

Nieuwe Kansen

Oratie van Michel Mandjes,
Hoogleraar Toegepaste Kansrekening
Korteweg-de Vries Instituut

uit te spreken op
vrijdag 2 februari 2007

Meneer de rector magnificus,
Meneer de decaan,
Geachte toehoorders,
Dames en heren,

Toen mij gevraagd werd na te denken over een uit te spreken oratie, bekwam me een gevoel van jongensachtig enthousiasme, maar daarnaast ook een zekere bezorgdheid, die ongetwijfeld verklaart waarom ik een en ander heb uitgesteld tot zes jaar na mijn benoeming. Een oratie dient een breed publiek aan te spreken, variërend van de naaste collega's tot familie en vrienden. Hierbij zijn de collega's vooral geïnteresseerd in de wetenschappelijke intenties van de nieuwe hoogleraar, de zogenaamde *leeropdracht*, en zijn ideeën ten aanzien van het academische klimaat aan deze universiteit, in dit land en in de wereld. De minder ingewijden geven de voorkeur aan een meer laagdrempelige inhoud, en uit hun mond heb ik in het verleden dan ook regelmatig vernomen dat de oratie volledig langs hen heen gegaan was. Zie hier het complexe probleem waarvoor de spreker zich gesteld ziet.

Ik moet ook bekennen dat, hoewel lid van de wetenschappelijke gemeenschap, mij ook menige oratie zwaar te moede werd. De hierboven beschreven spagaat leidt al gauw tot een knieval naar een van beide zijden. Bovendien leek het gemeengoed een wat voorspelbare structuur te hanteren waarin eerst een blik op het vakgebied wordt geworpen, dan de wetenschappelijke geloofsbrieven aan bod komen, vervolgens de visie op het onderzoeksbeleid en het onderwijs, om af te sluiten met enkele passages van meer persoonlijke aard.

Deze bespiegelingen in acht nemend leek het me een goed idee om voor mijn oratie eens een radicaal afwijkende vorm te kiezen. Mijn plan is U

mee te nemen naar het Korteweg-De Vries Instituut, afgekort KdV, en U een dag uit mijn leven te laten meeleven. Op die dag zullen we zien waar mijn vakgebied zich mee bezig houdt, wat mijn plannen zijn, wat mij drijft, en hoe ik denk over de ontwikkelingen binnen de academische wereld. Ik hoop dat na afloop zowel mijn collega's als alle minder ingewijden een goede indruk hebben gekregen van dit alles.

Vroeger begon elke werkdag met koffie, tegenwoordig met email. Email is de zegening en plaag van deze tijd. Het is je inmiddels moeilijk voor te stellen dat tot 15 jaar geleden alle communicatie met vakbroeders plaatsvond per post (of later fax), en één of twee keer per jaar tijdens een conferentie. Intensieve samenwerking was daardoor welhaast onmogelijk, tenzij men zich op dezelfde plek bevond. Die situatie is nu anders: ik werk samen met mensen uit alle windstreken, en er zijn geregeld dagen dat ik communiceer met prof. Peter Taylor in Australië, dr. Krzysztof Debicki in Polen, prof. Ilkka Norros in Finland, prof. Offer Kella in Israel, prof. Peter Glynn in California en dr. Alan Weiss in New Jersey.

Large deviations – lange-termijn afhankelijkheid. Misschien is dit een goed moment om iets te zeggen over hoe een dergelijke samenwerking er dan in de praktijk uit ziet. Laten we eens kijken naar bijvoorbeeld het werk dat ik met Ilkka Norros doe, en dan is het goed te beginnen met de inhoud ervan. Die heeft betrekking op een gebied van de kansrekening dat bekend staat als *large deviations*, of, in het Nederlands, de *theorie van grote afwijkingen*.

Als we terug gaan naar de basis van de kansrekening, dan weten we dat het steekproefgemiddelde van een rijtje onafhankelijke en gelijkverdeelde stochastische variabelen, onder milde voorwaarden convergeert naar de verwachting van die stochastische variabele. We schrijven dit als

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu \text{ bijna zeker,}$$

Dit een resultaat dat terug gaat naar de tijden van de 'grote Russen' Kolmogorov [14] en Khinchine [13], in de eerste helft van de vorige eeuw. Het is een heel krachtige stelling, want doorgaans wil men iets weten over het gemiddelde gedrag van een systeem. In heel veel toepassingen, is het echter niet zo zeer het gemiddelde dat de kwaliteit bepaalt, maar veeleer de zeldzame calamiteiten. Ik heb me in het verleden enkele jaren beziggehouden met het ontwerp van communicatienetwerken; dan

hadden die calamiteiten betrekking op de situatie dat er onvoldoende capaciteit aanwezig was om de klanten adequaat te helpen.

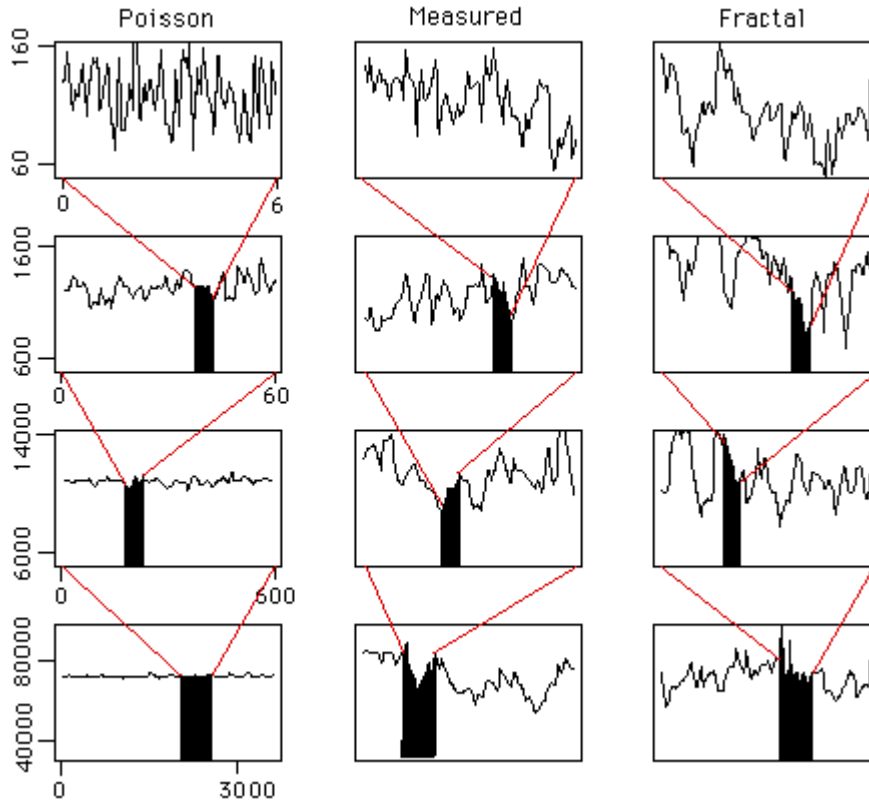
Met deze overweging in het achterhoofd, zou het interessant zijn de kans op een *zeldzame gebeurtenis* te analyseren:

