prijzen meedingen.

Verwachte uitbetaling

Elk van de 18 prijzen uit tabel 1 geeft een bijdrage tot de verwachtingswaarde van de wekelijkse uitbetaling aan prijzengeld. Definieer de stochastische variabele X_i als:

X_i = het aantal winnaars van de i-de prijs in een willekeurig gekozen week.

Voor de prijzen i = 4, ..., 17 is de bijdrage tot de verwachte wekelijkse uitbetaling

$$E\,(X_i) \text{ x (geldbedrag van prijs } i) \qquad (3)$$

waarbij we voorlopig de afspraak maken om de als loten uitgekeerde prijzen 14, 15, 16 en 17 als de geldbedragen *f* 3, *f* 2, *f* 1,50 en *f* 1 te tellen. Later zullen we hiervoor corrigeren. Verder stellen we voorlopig

$$E\,(X_i) = R_{p_i} \text{ voor } i = 4, ..., 17 \qquad (4)$$

Ook zullen we straks corrigeren voor die prijzen i waarvoor de kleur meetelt omdat onder de R loten een aantal gratis loten van *f* 1 zit waarop alleen een prijs zonder kleur kan vallen. Voor de prijzen 0 (jackpot), 1 (hoofdprijs), 2 en 3 geldt dat deze door meerdere winnaars gedeeld worden. Voor i = 1, 2, 3 nemen we de volgende bijdrage tot de verwachte wekelijkse uitbetaling:

$$P\,(X_i > 0) \text{ x (geldbedrag van prijs } i\text{)} \qquad (5)$$

week	jackpot	hoofdprijs	2e prijs	3e prijs	4e prijs	5e prijs	
32	0	0	0	0	15	77	
33	1	0	0	1	7	105	
34	0	0	0	0	11	38	
35	0	0	0	0	13	113	
36	0	0	1	0	16	54	
37	0	0	0	2	17	77	
38	0	0	0	1	8	29	
39	0	0	0	2	9	84	
40	0	0	0	4	26	116	
41	0	1	0	1	13	60	
42	0	0	0	0	13	69	
43	0	1	0	1	13	81	
44	0	0	0	2	19	72	
45	0	0	1	3	18	61	
46	0	1	1	0	17	92	
47	0	0	0	0	6	46	
48	0	0	0	1	8	60	
49	0	0	0	0	7	72	
50	0	0	0	0	5	45	
52	0	0	1	4	21	156	
1	0	0	1	0	22	102	
2	0	0	0	0	11	74	
3	0	0	0	3	20	106	
4	0	0	0	1	9	52	
5	0	1	0	1	14	91	
Totaal	1	4	5	27	338	1932	2307

Tabel 2. Gepubliceerde aantallen winnaars.

Onder de aanname van *random* ingevulde rijtjes is $P(X_i = 0) = (1 - p_i)^R \approx e^{-Rp_i}$. Dit geeft de Poisson benadering

$$P(X_i > 0) \approx 1 - e^{-Rp_i} \text{ voor } i = 1, 2, 3 \tag{6}$$

in overeenstemming met het feit dat er sprake is van een zeer groot aantal (R) experimenten elk met een zeer kleine succeskans (p_i).

Kans op jackpot

Voor de jackpot geldt ook dat deze door meerdere winnaars gedeeld wordt, maar met de complicatie dat de jackpot geen vast bedrag is. De jackpot begint met 4 miljoen gulden en loopt elke week met een half miljoen gulden op zolang de jackpot niet valt. Noteren we met p_J de kans dat de jackpot valt in een willekeurig gekozen week, dan is een schatting voor de verwachte hoogte van die jackpot gelijk aan

$$\sum_{k=1}^{\infty}\left[4+\tfrac{1}{2}(k-1)\right]p_J(1-p_J)^{k-1}$$

$$=3.5+\frac{1}{2p_J} \text{ miljoen gulden.}$$

Voor het verwachte aantal weken dat het duurt voordat de jackpot valt nemen we als schatting de verwachtingswaarde $1/p_J$ van de geometrische verdeling $\{p_J\,(1-p_J)^{k-1},\ k \geq 1\}$. Dit geeft de schatting

