

Peter Sander

Faculteit Technologie Management
Technische Universiteit Eindhoven
Postbus 513
5600 MB Eindhoven
p.c.sander@tm.tue.nl

Afscheidsrede

Van doodtrap

Hoe kan een consument er zeker van zijn dat een nieuw aangeschaft apparaat goed functioneert? Kan hij er vanuit gaan, dat het allemaal goed werkt, of doet hij er beter aan meteen een verzekering af te sluiten tegen defecten, zoals bijvoorbeeld bij digitale camera's heel gebruikelijk is? Op 10 oktober 2004 heeft Peter Sander afscheid genomen van de Technische Universiteit Eindhoven. Hij is vanaf 1977 verbonden geweest aan deze universiteit, waar hij eerst lector en later hoogleraar Toepassingen van de Statistiek is geweest. Zijn interesse gaat in het bijzonder uit naar de bijdrage van de statistiek aan het ontwerpen, fabriceren en instandhouden van industriële producten.

Hij begon met planten en dieren, met al wat om hem was — en als hij er lang op gestaard had werden het cijfers. Alles viel uiteen tot cijfers, — bladen vol cijfers. Dat vond docter Cijfer heerlijk, en hij zeide, dat het hem licht werd, als de cijfers kwamen, — doch voor Johannes was dat duisternis.

Uit: De Kleine Johannes door Frederik van Eeden (1870–1930)

Toen ik in de zestiger jaren in Utrecht wiskunde studeerde, had ik het voorrecht om bij professor Freudenthal (1905–1990) het college Waarschijnlijkheidsrekening en Statistiek te volgen. College lopen bij Freudenthal was een genoegen. Niet omdat hij volgens de gebruikelijke maatstaven een briljant docent was, maar wel omdat hij de wiskunde voor je ogen liet ontstaan, inclusief de onvermijdelijke missers en doodlopende wegen. Bij Freudenthal gebeurde altijd wat. Ook op mondelinge tentamens, zoals ik uit eigen ervaring kan vertellen. Freudenthal gaf onder andere het vak Grondslagen van de meetkunde. Ik had een afspraak gemaakt voor een mondelinge tentamen en dat liep heel redelijk. Tot hij

weer een nieuw probleem aan de orde stelde en zijn betoog uiteraard afrondde met een vraag. Ik had zijn redenering echter absoluut niet begrepen en dat liet ik mij ook spontaan ontvallen. Hierop liet hij zich achterover vallen in zijn stoel, dacht een paar minuten na, in elk geval leek het zo lang, en zei toen: “ja, ja, ja, dat is ook allemaal onzin”. Daarna stelde hij nog een vraag over heel iets anders en liet me toen vertrekken met een negen. Tot op de dag van vandaag heb ik het gevoel dat mijn onbegrip meer aan mij lag dan aan hem.

Als opening van dit afscheidscollege wil ik u graag een voorbeeld geven uit het college van Freudenthal. Het aardige van dit voorbeeld is zowel gelegen in de oorsprong van de waarnemingen, als in het feit dat de waarnemingen perfect kunnen worden gemodelleerd met een van de beroemdste kansverdelingen uit de statistiek, namelijk de Poissonverdeling, genoemd naar de Franse wiskundige Siméon-Denis Poisson, die leefde van 1781–1840. De praktische waarde van de methode zal u aan het einde van het college duidelijk worden.

Poisson was geïnteresseerd in het gebruik

van kansrekening bij de rechtsspraak en daarbij leidde hij de later naar hem genoemde verdeling af [6]. De Poissonverdeling bleef in de kast liggen tot Ladislaus Josephowitsch von Bortkiewicz, een in Rusland geboren Duitse hoogleraar van Poolse afkomst, hem er in 1898 uithaalde. Von Bortkiewicz modelleerde een onalledaagse situatie met behulp van de Poissonverdeling. Het gaat over het aantal soldaten uit het Pruisische leger dat per jaar en per troepenonderdeel overleed ten gevolge van een trap van een paard. De aantallen doden zijn gebaseerd op 10 korpsen over een periode van 20 jaar, namelijk van 1875–1894 [9]. In totaal werden 122 soldaten door een paard doodgetrapt. Tabel 1 geeft de aantallen.

Von Bortkiewicz vroeg zich af of de Poissonverdeling als kansmodel zou passen bij de waargenomen aantallen doden. De Poissonverdeling is niets anders dan een formule waarmee onder bepaalde voorwaarden de kansen op 0, of 1, of 2, enzovoorts successen kunnen worden berekend. In het voorbeeld identificeren we een geslaagde doodtrap als een succes. De formule luidt:

$$P(k \text{ successen}) = e^{-\lambda} \frac{\lambda^k}{k!} \text{ voor } k = 0, 1, 2, \dots$$

Hierin is $e = 2.718281828 \dots$, een notatie die in 1731 door Euler werd ingevoerd.

Voor de parameter λ kunnen we in beginsel een willekeurig positief getal kiezen, maar er zijn theoretische argumenten om in ons



Peter Sander

tot erger

voorbeeld λ gelijk te kiezen aan het gemiddeld aantal doden per korpsjaar. Daar er 200 korpsjaren zijn en 122 doodgetrapte soldaten, zijn er gemiddeld $122 : 200 = 0.61$ doden per korpsjaar.

Nu we λ gelijk aan 0.61 hebben gekozen, kunnen we met behulp van de Poissonverdeling voor elke waarde van k de kans uitrekenen op k doodgetrapte soldaten. Als we deze kansen vermenigvuldigen met het aantal korpsjaren, dus met 200, dan vinden we de verwachte frequenties. Tabel 2 geeft de resultaten.

Als we de verwachte frequenties vergelijken met de werkelijke frequenties, dan zien we dat de getallen sprekend op elkaar lijken. We concluderen hieruit, in navolging van Von Bortkiewicz, dat de werkelijke aantallen doodgetrapte Pruisische soldaten prima beschreven kunnen worden met een Poissonverdeling.

Ik kan me voorstellen dat u de overeenkomst tussen de waargenomen frequenties en de verwachte frequenties zo goed vindt, dat u zich afvraagt of Von Bortkiewicz met de getallen heeft geknoeid. Wel, er zijn in de literatuur inderdaad artikelen te vinden waarin Von Bortkiewicz wordt beticht van vals spel. In werkelijkheid waren er namelijk 14 legerkorpsen, maar hij rapporteert slechts over 10 korpsen. Ra, ra, ra, waar zijn die vier korpsen gebleven?