$$\mathbb{P} \left(\frac{X_1 + \cdots + X_n}{n} \geq \mu + \epsilon \right).$$

De tak van de kansrekening [9, 10, 12, 17] die dergelijke kansen bestudeert heet dus *large deviations*, en ontwikkelde zich, na eerste resultaten in de jaren '30 van de vorige eeuw door de Zweedse onderzoeker Cramér [7], in hoog tempo in de jaren '50 en '60, interessant genoeg in parallel door een Russische school en een Westerse school. De relevantie met het oog op het ontwerp van communicatienetwerken werd pas later, vanaf de jaren '80, onderkend en ten volle uitgebuit [23].

Naast de ontwikkeling van *large deviations*, is er nog een tweede interessante trend zichtbaar. Die heeft betrekking op de modellering van verkeer in die telecommunicatie-netwerken. Waar aanvankelijk basale modellen met beperkte correlatiestructuur werden gebruikt (zoals het traditionele Poissonmodel), bewezen metingen telkens weer dat verkeer, bijvoorbeeld Internetverkeer, veel grilliger is. Eén van de bevindingen van enkele onderzoekers van het vermaarde AT&T Bell Labs was dat Internetverkeer een *fractale* structuur heeft [3, 24]. Eén en ander wordt geïllustreerd in het plaatje.

Hierin heeft de linkerkolom betrekking op Poissonverkeer: als men 'uitzoomt' wordt het verkeerspatroon steeds gladder. In het gemeten verkeer, als zichtbaar in de middelste kolom, kan men onmogelijk zeggen welke tijdschaal is afgebeeld; na een zekere herschaling van de verticale as zien alle plaatjes er eender uit. Er is een sterke gelijkenis met de rechterkolom, waar fractaal verkeer wordt getoond. Men noemt dit fenomeen *self-similarity*.



Figuur 1: Verkeersaanbod op meerdere tijdschalen, onder de Poisson-aanname (links), zoals gemeten (middel), en onder een fractaal model (rechts).

Norros heeft pionierswerk verricht op het gebied van self-similar modellen. Hij heeft gekeken naar een belangrijk speciaal geval: fractionele Brownse beweging [16]. Dit is een bijzonder inrtigerend stochastisch proces: het heeft continue paden die nergens differentieerbaar zijn (aldus recht doend aan het grillige karakter), en bovendien is het self-similar. Preciezer geformuleerd bedoelen we daar het volgende mee. Als we met $A(t)$ definiëren de hoeveelheid gegenereerd verkeer in een willekeurige periode van lengte t , dan geldt dat de variantie van de hoeveelheid verkeer tijdens een interval van lengte αt gelijk is aan α^{2H} keer de variantie van de hoeveelheid verkeer tijdens een periode van lengte t :

$$\text{Var}A(\alpha t) = \alpha^{2H} \text{Var}A(t), \quad \alpha, t \geq 0.$$

Hierin is H de zogenaamde Hurst-parameter, die weergeeft hoe sterk de correlatie binnen het verkeersaanbod is. Deze formule laat zien dat, als we

de tijd herschalen met een factor α , we hetzelfde stochastische proces zien, mits ook de verticale as op de juiste manier wordt aangepast.

Het werk dat ik met Norros, en anderen, heb gedaan heeft betrekking op wachtrijen gevoed door fractionele Brownse beweging [18, 19]. Daarnaast hebben we veel aandacht besteed aan een aanzienlijk algemenere klasse van inputprocessen, te weten de zogenaamde *Gaussische* processen. Voor die Gaussische processen wordt krachtig en elegant wiskundig gereedschap ontwikkeld, gebruik makend van mooie structureigenschappen, zoals de onderliggende *reproducing kernel Hilbert space*. De wachtrijen die je zo krijgt kunnen dan de bufferinhoud van een netwerkelement in het Internet modelleren, en met de resultaten kan geschat worden hoeveel vertraging het verkeer ondervindt in dat knooppunt.

Er waren al veel resultaten voor het geval dat een enkel knooppunt wordt beschouwd. Netwerken bestaan echter uiteraard niet uit geïsoleerde knooppunten, en daarom hebben we ons gericht op ingewikkeldere systemen, waarin de output van een eerste wachtrij wordt gerouteerd naar een volgende wachtrij. Deze systemen zijn aanmerkelijk lastiger, en we hebben ons van zwaar geschut moeten bedienen om de large deviations van de bufferinhoud te karakteriseren. In de analyse dient men optimaliseringsproblemen over een oneindig-dimensionale toestandsruimte op te lossen, waarbij men handig gebruik moet maken van de structuur van het probleem.

We hebben daarnaast ook veel aandacht besteed aan knooppunten die niet werken volgens het principe van first-in-first-out, maar waarin één klasse voorrang heeft boven een andere. Dat is typisch iets dat men zou wensen in een communicatienetwerk: geef vertragingssgevoelig verkeer (telefonie bijvoorbeeld) voorrang boven vertragingstolerant verkeer (email bijvoorbeeld). Met mijn promovendus Pascal Lieshout, die deels op het CWI werkt, en aan de UvA gaat promoveren, kijk ik naar dit soort problemen [15].

De antwoorden die men vindt zijn vaak in termen van wat wordt genoemd logaritmische asymptotiek. Bekijk bijvoorbeeld zo'n wachtrijstelsel als ik hierboven beschreef, met n verkeersstromen die van het knooppunt gebruik maken. Het verkeer in het knooppunt wordt met constante snelheid verder het netwerk in gepompt; zeg dat die snelheid nc is. Als Q nu de bufferinhoud aanduidt, op een willekeurig moment, dan vinden we een positief getal ζ zo dat de kans dat een bufferniveau nb wordt overschreden exponentieel daalt in n :

$$\mathbb{P}(Q > nb) \approx \exp(-n\zeta).$$

De samenwerking met Norros is een prototypische Internet-samenwerking. We hebben elkaar ooit, toen ik nog bij het onderzoekslab van KPN werkte, bij een Europees project ontmoet, en toen we er jaren later achterkwamen dat we inmiddels in soortgelijke problemen geïnteresseerd waren hebben we contact met elkaar opgenomen. We bezoeken elkaar elk jaar een weekje, zien elkaar soms bij workshops en congressen, maar de rest van de communicatie gaat via email.