$$\frac{3.5 + 1/(2p_J)}{1/p_J} \text{ miljoen gulden} \tag{7}$$

voor de gemiddelde bijdrage van de jackpot aan de wekelijkse uitbetaling. Resteert nog p_J te schatten. Hiervoor nemen we analoog aan (6) de Poisson benadering

$$p_J \approx 1 - e^{-Rp_i} \tag{8}$$

In de Lotto met ongeveer 3 miljoen rijtjes per week is de kans ongeveer 94% dat de jackpot blijft staan. Het is dan ook niet verwonderlijk als de jackpot een (te) groot aantal opeenvolgende weken blijft staan. De kans dat de jackpot niet valt in 23 of meer opeenvolgende weken is ongeveer gelijk aan $(0.94)^{23}$ oftewel ongeveer 24%. Een hoge jackpot bevordert deelname aan de Lotto, maar niet als deze gemiddeld slechts één keer in de 17 weken valt.

Random rijtje?

In de afleiding van de formules (6) en (8) hebben we impliciet aangenomen dat de deelnemers hun rijtjes *random* invullen. In werkelijkheid is dit niet het geval. Mensen gebruiken geboortedata, geluksgetallen, rekenkundige rijen, etc. om hun getallen te kiezen (vgl. Tijms, 1999). Het feit dat een groot deel van de rijtjes niet *random* wordt ingevuld, heeft gevolgen voor de kansverdeling van X_i (= het aantal winnaars van prijs i). De kansverdeling als geheel kan niet door een Poisson verdeling worden benaderd. Het gaat echter alleen om de kans $P(X_i = 0)$. Uit een empirische studie voor de California Lotto 6/53 beschreven door Kadell and Ylvisaker (1991) blijkt echter dat de Poisson benadering voor deze kans redelijk accuraat is (de empirisch bepaalde kans is iets groter dan de Poisson benadering). De uitdrukkingen (1) tot en met (8) leiden tot tabel 3 waarin

Prijs	Bijdrage (in guldens)	Prijs	Bijdrage (in guldens)
0	721 729	9	86 267
1	279 085	10	13 803
2	48 713	11	46 009
3	85 962	12	82 816
4	22 383	13	276 055
5	74 609	14	24 845
6	13 989	15	82 816
7	46 631	16	108 697
8	25 880	17	362 322

Tabel 3. Ongecorrigeerde bijdragen tot de uitbetaling.

voor de verschillende prijzen de (ongecorrigeerde) bijdragen tot de verwachte wekelijkse uitbetaling aan prijzengeld worden gegeven. Optellen van de bedragen in tabel 3 geeft de schatting *ƒ* 2 402 611 voor de verwachte wekelijkse uitbetaling aan prijzengeld. Dit leidt tot de eerste schatting:

$$\text{ongecorrigeerde uitkeringspercentage} = \frac{2\,402\,611}{4\,797\,614} \times 100\% = 50.1\%$$

Aanpassinguitkeringspercentage

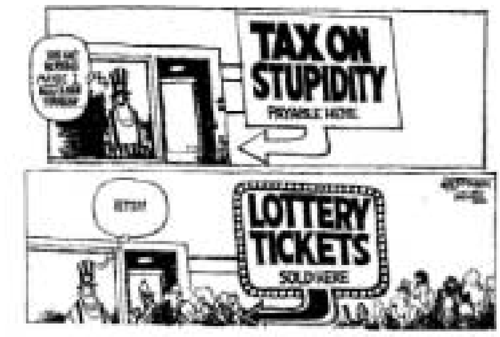
De berekening van bovenstaande schatting voor het uitkeringspercentage dient iets aangepast te worden. De wekelijkse inleg moet gecorrigeerd worden voor het feit dat een aantal van de ingevulde rijtjes afkomstig zijn van gratis loten die uitgekeerd zijn als de prijzen 14 t/m 17 (zie tabel 1). Verder is het zo dat de gratis *ƒ* 1-loten niet meedingen voor prijzen waarvoor ook de getrokken kleur goed moet zijn. Tenslotte hebben we de als gratis loten uitgekeerde prijzen 14 t/m 17 gerekend als uitbetalingen ter waarde van *ƒ* 3, *ƒ* 2, *ƒ* 1,50 en *ƒ* 1. Deze geldbedragen dienen uiteraard gecorrigeerd te worden met het uitkeringspercentage van de Lotto. Een en ander leidt tot de invoering van drie parameters die we iteratief gaan schatten. Deze parameters zijn