De eerste die hier over viel was Whitaker [12]. Von Bortkiewicz antwoordde dat die vier korpsen niet waren meegenomen, om-

dat deze een geheel andere organisatie hadden dan de beschreven korpsen [10]. Zo telde elk van de beschreven korpsen 25 squadrons, terwijl een van de weggelaten korpsen maar liefst 40 squadrons telde. Het argument van Von Bortkiewicz snijdt natuurlijk hout. Stel namelijk dat die vier extra korpsen veel groter zijn dan die andere 10; zelfs zo groot dat in elk van die vier korpsen gemiddeld wel 100 doden per jaar vallen. Dan geven dus 10 korpsen elk hooguit vijf doden per jaar, en vier korpsen elk zo'n 100 doden per jaar. Dit past natuurlijk in het geheel niet bij een Poissonverdeling.

Dames en heren, met dit voorbeeld heb ik u laten zien dat kansmodellen de potentie hebben om werkelijke situaties uitstekend te beschrijven. Het is een van de eerste voorbeelden die bewust mijn interesse wekten in de toepasbaarheid van kansmodellen. Lang daarvoor waren mijn ervaringen met kansen eerder teleurstellend van aard, als ik voor de zoveelste keer bij *Mens Erger Je Niet* op een cruciaal moment weer eens geen zes had goegoid. De frustratie van mijn opa was overigens nog een maat groter: met enige regelmaat vernielde hij uit pure woede op zaterdagavond het spel, waarna hij op maandag weer een nieuw kocht.

Het doel van de statistiek

Van oudsher is de statistiek een hulpmiddel bij het nemen van beslissingen. Vanuit deze invalshoek is de definitie van Barnett aantrekkelijk [2]: "Statistics is the study of how infor-

mation should be employed to reflect on, and give guidance for action in, a practical situation involving uncertainty."

Dit betekent dat wij een verschijnsel waarnemen (information), dat de waarnemingsuitkomsten ons niet al vooraf bekend zijn (uncertainty), en dat die waarnemingen ons inzicht verhogen (reflect on) en van nut zijn bij het nemen van maatregelen (give guidance for action).

Omdat deze definitie de indruk kan wekken dat de statistiek pas een rol gaat spelen als de waarnemingen zijn gedaan, voeg ik er gelijk aan toe dat de statistiek in het bijzonder waardevol is bij het structureren van problemen en het opzetten van experimenten. Als een probleem niet goed is gestructureerd en de experimenten niet goed zijn opgezet en uitgevoerd, dan is het erg optimistisch om er op te vertrouwen dat de statistiek toch nog wel kans zal zien om met slimme rekenpartijen allerlei zinvolle verschijnselen op te sporen die

Aantal doodgetrapte soldaten per korpsjaar	Waargenomen frequentie
0	109
1	65
2	22
3	3
4	1
≥ 5	0
Totaal	200

Tabel 1 Frequentieverdeling van het aantal doodgetrapte Pruisische soldaten per korpsjaar

Aantal doodgetrapte soldaten per korpsjaar	Waargenomen frequentie	Frequentie volgens Poissonverdeling
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6
≥ 5	0	0.1
Totaal	200	200

Tabel 2 Vergelijking tussen de waargenomen frequenties en de verwachte frequenties

in de getallenbrij zijn verborgen.

Veelal wordt de statistiek gebruikt als beslissingen moeten worden genomen over een populatie — en daarmee bedoel ik een verzameling van elementen waarop men de conclusies van het onderzoek wil betrekken — terwijl onvoldoende informatie beschikbaar is. Met behulp van de statistiek kan dan worden gepoogd tegen zo gering mogelijke kosten voldoende informatie over die populatie in handen te krijgen om een verantwoorde beslissing te kunnen nemen. Daartoe wordt een steekproef uit de populatie onderzocht en vervolgens wordt gepoogd om de toevallige uitkomsten van de steekproef te generaliseren tot uitspraken over de populatie. Laat ik dit toelichten.

Voorbeeld 1

Als een fabrikant zich afvraagt of een partij lucifers al dan niet aan de eisen voldoet, dan ligt het voor de hand om een willekeurige steekproef uit de productie te nemen en voor elk van de lucifers uit de steekproef vast te stellen of hij goed brandt. De lucifers uit de steekproef zijn na de test alleen geschikt om weggegooid te worden. Het restant van de populatie lucifers is wel relevant en over de kwaliteit van die lucifers kunnen we mogelijk iets zeggen op grond van de uitkomst van de steekproef. Als van de 100 lucifers in de steekproef er 3 niet voldoen, dan verwachten we in de populatie ook ongeveer 3% lucifers van onvoldoende kwaliteit.

De steekproef mag natuurlijk niet worden gemanipuleerd.

Voorbeeld 2

Als u zaterdagochtend op de markt 3 kilo appels koopt, en de marktkoopman weegt die 3 kilo voor u af, dan hoopt u dat die prachtige appels aan uw kant van de kraam een eerlijk beeld geven van de appels die de koopman aan zijn kant van de kraam afweegt.

De statistiek stelt een onderzoeker in staat om:

- Informatie over een te onderzoeken verschijnsel op een efficiënte en effectieve manier te verzamelen, onder meer door de goede experimenten te doen.
- Verzamelde informatie op een zodanige wijze te bewerken, dat zinvolle conclusies mogelijk zijn over de populatie waaruit de informatie afkomstig is.
- De opzet van een onderzoek, de meetuitkomsten en de conclusies op een inzichtelijke wijze te presenteren met behulp van grafieken, tabellen en kengetallen.

Het aardige van de statistiek zit hem in het feit dat we alleen kijken naar situaties waarin toeval een belangrijk rol speelt. Immers: als je alles van tevoren weet, dan is er geen aardigheid meer aan.

Toepassen van statistiek

Zoals Freudenthal mijn interesse in de statistiek heeft gewekt, zo heeft professor Leppink mij het grote verschil geleerd tussen theoretische kennis hebben en die kennis kunnen toepassen. Leppink was hoogleraar Mathematische Statistiek en de oprichter van het Instituut voor Mathematische Statistiek van de Rijksuniversiteit Utrecht. Een van de doelen van dat instituut was om het gebruik van de statistiek door onderzoekers van de universiteit aan te moedigen. Dat heeft heel wat jaren perfect gewerkt. Hierdoor werden de medewerkers van het instituut betrokken bij de meest uiteenlopende onderzoeken. Zij dienden dan ook een brede belangstelling te hebben, want statistiek toepassen zonder de moeite te nemen je in te werken in het vakgebied van de onderzoeker, is een kansloos gebeuren. Dan neem je je vak niet serieus. Ik zal nu aan de hand van enkele aspecten toelichten wat het toepassen van statistiek zo boeiend maakt.