Oplosmethoden uit de variatierekening - Markovse modellen. Een andere persoon met wie ik nog steeds, en voornamelijk via email, werk is Alan Weiss.

Weiss is een van de grondleggers van het gebruik van large deviations in de context van communicatienetwerken [23]. We zijn een tijdje collega's geweest, toen ik bij Bell Labs werkte. Zijn belangrijkste bijdrage, naar mijn smaak, is de karakterisering van de decay rate ζ als oplossing van een *variationeel probleem*, een concept waarvan Norros en ik ook veelvuldig gebruik hebben gemaakt.

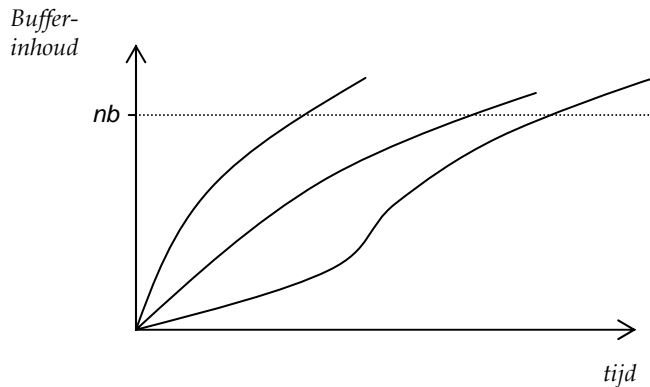
In meer concrete bewoordingen is het idee dat ζ uitgedrukt kan worden als het minimum van een 'kostenfunctie' over een padenruimte:

$$\zeta = \inf_{f \in \mathcal{S}} I(f),$$

dus het minimum van $I(f)$ over f in S . Hierbij bestaat de verzameling S uit alle paden (in de tijd) die starten met een lege wachtrij, en eindigen bij het overschrijden van bufferniveau nb . De functionaal $I(\cdot)$ geeft uitdrukking aan de kosten: een pad f heeft zekere 'entropie' (die geïnterpreteerd kan worden als 'kosten') $I(f)$. Hoe verder f afwijkt van het gemiddelde gedrag van het aankomstproces, des te hoger die entropie. Het gaat er dus om om een pad te vinden, dat start met een lege systeem en eindigt bij een volle wachtrij, en dat zo 'goedkoop' mogelijk is.

Het vinden van zo'n pad vereist de nodige handigheid met variatierekening. Het optimaliserende pad heeft de interessante interpretatie van *meest waarschijnlijke pad*: de gebeurtenis die we bekijken treedt zeer zelden op, maar *als* hij optreedt dan is het zeer waarschijnlijk dat dit gebeurt door middel van een pad dat in de buurt ligt van dat meest

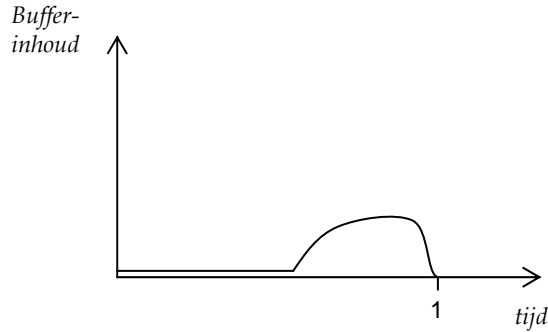
waarschijnlijke pad. Dit verklaart ook waarom men voor deze klasse van problemen vaak de term 'padsgewijze large deviations' bezigt. In de loop der tijd is er voor verschillende stochastische processen een onderliggende theorie ontwikkeld, en grote namen in dit verband zijn Mogulskii, Schilder, Donsker en Varadhan.



Figuur 2: Paden die leiden tot een volle buffer

Als gezegd is het vinden van zo'n pad f in S dat $I(f)$ minimaliseert doorgaans geen sinecure. Met Weiss ben ik er in geslaagd de klasse van Markovse modellen waarvoor zulks mogelijk is aanzienlijk op te rekken [20]. Uiteindelijk blijkt een en ander voor niet-Markovse modellen, zoals fractionele Brownse beweging, aanzienlijk lastiger te zijn.

Zoekt men bijvoorbeeld het meest waarschijnlijke pad dat er voor zorgt dat een wachtrij gevoed door fractionele Brownse beweging gedurende één eenheid tijd niet leeg is geweest, dan kan men na veel rekenwerk aantonen dat dit pad er uit ziet als op het plaatje. Met andere woorden: de buffer nagenoeg leeg is tot aan een tijdstip s^* , en daarna is er een strikt positieve bufferinhoud - deze uitkomst is uitermate ongewoon, en we konden het eerst nauwelijks geloven.



Figuur 3: Meest waarschijnlijk pad van busy period voor fractionele Brownse beweging

De decay rate (ofwel ‘norm’) die bij dit pad hoort heeft deze fascinerende gedaante:

$$\|\chi_{[0,s^*]}\|^2 + \frac{(1 - \chi_{[0,s^*]}(1))^2}{\text{Var}(Z_1 - \mathbb{E}[Z_1 | \mathcal{F}_s, s \in [0, s^*]])}$$

waarbij de eerste term betrekking heeft op de norm van het rechte pad χ tussen 0 en s^* , en de tweede term in rekening brengt wat de kosten zijn van het pad tussen s^* en 1.

De zojuist beschreven problematiek raakt aan het onderwerp van een OiO-project dat ik op dit moment begeleid, met promovendus Abdel Es-Saghouni. Hij bekijkt onder andere de *correlatiestructuur* van de bufferinhoud van de wachtrij, een tot dusver onopgelost probleem. Mijn verwachting is dat het verder ontwikkelen van de theorie van grote afwijkingen, en haar toepassingen in de wachtrijtheorie prominent op mijn (en ook de internationale) onderzoeksagenda zullen blijven staan.