u = uitkeringsfractie
c_1 = fractie rijtjes van gratis loten
c_2 = fractie rijtjes van gratis *ƒ* 1-loten

De prijzen 15 en 17 zijn gratis *ƒ* 1-loten en deze loten dingen alleen mee voor prijzen waar de getrokken kleur er niet toe doet, oftewel de gratis *ƒ* 1-loten dingen niet mee voor de prijzen 0, 2, 4, 6, 8, 10, 12, 14 en 16. De kans dat op een *ƒ* 1-rijtje prijs i met i oneven valt is uiteraard gelijk aan $p_{i-1} + p_i$. De formules (1) tot en met (8) vereisen nu een kleine aanpassing. In formule (1) moeten we np_i vervangen door $n(1\text{-}c_2)p_i$ voor i = 0, 2, 4 en door $n(1\text{-}c_2)p_i + nc_2(p_{i-1} + p_i)$ voor i = 1, 3, 5. Opgeteld geeft dit echter weer dezelfde vergelijking (1), hetgeen achteraf ook logisch is. In formule (2) moeten we R vervangen door $R(1\text{-}c_1)$, terwijl we R in formule (4) moeten vervangen door

Prijs	Bijdrage (in guldens)	Prijs	Bijdrage (in guldens)
0	690 869	9	88 737
1	285 808	10	11 827
2	42 856	11	47 326
3	86 729	12	70 960
4	19 178	13	283 959
5	76 746	14	11 111
6	11 987	15	44 461
7	47 966	16	48 609
8	22 175	17	194 517

Tabel 4. Gecorrigeerde bijdragen tot de uitbetaling.



Illustratie: Jim Borgman.

$R(1\text{-}c_2)$ voor die prijzen i met i even. In formule (6) met i = 2 en in formule (8) wordt R ook vervangen door $R(1\text{-}c_2)$. In formule (4) met i oneven en in formule (6) met i oneven moet R_{p_i} door $R(1\text{-}c_2)p_i + Rc_2(p_{i-1}+p_i)$ vervangen worden. Tenslotte in formule (3) wordt voor i = 14, 15, 16 en 17 in plaats van de geldbedragen *ƒ* 3, *ƒ* 2, *ƒ* 1,50 en *ƒ* 1 de geldbedragen $u \times ƒ3$, $u \times ƒ2$, $u \times ƒ1{,}50$ en $u \times ƒ1$ genomen. Bij het berekenen van het ongecorrigeerde uitkeringspercentage nemen we in eerste instantie $u = 1$ en $c_1 = c_2 = 0$. De eerste ronde van de berekeningen leiden tot de nieuwe schattingen $u = 0.5008$, $c_1 = 0.1670$ en $c_2 = 0.1392$. Vervolgens herhalen we de berekeningen op basis van de aangepaste formules (1) tot en met (8) met deze schattingen. Dit leidt tot nieuwe schattingen voor u, c_1 en c_2. Vervolgens herhalen we de berekeningen met deze nieuwe schattingen net zo lang tot de schattingen voor u, c_1 en c_2 voldoende geconvergeerd zijn. Dit gebeurde al na vier iteratiestappen en de uiteindelijke schattingen van u, c_1, c_2 en n zijn $u = 0.5219$, $c_1 = 0.1670$, $c_2 = 0.1432$ met n = 79 960 227. Ter controle: de verwachte waarden

$$n\,(1 - c_2)\,p_4 = 320$$
$$n\,(1 - c_2)\,p_5 + nc_2\,(p_4 + p_5) = 1919$$

voor de totale aantallen winnaars van de prijzen 4 en 5 stemmen overeen met de werkelijke waarden 338 en 1932 uit tabel 1. In tabel 4 geven we de uiteindelijke bijdragen van de verschillende prijzen aan de verwachtingswaarde van de wekelijkse uitkering van de lotto.