Modelleren van het probleem

Dit is een van de boeiendste en ook moeilijkste facetten uit het leven van een statisticus. Toepassen van statistiek betekent dat

het probleem van de onderzoeker gewoonlijk moet worden aangescherpt en vervolgens vertaald in een kwantitatief model dat zich leent voor een statistische analyse. Bij deze vertaling wordt het oorspronkelijke probleem ontleed in relevante en minder relevante aspecten. Door de minder relevante aspecten buiten beschouwing te laten, wordt het probleem behapbaar. Dit is een delicaat gebeuren, want als relevante aspecten ten onrechte worden weggelaten, dan representeert het kwantitatieve model de werkelijkheid in onvoldoende mate. Anderzijds, als niet relevante aspecten worden meegenomen in het model, dan wordt het model nodeloos gecompliceerd.

Uiteraard laat de statisticus zich bij het bouwen van een model mede sturen door zijn theoretische kennis van de statistiek, zeg maar: door alle modellen die hij kent; maar hij mag natuurlijk het oorspronkelijke probleem geen geweld aan doen. Zo nodig moeten nieuwe terreinen worden bewandeld en nieuwe statistische theorieën worden ontwikkeld.

Als het model eenmaal is opgesteld, dan wordt veelal een experiment uitgevoerd. Vervolgens worden de meetuitkomsten geanalyseerd en gepresenteerd. Tenslotte worden de resultaten van de statistische analyse vertaald naar het oorspronkelijke probleem.

In het bijzonder bij de volgende twee vertaalslagen is een intensieve samenwerking tussen de onderzoeker en de statisticus noodzakelijk:

- van het probleem naar het statistische model
- van de uitkomst van de statistische analyse naar het antwoord op de oorspronkelijke vraag.

Een onderzoeker heeft gewoonlijk onvoldoende kennis om de mogelijkheden van de statistiek optimaal te benutten, en een statisticus die denkt dat hij zelfstandig het probleem van de onderzoeker kan vertalen in een statistisch model, zelfstandig kan besluiten welke statistische analyses zinvol zijn, en zelfstandig de uitkomsten kan terugvertalen naar het oorspronkelijke probleem, heeft er weinig van begrepen.

Experimenteren

Zonder twijfel is het grootste nut van de statistiek gelegen in de besparing in tijd en geld die kan worden bereikt door experimenten op te zetten die het gewenste inzicht verschaffen met behulp van een minimum aan inspanning. Een niet efficiënt experiment leidt mo-

gelijk wel tot de juiste inzichten, maar tegen onnodig hoge kosten. Zo denken veel onderzoekers dat de invloed van meerdere factoren moet worden onderzocht door achtereenvolgens elk van die factoren te variëren bij constant houden van de andere factoren. Dit is echter meestal een weinig gelukkige methode. Deze methode leidt tot nodeloos veel experimenten, en veelal tot verkeerde conclusies.

Voorbeeld 3

Stel dat de opbrengst van een chemisch proces wordt beïnvloed door twee factoren: de temperatuur en de reactietijd. Het zou kunnen zijn dat onafhankelijk van de temperatuur de opbrengst bij een reactietijd van 6 minuten 10 eenheden hoger is dan bij een reactietijd van 2 minuten. Zie figuur 1.

Het zou echter ook kunnen zijn dat er een meer complexe relatie is tussen reactietijd, temperatuur en opbrengst. Het zou kunnen zijn dat een reactietijd van 2 minuten een relatief hoge opbrengst geeft bij een temperatuur van 80 °C, maar een relatief lage opbrengst bij een temperatuur van 40 °C, terwijl dit bij een reactietijd van 6 minuten juist andersom is. Zie figuur 1.

Wie in een dergelijke situatie op een efficiënte wijze de maximale opbrengst wil verkrijgen, doet er wijs aan de reactietijd en de temperatuur op een intelligente wijze simultaan te variëren.

Mijn ervaring is dat, in het bijzonder in het bedrijfsleven, heel wat experimenten worden gedaan zonder dat men serieus heeft overdacht wat het probleem precies is en welk experiment verstandig lijkt. Na afloop van het experiment komt men dan vaak tot de conclusie dat niet het gewenste resultaat is bereikt. Kennelijk neemt men het woord nadenken veel te letterlijk. Men denkt na, nadat het experiment is uitgevoerd. Men zou vóór het uitvoeren van het experiment moeten denken. Ik propageer om naast het woord nadenken ook het woord vóórdenken te gebruiken. In mijn onderwijs poog ik het voorddenken te stimuleren door het luiheidprincipe te hanteren: gij zult slechts handelen nadat gij hebt vastgesteld dat die handeling een relevante bijdrage levert aan de oplossing van het probleem. Bij twijfel: denk! Studenten houden van dit principe, zij gebruiken het graag, in het bijzonder als ik ze vertel dat de voortgang van hun werk mij tegenvalt.

Metten

Experimenteren betekent ook metten. Als je niet nauwkeurig genoeg kunt metten, dan kun

je ook niet vaststellen of waargenomen verschillen aan de meetmethode liggen of aan de situatie. Hoe nauwkeurig moet worden gemeten, hangt af van de doelstelling.

Voorbeeld 4

Op de Olympische winterspelen van 2002 in Salt Lake City schaatste Jochem Uytdehaage een wereldrecord op de 10 km heren: 12 min 58.92 seconde. De tijd wordt gemeten in honderdsten van een seconde. In 0.01 seconde reed Jochem ruim 12.8 centimeter. De 10 km bestaat uit 25 rondjes van 400 meter, dus de lengte van de baan mag hooguit 5 millimeter afwijken van 400 meter, want anders klopt de 10 km tijd niet meer. Die nauwkeurigheid is niet van belang als we alleen willen bepalen wie de gouden medaille moet krijgen, maar wel als we de tijd als wereldrecord willen erkennen.