Europa. Ik merk dat ik afdwaal. De zegeningen van email lijken me nu ruim genoeg bezongen, en het is nu tijd om aan te geven waarom ik email ook als plaag ervaar. De hoeveelheid tijd die ik besteed aan het behandelen van nagenoeg zinloze email is zorgwekkend. Er komt bijvoorbeeld zeer veel email voort uit Europese projecten, zogenaamde ‘Networks-of-Excellence’.

Deze ronkende naam ten spijt, heb ik deze projecten vaak als bureaucratische gedochten ervaren. Het wrange is dat Europese projecten zeker de toekomst hebben: samenwerkingen die al bestonden op een informeel niveau kunnen zo bestendig worden of verder uitgebreid. De

financieringsstructuur, waarbij 'Brussel' zich graag spiegelt aan de Amerikaanse NSF, is echter af en toe ronduit lachwekkend. De inspanningen die men zich moet getroosten om aan de administratieve verplichtingen te voldoen, staan vaak in schril contrast met de opbrengsten.

Mijn advies zou zijn om de beoordelingsprocedures volledig te herzien, en hierin de wetenschappers zelf een belangrijke rol te laten spelen. Zorg er verder voor dat daadwerkelijke excellentie wordt beloond, en hecht minder aan zuiver beleidsmatige criteria zoals bijvoorbeeld geografische spreiding. Alleen als er genoeg wetenschappelijke kwaliteit is, en er zorg gedragen wordt voor de transfer van de opgedane kennis naar de samenleving, kunnen Europese projecten tot een krachtig instrument uitgroeien.

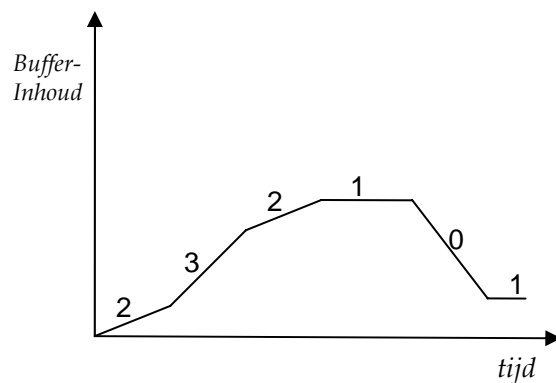
Modellen voor het delen van capaciteit - ad-hoc netwerken. Als de administratie weer enigszins op orde is, is het meestal tijd voor de wiskunde. In mijn geval betekent dat: kladblok en papier. Vaak heb ik ook een bijeenkomst met één van mijn promovendi. Naast de OiO's die ik al eerder noemde, is dat verder Frank Roijers, die hier over een paar jaar gaat promoveren. Frank werkt deeltijds op het Korteweg-de Vries Instituut, en deeltijds bij TNO Informatie en Communicatie Technologie.

Hij is het levende bewijs dat onderzoek zoals dat aan onze universiteit gedaan wordt, een directe link kan hebben met een operationele context. In zijn project richten we ons op zogenaamde *ad-hoc netwerken* [5]. Dat zijn netwerken waarin de fysieke infrastructuur ontbreekt: er zijn geen kabels of andere transmissieapparatuur. In plaats daarvan worden de knooppunten gevormd door de mobieltjes, of laptops, van de gebruikers van het netwerk. Als twee van die gebruikers met elkaar willen communiceren, dan zoeken zij een pad van mobieltjes waarover hun verkeer dan vervoerd kan worden. Een eerste interessante vraag is: gegeven het feit dat elk mobieltje maar een eindig bereik heeft, wat is dan de kans dat er zo'n pad is? Dat is een vraag met stochastische en combinatorische aspecten, en waarin ook percolatie-theorie een rol zou kunnen spelen.

Verder is het zo dat in zo'n netwerk vaak een knooppunt is aan te wijzen dat als bottleneck fungeert [4]: een positie in het netwerk zonder welke geen communicatie mogelijk zou zijn. Als we dat knooppunt in meer detail bekijken, zien we dat de bedieningscapaciteit beschikbaar voor het

knooppunt gedeeld moet worden tussen het *binnenhalen* van aankomend verkeer, en het *bedienen* van uitgaand verkeer.

Het standaardprotocol voor ad-hoc netwerken werkt als volgt. Stel dat er op zeker moment N zendende mobieltjes zijn. Die krijgen elk $1/(N + 1)$ van de capaciteit krijgen, samen dus $N/(N + 1)$. De resterende $1/(N + 1)$ wordt benut om het knooppunt te bedienen. De buffer groeit dan dus aan met een snelheid $(N - 1)/(N + 1)$. Met andere woorden: de hoeveelheid verkeer in het knooppunt vormt een wachtrij die alleen maar leeg loopt als er even géén zendende mobieltjes zijn, zoals zichtbaar in het plaatje.



Figuur 4: Evolutie van bufferinhoud voor een bottleneck in een ad-hoc netwerk

Het hier beschreven model is een zogenaamd *fluid model* [1, 6]. Het is echter een fluid model met een zeer ongebruikelijk koppeling tussen de input en output, en daarom niet standaard oplosbaar. Een doel zou kunnen zijn de verdeling te bepalen van de tijd die een boodschap erover doet door de bottleneck heen te komen; deze kennis kan dan weer gebruikt worden bij het verder verfijnen van de protocollen – heeft het bijvoorbeeld zin meer dan een fractie $1/(N + 1)$ van de capaciteit toe te kennen aan het bedienen van de wachtrij?

We hopen het fluid model te kunnen analyseren met de daartoe geëigende methoden, bijvoorbeeld het opstellen van een stelsel differentiaalvergelijkingen voor het tijdsafhankelijke gedrag, of wellicht via Wiener-Hopf-technieken. Fluid modellen hebben ook al een traditie van enkele decennia en vormen een uitermate boeiend gebied. Er zijn echter nog veel fundamentele wiskundige vragen die niet opgelost zijn, en die we samen met Frank hopen op te lossen. Ik begeleid Frank samen met Hans van den

Berg van TNO, met wie ik al meer dan 10 jaar een heel plezierige samenwerkingsrelatie heb.

CWI en EURANDOM. Veel van mijn werk in de komende jaren zal in samenwerking zijn met verschillende verwante groepen in binnen- en buitenland. Vanzelfsprekend zal ik nauwe contacten blijven onderhouden met mijn vorige werkplek, de groep PNA2 van het CWI.