Gecorrigeerde uitkeringspercentage

De verwachte waarde van de totale uitkering in een week is 2 085 882 gulden. Voor de gemiddelde wekelijkse inleg vonden we uiteindelijk op grond van de aangepaste versie van formule (2) het inlegbedrag *ƒ* 3 996 365. Dit geeft de schatting

$$\text{gecorrigeerde uitkeringspercentage} = \frac{2\,085\,882}{3\,996\,365} \times 100\% = 52.2\%$$

Dit is een iets hoger percentage dan het ongecorrigeerde uitkeringspercentage van 50.1. Het werkelijke uitkeringspercentage wordt echter naar beneden beïnvloed door twee andere factoren. Ten eerste het feit dat het merendeel van de rijtjes met de hand en dus niet random worden ingevuld. Dit betekent dat de werkelijke kans dat de jackpot en andere grote prijzen vallen, kleiner is dan in de modelsituatie van *random* ingevulde rijtjes (om dit in te zien denk het extreme geval in dat alle deelnemers hetzelfde rijtje kiezen, in welk geval er praktisch gesproken nooit winnaars van de hoogste prijzen zullen zijn). Een ander aspect dat het uitkeringspercentage iets zal drukken, is het feit dat niet iedere winnaar van de gratis loten deze loten ook werkelijk hergebruikt. Het is een bekend verschijnsel dat niet alle prijzen geclaimd worden. In de Amerikaanse *Powerball Lottery*, de grootste lotto ter wereld, wordt 30% van de kleinste prijzen en 12% van het totale prijzengeld niet opgeëist. Zou slechts 75% van de gratis loten ingeleverd worden, dan zou het uitkeringspercentage van 52.2 dalen naar 49.6.

Conclusie

Onze berekeningen bevestigen de uitspraak van de Lotto dat zij ongeveer de 50% van de inleg als prijzengeld uitkeert aan de deelnemers (een lager percentage aan prijzengeld dan het uitkeringspercentage van meer dan 65 bij de Staatsloterij,

maar een stuk hoger dan het uitkeringspercentage van nog geen 25 bij de Nationale Postcode Loterij). Wat we echter niet kunnen onderschrijven, is een eerdere reclameslogan van de Lotto: 'Lotto de beste kans om miljonair te worden.' De kans om met één rijtje de jackpot te winnen is 2.05×10^{-8}. Deze kans is slechts 2/3 van de kans om bij 25 worpen met een zuivere munt alleen maar kop te gooien. Zou je elke week trouw 12 rijtjes invullen, dan zou je meer dan 54 duizend jaar van leven moeten hebben om met een kans van tenminste 50% ooit in je leven een keer de jackpot te winnen! Dit laatste is eenvoudig na te rekenen met het Poisson kansmodel waarvan vele aardige toepassingen gegeven worden in de hieronder genoemde literatuur.

Noot

1 Het materiaal in deze bijdrage is erg geschikt voor onderwijsdoeleinden, waarbij ook gedacht wordt aan praktische opdrachten in het Studiehuis op het vwo. De Lotto is bij uitstek geschikt voor didactische doeleinden in de kansrekening, zie ook Tijms (1999) voor ander lesmateriaal over loterijen.

Literatuur

Kadell, D. and Ylvisaker, D. (1991). Lotto play: the good, the fair and the truly awful. *Chance*, 4, 22-25.

Tijms, H.C. (1999). *Spelen met Kansen*. Utrecht: Epsilon Uitgaven.

Ton Dieker en Henk Tijms zijn werkzaam bij de Afdeling Econometrie en Operationele Research van de Vrije Universiteit Amsterdam, <tijms@econ.vu.nl>.