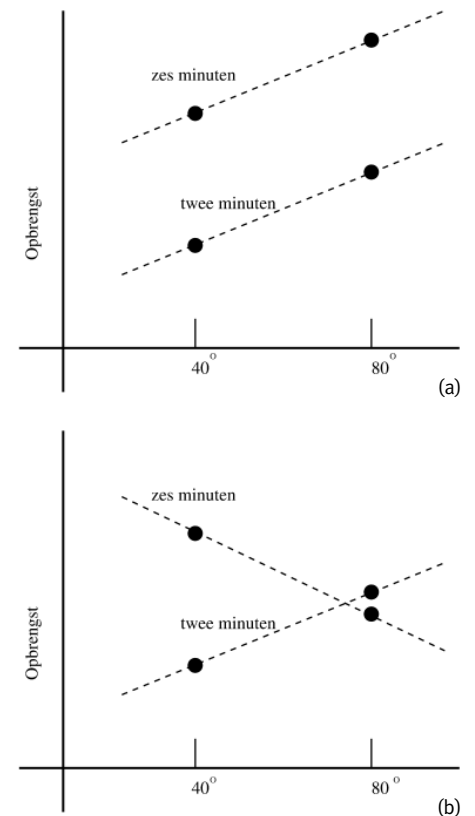
Een heel ander probleem doet zich voor bij het meten van complexe begrippen, daarmee bedoel ik begrippen die zijn opgebouwd uit verschillende kenmerken.

Voorbeeld 5

Op diezelfde Olympische winterspelen werd in een andere hal gestreden om de gouden medaille bij het paardrijden, naar mijn smaak een van de aardigste onderdelen van het kunstschaatsen. Om voor de hand liggende redenen is er bij dat nummer nogal eens verschil van mening over welk paar de gouden medaille verdient. In 2002 had het Russische paar Jelena Bereznaja en Anton Sikhuralidze met vijf stemmen tegen vier de voorkeur van de jury, maar onder druk van de publieke opinie werd ook aan het Canadese paar Jamie Sale en David Pelletier een gouden medaille uitgereikt. Hoe moeilijk is meten en hoe nauwkeurig lukt dat?

Analyseren en presenteren van meetuitkomsten

Als de meetuitkomsten beschikbaar zijn, dan volgt de analyse. Gezien de overvloed aan statistische programmatuur is het bijzonder verleidelijk om allerlei dingen uit te rekenen. Zonder inzicht in achter de getallen verborgen structuren kunnen zulke rekenpartijen tot volstrekt verkeerde gedachten leiden. Daarom hoort elke analyse te beginnen met een visuele analyse: grafieken. Kwantitatieve analysetechnieken komen pas aan bod ná een grafische analyse. In elke complexe situatie zijn grafieken onmisbaar. Daarom zit er bij een doos vol planken, schroeven en moeren altijd een werktekening, want met alleen tekst krijgt uw buurman de kast nooit in elkaar.



Figuur 1 De relatie tussen de temperatuur en de opbrengst is onafhankelijk (a) respectievelijk afhankelijk (b) van de reactietijd.

Ik gebruik een beroemde dataset om u duidelijk te maken dat een grafische weergave de essentie van waarnemingen veelal helderder kan weergeven dan tabellen. Tabel 3 geeft de dataset.

Aan zo'n tabel met getallen valt niet veel af te lezen, daarom is het gebruikelijk om sets met getallen samen te vatten in enkele kernwaarden. In dit geval doe ik dat door u te vertellen dat voor elk van de vier paren (x_1, y_1) , (x_1, y_2) , (x_1, y_3) , (x_2, y_4) het volgende geldt:

Aantal waarnemingen:	11
Gemiddelde van de x -waarden:	9.0
Gemiddelde van de y -waarden:	7.5
Standaardafwijking van de x -waarden:	3.3
Standaardafwijking van de y -waarden:	2.0
Regressielijn:	$y = 3 + 0.5x$
Correlatie tussen x en y :	0.82

De paren worden in figuur 2 grafisch weergegeven. De figuur toont verschijnselen die uit tabel 3 niet zijn af te lezen.

Het zal u duidelijk zijn: grafieken zijn onmisbaar om een goed inzicht te krijgen in complexe verschijnselen en om veronderstellingen te verifiëren die in het gebruikte model zijn besloten.

x_1	y_1	y_2	y_3	x_2	y_4
10	8.04	9.14	7.46	8	6.58
8	6.95	8.14	6.77	8	5.76
13	7.58	8.74	12.74	8	7.71
9	8.81	8.77	7.11	8	8.84
11	8.33	9.26	7.81	8	8.47
14	9.96	8.10	8.84	8	7.04
6	7.24	6.13	6.08	8	5.25
4	4.26	3.10	5.39	19	12.50
12	10.84	9.13	8.15	8	5.56
7	4.82	7.26	6.42	8	7.91
5	5.68	4.74	5.73	8	6.89

Tabel 3 De dataset van Anscombe [1]

Terugvertalen naar de praktijk

Naar mijn mening is de meest voorkomende fout bij het toepassen van statistische methoden, dat uit het oog wordt verloren waarvoor deze methoden dienen. Dit wordt in de hand gewerkt door de wijze waarop auteurs van leerboeken statistische methoden en technieken behandelen. De nadruk ligt in de leerboeken, de goede uitzonderingen daargelaten, altijd op de statistiek en niet op het oorspronkelijke probleem.

Onderzoekers stellen gewoonlijk hypothesen op die met behulp van statistische methoden worden geverifieerd. Daarbij volgt men trouw een rekenschema en concludeert aan het einde dat de uitkomst al dan niet significant is, en als gevolg hiervan wordt een hypothese wel of niet verworpen. Menig onderzoeker vergeet daarbij dat er twee soorten significantie bestaan: statistische significantie en praktische significantie; en deze twee hebben niets met elkaar te maken. Elk verschil, hoe klein ook, is statistisch significant als er voldoende waarnemingen worden gedaan. In de praktijk is een verschil echter alleen een verschil als het verschil maakt.

In wetenschappelijke tijdschriften is met regelmaat te zien dat voor de praktijk van alledag irrelevante zaken zinloos in de belangstelling staan doordat de uitkomsten van een experiment statistisch significant zijn bij de gehanteerde steekproefomvang.

Diezelfde tijdschriften doen vermoeden, dat talloze voor de praktijk van alledag interessante verschijnselen òn voldoende belangstelling krijgen doordat de uitkomsten van de betreffende experimenten niet statistisch significant zijn bij de gegeven steekproefomvang en de gekozen statistische analysemethode.