Het CWI is voor mij de ideale plaats geweest om me wetenschappelijk te ontwikkelen. Ik hoop dat de plezierige samenwerking, met name met prof. Rob van der Mei en dr. Sindo Núñez Queija, nog lang niet ten einde is. Het is onmiskenbaar van het grootste belang dat er zo'n instituut bestaat als het CWI, en ik wil graag bijdragen aan de toekomst ervan.

Daarnaast zal ik als adviseur betrokken blijven bij het instituut EURANDOM in Eindhoven. EURANDOM is een instituut waarin jonge onderzoekers, postdocs, optimale mogelijkheden zich verder in de kansrekening en statistiek te verdiepen. Ik ben de directeur van EURANDOM, prof. Onno Boxma, zeer erkentelijk dat hij me de gelegenheid heeft gegeven deze positie te bekleden.

Ook hoop ik dat de banden met de Vrije Universiteit, mijn alma mater, even hecht blijven als ze nu zijn. Het is een goed moment stil te staan bij de rol die prof. Henk Tijms in mijn loopbaan heeft gespeeld. Ik heb ook aan de Universiteit Twente gewerkt – ik ben de UT zeer dankbaar voor het feit dat men het heeft aangedurfd me destijds te benoemen, wat ik als grote eer en erkenning beschouw. En het heeft me een heel plezierige samenwerking opgeleverd met dr. Werner Scheinhardt, die tot op de dag van vandaag voortduurt.

Zoals al blijkt uit het voorgaande, hebben de vakgebieden van de toegepaste kansrekening, stochastische besliskunde en wachtrijtheorie een lange traditie in Nederland. In deze zin ben ik schatplichtig aan een indrukwekkende reeks voorgangers. Zonder anderen te kort te willen doen, wil ik graag de naam van prof. Cohen noemen. Hij is vooral bekend om zijn nauwgezette analyse van een belangrijke klasse van wachtrijsystemen, zich met name beroepend op methoden uit de complexe-functietheorie [7]. Na zijn dood, in 2000, mocht ik zijn toga overnemen – ik ben er heel trots op dat ik deze mag dragen.

Ook aan deze universiteit heeft de stochastiek een lange en rijke traditie. Mijn dank is groot aan de mensen die zich ingezet hebben voor mijn komst, met name prof. Jan Wiegerinck en prof. Chris Klaassen. Ik zal mijn

uiterste best doen het vertrouwen dat zij in mij hebben gesteld niet te beschamen.

Aan de UvA zijn er verschillende disciplines waarvan ik wil onderzoeken of nauwere samenwerking aantrekkelijk is. Op voor mij bekend terrein zou het contact met besliskundigen prof. Nico van Dijk en prof. Jan van der Wal geïntensiveerd kunnen worden. Daarnaast liggen er interessante kansen voor de toegepaste kansrekening bij informatica, econometrie en actuariaat. Gezien de breedte van mijn leeropdracht zie ik het als taak deze mogelijkheden verder te verkennen.

Onderwijs. Het spreekt voor zich dat sinds mijn overgang van het CWI naar de UvA onderwijs een steeds belangrijker rol is gaan spelen in mijn bezigheden. Ik geef nu colleges voor eerstejaars en derdejaars. Dat eerste was nieuw voor mij, en heb ik als zeer verfrissend ervaren, en daarnaast ook nuttig, want lesgeven dwingt de docent nog eens goed na te denken over de fundamenteën van zijn vak. Ik heb nog steeds veel baat bij de sterke docenten die ik zelf heb gehad, en die mijn belangstelling voor wiskunde aanwakkerden, zoals dr. Rein Nobel en prof. Piet Holewijn.

Het is verleidelijk om ten aanzien van het onderwijs mee te gaan huilen met de wolven in het bos. Het niveau zou sinds de invoering van de Mammoetwet slechts een zorgwekkende neerwaartse trend vertonen, de eisen zouden steeds lager worden gesteld, en de moderne onderwijsvormen zouden louter beknellend werken. Ik acht het echter zeer kortzichtig hierom maar te opteren voor de onderwijsmethoden van een halve eeuw geleden [25]: de wereld is veranderd, en de student is veranderd, en het is noodzakelijk het onderwijs hierop aan te passen.

Het valt me echter steeds weer op dat in de *top* van de populatie – dus de bovenmodale wiskundestudenten – nauwelijks verandering optreedt. In elk jaar zitten wel een paar zeer getalenteerde studenten, die het later tot een promotie schoppen. Het *gemiddelde* niveau lijkt echter wel te dalen, om nog maar te zwijgen van het wiskundig niveau binnen studierichtingen die de wiskunde slechts als hulpmiddel gebruiken (economie bijvoorbeeld). Evidente oorzaak is de achteruitgang van het wiskunde-onderwijs op de middelbare school, maar ook het negatieve imago van exacte vakken onder scholieren.

Ik zie een verdere popularisering van wiskunde-onderwijs niet als zaligmakend. Een bepaalde groep leerlingen zal het inderdaad aanspreken als benadrukt wordt waar je die wiskunde zoal voor kunt gebruiken. Aan de andere kant valt het mij ook steeds weer op dat er daarnaast veel

studenten zijn die de schoonheid en het systematische van de wiskunde waarderen – dit aspect krijgt momenteel te weinig aandacht, denk ik. Betere aansluiting bij de toepassing, waar mogelijk, is dus prima, maar blijf aandacht houden voor de groep studenten die zich met name laat leiden door generieke aandacht voor abstracte vraagstukken.

Een mythe waarvan het hoog tijd wordt dat deze eens drastisch wordt ontzenuwd, is dat een wiskunde-opleiding een matige toegang tot de arbeidsmarkt zou geven. Ik zou graag het tegendeel willen beweren. Mijn waarneming is dat nagenoeg alle afgestudeerden binnen korte tijd werk vinden dat bij hen past, in het onderwijs, onderzoek of bedrijfsleven. Binnen de muren van KPN heb ik ook mogen ervaren dat een wiskundige achtergrond vaak wonderen doet bij het snel analyseren van problemen. Het verwerven van domeinkennis, dus achtergronden van toepassingsgebieden, hoort mijns inziens nauwelijks in de wiskunde-opleiding thuis. Die kan men zich namelijk zeer snel eigen maken wanneer men eenmaal bij een bedrijf werkt, terwijl de ervaring leert dat men, na het afstuderen, nauwelijks de gelegenheid krijgt nog enige wiskunde bij te leren.

Onderzoekssubsidies. Een substantieel deel van mijn tijd zal ik in de komende jaren blijven besteden aan het verwerven van onderzoekssubsidies, bijvoorbeeld om promovendi te kunnen aanstellen. Enkele decennia geleden liep die financiering met name via de universiteiten, maar tegenwoordig dient vrijwel al het geld geworven te worden via competities. Dit zijn de competities van NWO, maar daarnaast allerlei subsidieprogramma's die samenwerking met de 'toepassers' stimuleren, en natuurlijk ook projecten direct voor het bedrijfsleven of voor de overheid.

De competities van NWO zijn van groot belang, zeker voor een relatief fundamenteel instituut als het KdV. Het is daarom zorgelijk dat NWO zich in zwaar weer bevindt, en er slechts beperkte middelen zijn. De kansrekening en statistiek zijn bij een vorige grootschalige subsidieronde (de zogenaamde 'clustering') buiten de boot gevallen, maar ik zou er voor willen pleiten deze tak van wiskunde extra te ondersteunen – gezien de (hoge) kwaliteit van de Nederlandse onderzoekers op dit gebied, en gezien de relevantie van het veld. Hierin zou een centrale rol kunnen worden gespeeld door de stochastiekgroepen van de Amsterdamse instituten KdV en CWI, en mogelijk de VU, enerzijds, en EURANDOM anderzijds.

Over het werven van derde-geldstroom-onderzoek lopen de meningen sterk uiteen [11]. Sommigen redeneren dat financiële steun van het bedrijfsleven een bewijs is van de relevantie van het onderzoek. Er zijn echter typen onderzoek die wat verder van de praktijk afstaan, of minder directe economische waarde hebben, en die slagen er daardoor minder goed in subsidies te bemachtigen. Uiteraard zegt dit weinig over het wetenschappelijk kaliber. Daarnaast krijg ik soms de indruk dat met name wordt beloond in welke mate men het ambacht verstaat een goed, glad voorstel te schrijven.

Het is mijn doel onderzoeksprojecten te werven in het bedrijfsleven, maar daarnaast is het voor het KdV ook van groot belang dat ik bijdraag tot de totstandkoming van interessante stageplaatsen. Geld verdienen zal echter nooit leidend zijn: de inhoud moet in voldoende mate wetenschappelijk interessant zijn. Wellicht kan ik in dit verband een rol spelen binnen IBIS UvA, het florerende instituut voor industriële statistiek van het Korteweg-de Vries Instituut.

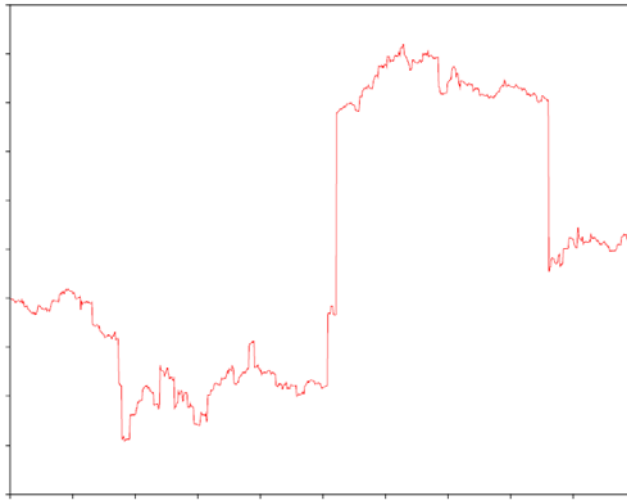
Een punt van zorg in dit alles betreft de budgetten die van overheidswege toegekend worden aan research en development. In Nederland betreffen die slechts 0.78% van het bruto binnenlands product, waar de norm, als vastgelegd in de Barcelona-convenanten, 1% is [21]. Het wordt voor mij daarom ook steeds onbegrijpelijker dat de politiek Nederland graag afschildert als een 'kenniseconomie'.

Een belangrijk gevaar dat ik observeer in mijn huidige contacten met bedrijven en in toegepast projectonderzoek, is de grote afstand tussen de voortreffelijke wetenschappelijk resultaten enerzijds, en hun toepassing (in concrete producten en diensten) anderzijds, de zogenaamde kennisparadox. Nieuwe subsidieprogramma's als Smartmix en Casimir, die die kloof moeten dichten, zijn hoopgevend, maar zie ik als slechts een eerste stap. Ik spreek mijn hoop uit dat de door NWO geclaimde uitbreiding van de tweede geldstroom [22] gerealiseerd zal worden.

Financiële toepassingen. Uit mijn onderzoeksactiviteiten blijkt mijn wens me te verbreden - mijn leerstoel heet 'Toegepaste kansrekening', en er zijn tal van domeinen waarin die toepassingen zouden kunnen liggen. Hiervoor sprak ik al over telecommunicatie, maar er zijn tal van andere mogelijkheden.

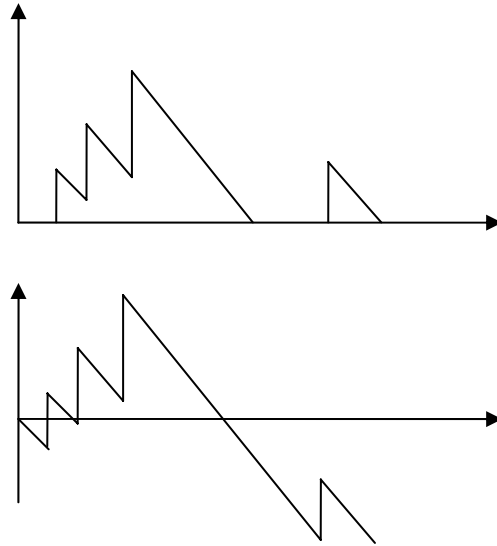
Via dr. Peter Spreij en zijn promovendus Vincent Leijdekker kom ik steeds meer in aanraking met het toepassen van stochastiek in de financiële wereld. Vincent werkt deeltijds bij ABN-AMRO, en deeltijds bij ons, en

brengt momenteel het stochastische gedrag van failliet gaande bedrijven in kaart. Eenvoudig gezegd: voor banken is het nuttig te analyseren wat het risico, de kans dus, is, dat meerdere van hun klanten binnen een vaste tijdshorizon failliet gaan. Met ander woorden: ook in de financiële sector hebben veel relevante vragen te maken hebben met zeldzame gebeurtenissen die eventueel geanalyseerd zouden kunnen worden met large deviations, of wellicht met andere probabilistische technieken.



Figuur 5: Een Lévy-proces met een continue component en sprongen.