Gebruik van voorkennis

Een onderzoeker begint een onderzoek altijd

met een bepaald beeld van de werkelijkheid. Soms is dat beeld nog vaag en soms is het al vrij scherp. Onderzoek is altijd geënt op bestaande kennis en levert aanvullende informatie. Na het onderzoek trekt de onderzoeker conclusies op grond van zijn subjectieve ervaring aangevuld met de resultaten van zijn laatste onderzoek. Savage [8] geeft drie beroemde voorbeelden om het gebruik van subjectieve kennis van de onderzoeker bij het interpreteren van waarnemingen te verduidelijken. Ik geef deze voorbeelden hier kort weer.

- In het eerste voorbeeld refereert Savage aan een voorbeeld van Fisher [4], waarin een mevrouw claimt dat zij kan proeven of eerst de melk of eerst de thee werd ingeschonken. In een experiment waarin zij tien kopjes thee krijgt aangeboden, geeft zij voor elk kopje thee inderdaad correct de volgorde aan waarin de melk en de thee werden ingeschonken.
- Het tweede voorbeeld betreft een musicus die beweert dat hij aan een bladzijde uit een partituur kan zien of het van Haydn of van Mozart is. De musicus wijst in een experiment inderdaad voor elk van 10 bladzijden correct de componist aan.
- In het derde voorbeeld verklaart een lichtelijk beneveld persoon dat hij kan voorspellen of de uitkomst van een worp met een muntstuk kruis of munt oplevert. De man krijgt de kans om zijn bewering hard te maken met behulp van een experiment. Hij voorspelt voor 10 achtereenvolgende worpen de uitkomst correct.

Voor elk van de drie voorbeelden geldt dat als de hoofdpersoon domweg gokt, de kans op 10 maal correct beslissen gelijk is aan ongeveer één promille. Op grond van extra — subjectieve — informatie hecht Savage aan de uitkomsten van de experimenten echter geheel verschillende waarde. In de situatie van de theedrinkende dame is hij niet overtuigd, maar sluit hij niet uit dat de volgorde inderdaad invloed heeft op de smaak. Dat een musicus aan de partituur kan zien of deze afkomstig is van Haydn of van Mozart, lijkt hem vanzelfsprekend en de uitkomst van het experiment bevestigt zijn mening. Dat de benevelde man 10 maal correct de uitkomst van een worp met een munt voorspelt, vindt Savage puur dronkemansgeluk, hij hecht daar in het geheel geen waarde aan.

De klassieke statistiek kan met deze voorbeelden niet goed uit de voeten. De klassieke statistiek poogt zich te beperken tot het analyseren van de uitkomsten van experimenten, zonder gebruik te maken van additionele informatie. Niettemin gebruikt de klassieke

statistiek de ervaring van onderzoekers toch, maar slechts gedeeltelijk. Wél wordt het type kansverdeling waaruit de waarnemingen afkomstig zijn veelal bekend verondersteld, bijvoorbeeld een Poissonverdeling, maar er wordt bijvoorbeeld géén kennis gebruikt over de vermoedelijke waarde van de parameters die in het model voorkomen, zoals de λ in de Poissonverdeling. Uitspraken over die parameters mogen alleen zijn gebaseerd op eigen experimenten.

Daar statistici zich de beperkingen van de klassieke statistiek natuurlijk uitstekend realiseren, is er een theorie ontwikkeld, de zogenaamde Bayesiaanse statistiek, die bedoeld is om op een consistente wijze de bestaande inzichten te kunnen bijstellen als nieuwe informatie beschikbaar komt. Zeker in vakgebieden waarin het doen van waarnemingen een kostbare zaak is, is een Bayesiaanse aanpak onmisbaar. Tenslotte wordt het maatschappelijk niet aanvaard als een olie-maatschappij 10 raffinaderijen laat ontploffen enkel en alleen om de statistisch gewenste steekproefgrootte te halen. Het ligt dan meer voor de hand om naast de waarnemingen ook de kennis van experts te gebruiken. Dit is precies wat de Bayesiaanse statistiek beoogt.

Kortom: alleen op de beurs mag voorkennis niet worden gebruikt.

Bedrijfszekerheid

Als u een product koopt, dan verwacht u dat dit product gedurende een redelijke tijd, gewoonlijk substantieel langer dan de garantieperiode, naar genoegen functioneert. De wet dwingt fabrikanten om uw verwachtingen te honoreren. Ik zal eerst ingaan op het begrip bedrijfszekerheid en daarna wijzen op de bescherming die de wet aan consumenten biedt in geval de leverancier geen deugdelijk product levert.

Al kort na mijn benoeming aan deze universiteit heb ik mij gericht op het gebruik van statistiek bij het ontwerpen, produceren en instandhouden van producten. De systeemgrenzen zijn daarbij van belang. Wie vergeet om het systeem te optimaliseren, maar zich concentreert op het optimaliseren van deelprocessen, komt makkelijk tot verkeerde conclusies. Zo leidt de taakstelling 'reducere de onderhoudskosten met 10%' licht tot machinestilstand en dus tot minder productie. De schade kost dan al gauw een veelvoud van de besparingen op het onderhoud.

Bedrijfszekerheid richt zich op het optimaliseren, gegeven allerlei randvoorwaarden, van de kans op de succesvolle werking van een systeem. Om dit doel te bereiken moet

het optreden van storingen worden tegengegaan. Kenmerkend voor storingen is dat ze op onvoorziene, en gewoonlijk ongelukkige, momenten optreden. De systeemspecificaties geven aan onder welke omstandigheden het systeem heel moet blijven, en onder welke omstandigheden het mag falen.

Voorbeeld 6

Op 6 oktober 1981 steeg een Fokker Friendship in Rotterdam op voor een vlucht naar Eindhoven. Het toestel kwam daar nooit aan. Het kwam in de trechter van een orkaan, de rechtervleugel brak af, en 8 minuten na vertrek lagen de resten van het toestel bij Moerdijk. Dat die vleugel afbrak was correct; het probleem was dat het toestel nooit in die orkaan terecht had mogen komen.

Het moment waarop producten falen wordt natuurlijk beïnvloed door variatie tussen componenten, variatie tijdens productie en variatie tussen gebruikers. Bij beschouwingen over productkwaliteit poogt men daarom variatie waar mogelijk uit te bannen. Toch is variatie niet alleen negatief. Een deel van onze genoegen halen we juist uit het feit dat er variatie is. Ieder heeft nu eenmaal zijn eigen smaak en zijn eigen financiële mogelijkheden. Soms is variatie zelfs essentieel. Stel u voor dat bij het maken van popcorn alle maïskorrels op hetzelfde moment zouden 'poppen'. Een explosie zou het gevolg zijn. Ook bij het reduceren van variatie dienen we dus eerst de rol van variatie in het systeem te begrijpen, daarna kunnen we eventueel besluiten om delen van die variatie te reduceren.