Binnen de financiële toepassingen is ook een belangrijke rol weggelegd voor *Lévy-processen*. Dat zijn stochastische processen die zich kenmerken door onafhankelijke stationaire aangroeiingen. Die klasse van processen is, vanwege deze mooie eigenschappen, bijzonder goed te analyseren. Ze omvat belangrijke speciale gevallen. Die kunnen continu zijn, zoals de beroemde Brownse beweging (waarvan de relevantie in de financiële wereld al werd benadrukt door Bachelier in 1900 [2] – vijf jaar *voor* Einstein dus!), maar ze kunnen ook *discontinu* zijn, zoals het ‘compound Poisson-proces’. Het plaatje toont een voorbeeld van een algemeen Lévy-proces, met intervallen waarop het proces continu is, maar op bepaalde punten maakt het sprongen. Het is bekend dat dit soort processen goed gebruikt kan worden voor het modelleren van allerlei processen uit de financiële wereld.



Figuur 6: Gereflecteerde proces en bijbehorende vrije proces

Wanneer men bijvoorbeeld ruïneringsproblemen, die een centrale rol spelen voor verzekeringsmaatschappijen, wil onderzoeken, is men geïnteresseerd in de kansverdeling van de extreme waarden van zo'n Lévy-proces. Opmerkelijk genoeg zijn er directe connecties tussen die verdeling en de stationaire verdeling van een gerelateerd wachtrijproces (dat gereflecteerd wordt op 0), zoals getekend in deze figuur.

Dit betekent dat allerlei resultaten uit de zeer krachtige en zeer sterk ontwikkelde theorie van Lévy-processen in de wachtrijtheorie gebuikt kunnen worden, en vice versa. Het blijkt bijvoorbeeld dat het aloude resultaat van Pollaczek en Khinchine over de stationaire verdeling van de M/G/1-wachtrij uitgebreid kan worden naar gereflecteerde Lévy-processen.

Met Prof. Onno Boxma en Prof. Offer Kella heb ik de laatste jaren naar dergelijke relaties gekeken. Hier liggen nog veel open problemen, zeker als men kijkt naar modellen waarin het Lévy-process gemoduleerd wordt door een Markov-keten. De analyse van die laatste klasse van modellen vereist de toepassing van een breed scala van geavanceerde technieken, variërend van martingaaltheorie tot spectraalmethoden. Met Ton Dieker, die hier vorig jaar onder mijn begeleiding is gepromoveerd, ben ik gekomen tot de volgende generalisatie van Pollaczek-Khinchine, die mijn favoriete formule is van de laatste jaren:

$$\begin{aligned}
& (\psi_{-X}(\beta) - \alpha \mathbf{I}) \mathbf{E} \left[e^{-\alpha \bar{F}^X - \beta \bar{X}}; \bar{I} \right] \\
&= \begin{pmatrix} \mathbf{0}_{ss} & \mathbf{0}_{sz} & \mathbf{0}_{s-} \\ \mathbf{0}_{zs} & \mathbf{K}_{zz} - \alpha \mathbf{I}_{zz} & \mathbf{K}_{z-} \\ \mathbf{0}_{-s} & \mathbf{K}_{-z}^\alpha & \beta \mathbf{I}_{--} + \mathbf{K}_{--}^\alpha \end{pmatrix} \text{diag} \begin{pmatrix} \mathbf{0}_s \\ \mathbf{P}_z(\bar{X} = 0) \\ \mathbf{v}_- \end{pmatrix}
\end{aligned}$$

Bio-stochastiek. Langs andere weg ben ik recentelijk in aanraking gekomen met toepassingen van kansrekening in het kankeronderzoek. Eén van de onderzoeksrichtingen hierin betreft de vraag welke genen verantwoordelijk zijn voor het ontstaan van kanker. Veel onderzoek is gewijd aan de zogenaamde ‘insertie mutagenese’ waarbij de hypothese is dat kanker ontstaat door mutaties van genen. Om dit te onderzoeken laat men eerst laboranten, op kunstmatige wijze, mutaties veroorzaken bij proefdieren, bijvoorbeeld door virussen te injecteren. Hierdoor krijgt het proefdier een gezwel, en dat kan dan weer nader geanalyseerd worden.

De centrale vraag betreft nu de *positie* van het virus in het DNA: heeft het virus een ‘voorkeur’, in de zin dat specifieke genen zijn gemuteerd? Als onze hypothese juist is en kanker inderdaad ontstaat door mutaties van genen, verwacht je clusters van mutaties; bij een andere oorzaak zouden die volledig willekeurige posities innemen. Het menselijk genoom bestaat uit een zeer groot aantal van die posities, of *nucleotiden*; in de orde van drie miljard. Het aantal van die posities waarin het virus zich genesteld heeft, zogenaamde ‘inserties’, is veel kleiner: in de orde van 10000.

Op het scherm ziet u de verdeling van die inserties over het genoom. Het lijkt te suggereren dat er inderdaad van die clusters zijn, maar kan men dat ook statistisch onderbouwen? In stochastische termen is de vraag nu of die inserties, zouden kunnen stammen uit een homogene verdeling. Om dit vast te stellen moet, onder de hypothese van homogene spreiding, bepaald worden hoe waarschijnlijk het specifieke patroon is dat we observeren. In verband met ‘multiple hypothesis testing correction’ komen we, ook al zijn we geïnteresseerd in een betrouwbaarheidsdrempel van 5%, terecht in het domein van de zeer kleine kansen (in de orde van 10^{-5}). Het is gebleken dat ook hier large deviations toegepast kan worden.

Moraal. De moraal van het verhaal is de volgende. In verschillende toepassingsgebieden kunnen de cruciale vragen vaak gedefinieerd worden in generieke termen. Die kunnen geanalyseerd worden door bestaande technieken uit de kansrekening toe te passen, of door die

bestaande technieken verder te ontwikkelen, of door een geheel nieuw instrumentarium op te zetten. De analyse van zeldzame gebeurtenissen door middel van large deviations, die we aantreffen bij de analyse van telecommunicatienetwerken, bij financiële modellen en bij kankeronderzoek, is daar maar een voorbeeld van. Bij het bekleden van mijn ambt zal het steeds mijn doel zijn bij te dragen op methodologisch gebied, maar ook die methoden te gebruiken bij het analyseren van concrete problemen.

Slot. Na al deze beschouwingen naderen we het einde van mijn oratie, en is nu de tijd aangebroken voor wat meer persoonlijke woorden.

Het is duidelijk dat ik niet de persoon zou zijn die ik ben, zonder de hulp van veel mensen om mij heen. Ik wil dit moment gebruiken mijn ouders te vertellen hoezeer ik het waardeer dat ze me altijd vrij lieten in mijn keuzes. Het feit dat ik hier nu sta, is, denk ik, het directe gevolg van het feit dat ik mijn hart kon volgen, en de dingen kon gaan doen die me echt boeiden. Mijn vrouw Miranda en onze dochter Chloe hebben mijn leven veel kleur gegeven. Van hun aanwezigheid en steun geniet ik elke dag.

Ik heb gezegd.

Referenties

- [1] D. ANICK, D. MITRA en M. SONDHI (1982). 'Stochastic theory of a data-handling system with multiple sources'. *The Bell System Technical Journal* **61**, pp. 1871–1894.
- [2] L. BACHELIER (1900), Théorie de la Spéculation (Thèse), *Annales Scientifiques de l'École Normale Supérieure* (1900) **III -17**, pp. 21–86.
- [3] J. BERAN, R. SHERMAN, M. TAQQU en W. WILLINGER (1995). 'Long-range dependence in variable bit-rate video traffic.' *IEEE Transactions on Communications* **43**, pp. 1566–1579.
- [4] H. VAN DEN BERG, M. MANDJES en F. ROIJERS (2006). 'Performance modeling of a bottleneck node in an IEEE 802.11 ad-hoc network'. In: T. Kunz and S.S. Ravi (Red.): *Ad Hoc Now 2006*. 5th International Conference on Ad hoc Networks & Wireless "Ad Hoc Now", Ottawa, Canada. *Lecture Notes in Computer Science* (LNCS) Series **4104**, pp. 321–336.
- [5] T. BHEEMA REDDY, I. KARTHIGEYAN, B. S. MANOJ en C. SIVA RAM MURTHY (2006). 'Quality-of-Service provisioning in ad hoc wireless networks: a survey of issues and solutions'. *Journal of Ad Hoc Networks* **4**, pp. 83–124.
- [6] J.W. COHEN (1974). 'Superimposed renewal processes and storage with gradual input'. *Stochastic Processes and their Applications* **2**, pp. 31–58.
- [7] J.W. COHEN (1982). *The Single Server Queue*. Gereviseerde editie. North-Holland, Amsterdam, Nederland.
- [8] H. CRAMÉR (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* **736**, pp. 5–23.
- [9] A. DEMBO en O. ZEITOUNI (1993). *Large deviations techniques and applications*. Jones and Bartlett Publishers, Boston MA, Verenigde Staten.
- [10] R. S. ELLIS (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York, Verenigde Staten.
- [11] F. GEERDINK (Red.) (2006). *Onderzoek is altijd belangrijker!*, 2e editie. Verslag van de conferentie over bewegingen in de wetenschappelijke loopbaan. Stichting SOFOKLES.
- [12] F. DEN HOLLANDER (2000). *Large Deviations*. Fields Institute Monographs 14, American Mathematical Society, Providence RI, Verenigde Staten.
- [13] A.YA. KHINCHIN (1927). *Fundamental laws of probability theory*. Moskou (in het Russisch).
- [14] A.N. KOLMOGOROV (1930). 'Sur la loi forte des grands nombres'. *Comptes Rendus de l'Académie des Sciences* **191**, pp. 910–912.
- [15] P. LIESHOUT, M. MANDJES en S. BORST (2006). 'GPS scheduling: Selection of optimal weights and comparison with strict priorities'. *Proc. ACM Sigmetrics/Performance. Performance Evaluation Review* **34**, pp. 75–86.
- [16] B. MANDELBROT en J.W. VAN NESS (1968). 'Fractional Brownian motions, fractional noises and applications'. *SIAM Review* **10**, pp. 422–437.
- [17] M. MANDJES (2007). *Large deviations for Gaussian queues*. Wiley, Chichester, Verenigd Koninkrijk.

[18] M. MANDJES, P. MANNERSALO en I. NORROS (2007). 'Gaussian tandem queues with an application to dimensioning of switch fabrics'. *Computer Networks* **51**, pp. 781--797.

[19] M. MANDJES, P. MANNERSALO, I. NORROS en M. VAN UITERT (2006). 'Large deviations of infinite intersections of events in Gaussian processes'. *Stochastic Processes and their Applications* **116**, pp. 1269--1293.

[20] M. MANDJES en A. WEISS (2004). 'Sample path large deviations of a multiple time-scale queueing model'. Ingediend.

[21] NWO. *Wetenschap gewaardeerd! NWO-strategie 2007-2010*. Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

[22] M. PERSSON. 'Geef onderzoekers rust, rust alstublieft'. *de Volkskrant*, 27 mei 2006, p. Kennis-5.

[23] A. SHWARTZ en A. WEISS (1995). *Large Deviations for Performance Analysis*. Chapman & Hall, Londen, Verenigd Koninkrijk.

[24] M. TAQQU, W. WILLINGER en R. SHERMAN (1997). 'Proof for a fundamental result in self-similar traffic modeling'. *Computer Communications Review* **27**, pp. 5-23.

[25] A. VERBRUGGE en M. VERBRUGGE. 'Help, het onderwijs verzuipt! Leraren, leerlingen en ouders moeten samen de strijd aangaan voor beter onderwijs'. *NRC Handelsblad*, 3 juni 2006, pp. 15-16.

Kansrekening wordt vaak afgedaan als een wiskundige discipline waarin gekunstelde problemen, zonder duidelijke relevantie, geanalyseerd worden, veelal in de sfeer van kaartspelen of het werpen met dobbelstenen. Men gaat daarbij volledig voorbij aan de sterk toegenomen rol van de kansrekening binnen vele toepassingsdomeinen, waarbij gepoogd wordt onzekerheid te kwantificeren of, beter nog, te beteugelen. Aan de hand van enkele uitstapjes naar praktische problemen op het Internet, de beursvloer en het menselijk lichaam, zal aanschouwelijk gemaakt worden hoe fundamentele kansrekening gebruikt kan worden greep te krijgen op de onzekerheid van de wereld om ons heen.

Michel Mandjes is hoogleraar Toegepaste Kansrekening aan het Korteweg-de Vries Instituut van de Universiteit van Amsterdam, en daarnaast deeltijds verbonden aan het Centrum voor Wiskunde en Informatica en EURANDOM.