De vakgroep Quality and Reliability Engineering besteedt veel aandacht aan de bedrijfszekerheid van consumentenelektronica. Door de hoge innovatiegraad en de drang van fabrikanten om als eerste een nieuw product op de markt te hebben, is de bedrijfszekerheid het kind van de rekening. U kunt zich dat wel voorstellen. Als het een half jaar kost om een innovatief product te ontwikkelen en te fabriceren, dan is de bereidheid gering om ook nog eens een half jaar uit te trekken om het product grondig te testen.

Het gevolg is dat producten bij de gebruiker lang niet altijd vlekkeloos functioneren. Het ligt daarom voor de hand dat fabrikanten pogen om snel na de introductie vast te stellen welke problemen zich bij de gebruiker voordoen en hoe frequent. Helaas is dit geen eenvoudige procedure. Ik noem twee oorzaken:

1. Als een apparaat niet goed werkt, dan ligt dat tegenwoordig meestal niet aan het falen van een component. Meestal ligt het

aan het complexe samenspel tussen componenten, modulen en/of apparaten. De oorzaak is daardoor moeilijk op te sporen, daar moeten specialisten aan te pas komen met speciale apparatuur. Het apparaat wordt dan gewoonlijk meegenomen naar een speciale werkplaats. Hierdoor wordt het product echter uit het systeem gelicht, en daardoor verdwijnt het samenspel dat tot het niet-correct functioneren leidde.

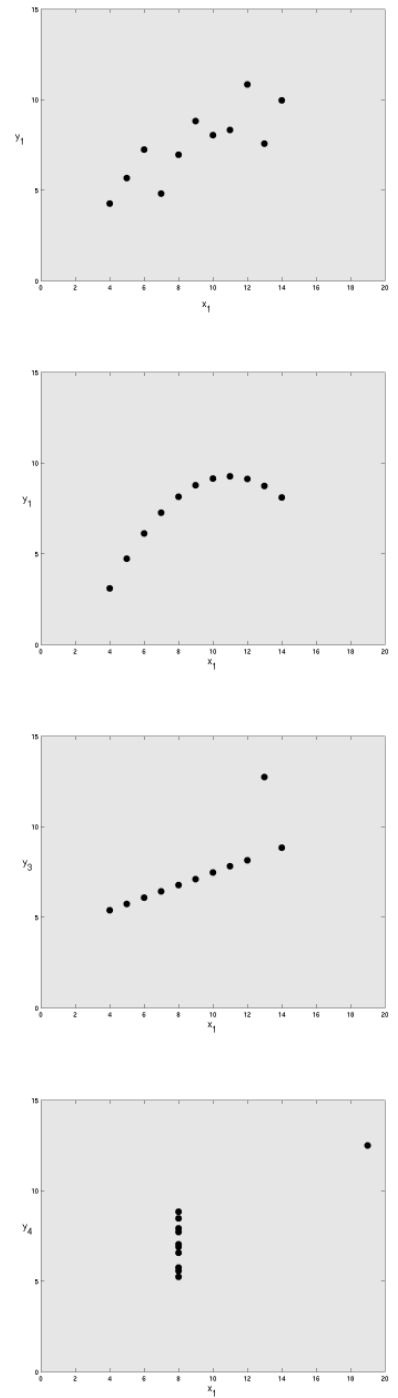
2. Door eindeloze logistieke processen is het heel normaal als de fabrikant na de start van de productie pas na een halfjaar de eerste terugkoppeling krijgt over veldproblemen. Het zou kunnen zijn dat dit komt omdat die problemen zo zeldzaam zijn, maar dat is helaas niet het geval. Na een halfjaar is het product vaak al niet meer te koop en daardoor ziet de ontwerper de veldinformatie als mosterd na de maaltijd. Voor productontwikkeling zou de Wet van Murphy als volgt kunnen worden geformuleerd:

De informatie die een wijziging in een ontwerp noodzakelijk maakt, wordt pas aan de ontwerper geopenbaard nadat de productie al is gestart.

De wet

Gelukkig beschermt de wet de consument tegen al te overmoedige acties van fabrikanten. Dit is niets nieuws, al zo'n 4000 jaar geleden werd van fabrikanten verwacht dat zij een kwalitatief fatsoenlijk product leverden. Het enige verschil met tegenwoordig is dat de straffen vroeger wat serieuzer waren. De oudste geschreven wettekst is zo'n kleine 4000 jaar oud en stamt uit Babylon. Daar leefden toen de Sumeriërs, een hoog ontwikkeld volk; zo kenden zij de stelling van Pythagoras al ruim 1000 jaar voor Pythagoras werd geboren. Ook het wiel is vermoedelijk door de Sumeriërs uitgevonden. Ik weet niet of zij al zover waren gevorderd met de bedrijfszekerheid, dat zij ook het reservewiel hebben uitgevonden.

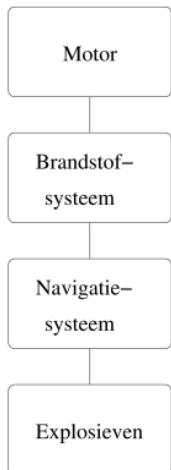
In 1763 voor Christus versloeg Hammoerabi de Sumeriërs in Rim-Sin en dat was het begin van het Babylonische Rijk. Hammoerabi liet de al eeuwen bestaande gebruiken vastleggen in wetten. Wie de originele versie daarvan wil lezen, moet naar het Louvre gaan; daar staat de bijna drie meter hoge stèle waar de 282 wetten van Hammoerabi zijn ingebeiteld. Inbeitelen van wetteksten is lang in zwang gebleven; zo'n 500 jaar na Hammoerabi beitelde Mozes de tien geboden in de Stenen Tafelen. Die tien geboden zijn wat de inter-



Figuur 2 Grafische weergave van de gegevens uit tabel 3

menselijke relaties betreft in wezen een korte samenvatting van de 282 wetten van Hammoerabi. Alleen voegde Mozes er niet-aardse wetten aan toe. Hammoerabi was zijn tijd ver vooruit, want hij hanteerde toen al een strikte scheiding van kerk en staat.

Het uitgangspunt van Hammoerabi is dat de sterke de zwakke niet mag onderdrukken. Dat wordt naar de huidige normen vaak wat zwaar aangezet, zoals in wet 229:



Blokdiagram van de V1

Als een bouwmeester voor een particulier een huis gebouwd heeft, en de constructie is niet sterk genoeg, waardoor het huis instort en de bewoner doodt, dan moet ook de bouwmeester worden gedood.

In gewone taal betekent dit niets anders dan dat de koper mag verwachten dat een deugdelijk product wordt geleverd. Wet 230 voerde de druk op de bouwmeester nog wat op:

Als de zoon van de bewoner wordt gedood, dan moet ook de zoon van de bouwmeester worden gedood.

Tegenwoordig vinden we dit oog-om-oog, tand-om-tand-principe niet humaan. Het doel om de misdadiger op te voeden tot een geëerd medeburger weegt nu zwaarder dan de behoefte om te straffen. Wet 232 spreekt ons meer aan:

Als goederen worden beschadigd, dan zal de bouwmeester alle schade vergoeden, bovendien zal hij het huis uit eigen middelen opnieuw opbouwen.

Daar het Tweestromenland de bakermat van onze beschaving is, spreekt het vanzelf dat onze wetten zijn gebaseerd op de wetten van Hammoerabi. Ik noem twee van onze wetten die aansluiten bij de gedachte dat de consument moet worden beschermd.

Ik begin met de wet op de productaansprakelijkheid volgens het Burgerlijk Wetboek Boek 6, Titel 3, Afdeling 3 [11]. De kern van deze wet is dat een fabrikant gedurende 10 jaar na verkoop van het product verantwoordelijk wordt gehouden voor schade die het gevolg is van een gebrek in zijn product. Het begrip gebrek is breed gedefinieerd. Volgens artikel 186 is een product gebrekkig indien het niet de veiligheid biedt die men daarvan mag verwachten, alle omstandigheden in aanmerking

genomen. Schade kan betekenen dood of lichamelijk letsel, maar schade kan ook gevolgschade zijn.

Hoe zinvol deze wet ook is, ik denk dat vaker een beroep zal worden gedaan op de conformiteitswet die ons beschermt als een product niet doet wat wij er in redelijkheid van mogen verwachten. Deze wet heeft lak aan garantiebepalingen, maar gaat uit van de gedachte dat een product aan redelijke verwachtingen behoort te voldoen. Letterlijk staat in artikel 17 sub 2 van het Burgerlijk Wetboek Boek 7, Titel 1, Afdeling 2 [11]:

De koper mag verwachten dat de zaak de eigenschappen bezit die voor een normaal gebruik daarvan nodig zijn en waarvan hij de aanwezigheid niet behoefde te betwijfelen, alsmede de eigenschappen die nodig zijn voor een bijzonder gebruik dat bij de overeenkomst is voorzien.

Ik heb sterk de indruk dat winkeliers hun uiterste best doen om deze wet uit de publiciteit te houden. Sterker nog, zij trachten u te verleiden om tegelijk met uw aankoop nog een aanvullende verzekering te kopen om de garantietermijn met één of twee jaar op te rekken. Onzin, want dan betaalt u voor dingen die in de conformiteitswet al zijn geregeld. Postbus 51 van het Ministerie van Justitie is daar helder over [7]. Ik citeer:

Vaak biedt een verkoper tegen een extra bedrag bovenop de koopsom een aanvullende garantie in de vorm van een serviceverzekering. Een koper doet er verstandig aan goed te overwegen wat hij hierdoor extra krijgt bovenop de wettelijke rechten die de consument al heeft. De serviceverzekering biedt mogelijk niet veel meer dan de koper al aan wettelijke rechten heeft.

Postbus 51 geeft het volgende voorbeeld:

De verkoper of de producent garandeert bijvoorbeeld dat een horloge zeker 40 jaar zal werken. Gaat het horloge eerder stuk, dan krijgt de koper een nieuwe. De koper kan, indien reparatie mogelijk is, ook verlangen dat het horloge wordt gerepareerd.

Het aardige van de wet is dat hij niet alleen goed is voor u, maar zeker ook voor mijn collega's in de vakgroep Quality and Reliability Engineering, want hierdoor is de relevantie van hun werkerrein in de wet verankerd.

Oorlogstuig

Ieder van mijn generatie herinnert zich dat op 4 oktober 1957 de Spoetnik werd gelanceerd, de eerste kunstmaan van de historie. De Spoetnik was een soort aluminium kogel met een diameter van 58 cm en een gewicht van 84 kg. De lancering maakte uiteraard bui-

tengewoon veel indruk in de wereld en was zeker voor de Amerikanen een schok. Tenslotte hadden zij praktisch alle raketgeleerden uit nazi-Duitsland naar de VS gehaald, met als kopstuk, natuurlijk, Wernher von Braun.

Al vanaf 1936 werkten Von Braun en andere raketspecialisten in Peenemünde, een legerbasis in de Baltische Zee, aan raketten [5]. Hitler zag aanvankelijk niet veel heil in die activiteiten. Dat veranderde pas in 1943, toen hij inzag dat de oorlog voor hem verkeerd dreigde af te lopen. Eén groep werkte aan de V1 en een andere, onder leiding van Von Braun, aan de V2. De V2 was een echte supersonische raket; hij vloog met zo'n kleine 5000 km per uur in 5 minuten van Den Haag naar Londen. Het belangrijkste doel van de V2's was trouwens niet Londen maar Antwerpen, omdat Antwerpen voor de geallieerden de belangrijkste aanvoerhaven voor materieel was.

De V1 was geen raket, maar een onbemand vliegtuigje. Letterlijk een vliegende bom, met een lengte van ongeveer 8 meter, met 900 kg springstof en een straalmotor. Hij vloog met zo'n 600 km/uur in ruim 20 minuten van Noord Frankrijk naar Londen. De V1 was veel goedkoper en sneller te produceren dan de V2. De namen V1 en V2 kwamen uit het brein van Goebbels, die niet voor niets minister van propaganda was. De V stond voor Vergeltung.

Ik concentreer mij op de bedrijfszekerheid van de V1 en begin daarom met een eenvoudige technische beschrijving. De V1 kon na de lancering niet meer worden bestuurd, dus de koers moest van tevoren zorgvuldig worden bepaald. Om de koers in het horizontale vlak te handhaven was in de neus een gyrokompass ingebouwd; een barometer moest zorgen dat de juiste hoogte werd aangehouden.

Als de V1 boven Londen was gearriveerd, dan moest de motor stoppen. Hiertoe was de neus voorzien van een kleine propeller. Als deze het vooraf berekende aantal omwentelingen had gemaakt, dan werden de hoogteroeien gedraaid, ging de neus naar beneden en stopte de benzinetoevoer vanzelf. Bij de inslag werd de stroomkring van het ontstekingsmechanisme gesloten en dit leidde tot de beoogde ontploffing.

Het werkte allemaal niet erg nauwkeurig, maar het doel was ook niet om strategische objecten te vernietigen, het doel was om dood en verderf te zaaien. Op 13 juni 1944 werden de eerste tien V1's afgevuurd. Het was geen groot succes. Vier crashten ter plekke, twee verdwenen spoorloos in het Kanaal, slechts vier haalden de overkant van het kanaal; drie daarvan kwamen in het open veld terecht en één kwam zo'n drie km van Tower Bridge,

zijn eigenlijke doel. Het resultaat was drie doden en een beschadigde treinbrug. Er werden ruim 9000 V1's op Engeland afgevuurd en ongeveer 2400 hiervan bereikten hun bestemming. De piloten van de RAF vonden het een sport om naast de V1 te vliegen, hun vleugel onder de vleugel van de V1 te brengen en het ding dan om te kieperen.

De hoofdreden voor de onbetrouwbaarheid van de V1 (en ook van de V2) was het ontbreken van elementaire kennis van kansrekening. Ik zal dat toelichten.

Als we de V1 terugbrengen tot zijn meest elementaire onderdelen, dan ziet hij er uit zoals in figuur 3 is weergegeven.

De V1 functioneert alleen naar behoren als elk van de vier delen correct functioneert. Dus én de motor, én het brandstofsysteem, én het navigatiesysteem, én de explosieven.

Von Braun dacht dat de kans dat een systeem werkt wordt bepaald door de zwakste schakel. Stel dat elke module met kans 0.8 correct functioneert. Volgens Von Braun zou dan 80% van de V1's in Londen ontploffen. De kansrekening leert echter dat onder deze omstandigheden slechts 41% in Londen zal ontploffen.

Robert Lüsser was de projectleider van de V1. Hij zag dat het toestel buitengewoon onbetrouwbaar was, maar hij had geen idee waardoor dat kwam. Ten einde raad haalde hij er een wiskundige bij, Perushka. Deze rekende hem voor hoe je de betrouwbaarheid van een seriesysteem bepaalt. Vanaf dat mo-

ment vloog de betrouwbaarheid van de V1 omhoog. Gedurende de laatste maanden van het gebruik van de V1 functioneerde zo'n 95% volgens het boekje. Het tragische voor Lüsser was dat toen de tegenmaatregelen van de Engelsen ook aardig op orde waren, waardoor nauwelijks meer dan 5% van de op Londen gerichte V1's door de diverse veiligheidslagen drong.

Ondanks de beperkingen van de V1 waren de geallieerden er uiteraard behoorlijk van onder de indruk. Een belangrijke strategische vraag luidde: konden de Duitsers hun vliegende bommen doeltreffend richten, of werden ze op goed geluk richting Londen geschoten? Om deze vraag te beantwoorden haalde men in Londen in navolging van Von Bortkiewicz de Poissonverdeling uit de kast [3]. Als de Duitsers de V1's niet goed konden richten, dan zou het aantal inslagen per hectare in Londen Poissonverdeeld moeten zijn. Met de doodgetrapte soldaten uit het begin van dit college in het achterhoofd, kost het u natuurlijk geen moeite om de geëigende aanpak te bepalen:

- u verdeelt Londen in stukken van gelijk oppervlak
- u telt hoeveel V1's in elk stuk neerkwamen
- u vergelijkt de gevonden frequenties met de verwachte frequenties volgens de Poissonverdeling.

Tijdens de oorlog verdeelde men een deel van Zuid-Londen in 576 stukken van 25 hectare en hield men bij hoeveel V1's in elk stuk neerkwamen. In het betreffende deel van Lon-

Aantal inslagen	Waargenomen frequentie	Frequentie volgens Poissonverdeling
0	229	226.7
1	211	211.4
2	93	98.5
3	35	30.6
4	7	7.1
≥ 5	1*	1.6
Totaal	576	576

Tabel 4 Frequentieverdeling van het aanslagen inslagen per stuk van 25 hectare in Zuid-Londen. * Het aantal inslagen was 7.

den werden 537 inslagen geteld. Tabel 4 geeft de verdeling van de V1's over die 576 stukken.

De Poissonverdeling met $\lambda = 537 : 576 = 0.9323$ inslagen per 25 hectare levert de verwachte frequenties in de rechter kolom. De overeenkomst tussen de waargenomen frequentie en de verwachte frequentie is dermate duidelijk dat men concludeerde dat de Nazi's niet konden richten. De tabel geeft dus een statistische onderbouwing voor de stelling dat de Duitsers niet uitwaren op het vernietigen van relevante doelen, maar dat dood en verderf zaaien de insteek was. Het nut van deze kennis was dat belangrijke veiligheidsinstallaties het beste konden worden beschermd door ze te verspreiden over het platteland; de kans was dan heel klein dat de Duitsers die installaties zouden raken. ◀

Literatuur

- 1 Anscombe, F.J. (1973): Graphs in statistical analysis. *American Statistician*, 27, 17–21.
- 2 Barnett, V. (1973): *Comparative Statistical Inference*. Wiley, London.
- 3 Clarke, R.D. (1946): An application of the Poisson distribution. *Journal of the Institute of Actuaries*, 22, 481.
- 4 Fisher, R.A. (1966): *Design of Experiments*. Oliver and Boyd, London.
- 5 Kokhuis, G.J.I. (1972): *Van V1 tot Ruimtevaart. Duwaer & Zonen*, Amsterdam.
- 6 Poisson, S-D. (1837): Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités. *Des Regles Generales Du Calcul Des Probabilites*, Paris, Bachelier.
- 7 Postbus 51 (2004): <http://www.justitie.nl/publicaties/brochures.en.factsheets/factsheets>
- 8 Savage, L.J. (1961): *The Subjective Basis of Statistical Practice*. Report, University of Michigan, Ann Arbor.
- 9 Von Bortkiewicz, L.J. (1898): *Das Gesetz der Kleinen Zahlen*. Teubner, Leipzig.
- 10 Von Bortkiewicz, L.J. (1915): Realismus und Formalismus in der mathematischen Statistik. *Allgemeines Statistisches Archiv*, 9, 225–256.
- 11 Wetten (2004): <http://wetten.overheid.nl>
- 12 Whitaker, L. (1914): On the Poisson law of small numbers. *Biometrika*, 10, 36–71.