

Gemiste Kansen

Rede

uitgesproken bij het aanvaarden van het ambt van
bijzonder hoogleraar in de toepassingen van de wiskunde
aan de Rijksuniversiteit te Leiden
op vrijdag 20 februari 1987

door

DR. R.D. GILL



1987

RIJKSUNIVERSITEIT TE LEIDEN

Mijnheer de Rector Magnificus,

Mijne Heren Bestuurders van deze Universiteit, Mijnheer de Dekaan van de Faculteit van Wis- en Natuurkunde,

Dames en Heren Curatoren van deze leerstoel,

Dames en Heren Leden van het Algemeen Bestuur van het Leids Universiteits-Fonds,

Dames en Heren Hoogleraren, Docenten en Leden van de Wetenschappelijke en van de Technische en Administratieve Staf,

Dames en Heren Studenten,

en voorts allen die door Uw aanwezigheid blijk geeft van Uw belangstelling,

Zeer gewaardeerde toehoorders,

De meesten van U zullen er niet bij hebben stilgestaan dat de benoeming van een *mathematisch statisticus* op een leerstoel voor de *toepassingen van de wiskunde* iets paradoxaals in zich heeft. De statistiek is immers bij uitstek een toepassing van de wiskunde?

Toch kan men hier wel degelijk van een paradoxale situatie spreken, en ik hoop dit in het komende uur enigszins voor U te verduidelijken. Allereerst moet ik de niet-wiskundigen onder U erop wijzen dat, zoals er in de wiskunde in het algemeen een onderscheid te maken is tussen zuivere en toegepaste wiskunde, men in de statistiek onderscheid maakt tussen de mathematische statistiek, waarin men zich bezighoudt met algemene theorievorming, en de toegepaste statistiek, waarin men zich met het specifieke van bepaalde toepassingen bezighoudt en probeert haar in concrete situaties voor doeleinden buiten het vak

te gebruiken. Extreem gesteld: de mathematische statistiek wordt wel degelijk geïnspireerd door een bepaalde categorie toepassingen van de wiskunde, maar in de eerste plaats is zij een tak van de zuivere wiskunde, met haar eigen dynamiek en haar eigen esthetische criteria van algemeenheid, abstractie en diepgang. Wel kan het onderzoek van de mathematisch statistici gevolgen voor de praktijk hebben - de meesten onder hen zullen dat tenminste heimelijk geloven, hoewel helaas sommige toepassers het tegendeel verkondigen - maar dat is voor hen van ondergeschikte belang.

Vandaag zal ik de klassieke tegenstelling tussen de zuivere en de toegepaste wiskunde bekijken, waarbij ik uiteraard de nadruk zal leggen op de plaats van de statistiek en statistici daarin. Dat er een tegenstelling is, dat een paradox kan bestaan in het plaatsen van de mathematische statistiek onder de toepassingen van de wiskunde, wil ik nogmaals onderstrepen met het volgende citaat, afkomstig van de bekende zuiver wiskundige Paul Halmos¹: 'applied mathematics is bad mathematics'. Dat iemand uit die hoek dit zegt hoeft geen verbazing te wekken; maar wat vindt U van het feit dat deze uitspraak in een recente oratie² van een van mijn mathematisch statistische vakbroeders zonder duidelijk blijk van afkeuring werd geciteerd?

Het voorgaande is natuurlijk extreem gesteld. Men zou hieruit kunnen opmaken dat er twee goed te onderscheiden soorten statistici zijn, terwijl velen bij verschillende gelegenheden zich verschillend kunnen voordoen. Soms leidt dit tot een gespleten persoonlijkheid, soms juist tot een verrijkte, maar er blijft een spanningsveld bestaan tussen theorie en praktijk. Ik hoop U te tonen dat de toepassing van de statistiek in de werkelijke wereld, en de interactie hiervan met de theorie van de statistiek, en breder gezien, met de wiskunde als geheel, veel subtieler in elkaar zit dan het makkelijk te maken onderscheid tussen theorie en toepassing doet vermoeden. Mathematische en toegepaste statistiek liggen veel dichterbij elkaar dan men meestal erkent. Men zou zelfs van een innige verstrengeling kunnen spreken - misschien dankt het vak juist hier zijn speciale bekoring aan.

Laten we nu op enige echte toepassingen van de statistiek ingaan, om de rol

van de wiskunde erin te bekijken, en mijn boven geschetste stelling te verhelderen. Hiertoe zal ik U eerst iets meer moeten vertellen over wat de statistiek is en wat haar doelen zijn.

We worden dagelijks door de kranten en beeldbuis geconfronteerd met uitspraken waarbij statistiek te pas of te onpas is toegepast. Statistiek gaat over het trekken van conclusies uit waarnemingen of gegevens in situaties waarbij toeval of kans zo'n grote rol heeft gespeeld dat een eenduidige conclusie niet mogelijk is. Er zijn alleen onzekere conclusies te trekken, waarbij de rol van de statistiek juist is de mate van onzekerheid te karakteriseren en te minimaliseren. (Ik ga hier voorbij aan de belangrijke deelgebieden 'beschrijvende statistiek' en 'data-analyse' waarbij men het zonder kansuitspraken wil of moet doen.) Dat volgens de ene expert de ramp met de Challenger een kans had van een op duizend lanceringen terwijl hij volgens een andere wel een op twintig was, zou wellicht tot overeenstemming te brengen zijn door de ongetwijfeld aanwezige onzekerheidsmarges bij deze twee uitspraken te betrekken. Overigens wil ik beslist niet suggeren dat in een dergelijke complexe situatie een degelijke statistische conclusie te trekken zou zijn. Men moet niet alleen over alle gegevens beschikken die betrekking hebben op de kans op een ramp, maar ook over een model waardoor deze relatie expliciet en kwantificeerbaar wordt; wellicht lag het grote verschil tussen de twee uitspraken aan het feit dat verschillende modellen gebruikt werden.

Het belang van een model - een denkkader zo U wilt - wordt nogmaals geïllustreerd door de mededelingen in de kranten van het aantal te verwachten extra sterftegevallen als gevolg van het ongeluk in een kerncentrale in Tsjernobyl. Een dergelijk getal heeft geen enkele betekenis of is in het gunstigste geval op vele manieren interpreteerbaar. Iedereen sterft immers precies één keer, ramp of geen ramp. Bedoelt men eigenlijk: extra sterftegevallen binnen een bepaalde (relatief korte) tijd? Of gaat het om het totale aantal mensen dat eerder overlijdt dan het geval zou zijn geweest, als het ongeluk zich niet had voorgedaan? Maar is het in dat geval niet relevant, hoevéél eerder dit zou zijn; zou met andere woorden het 'aantal verloren

mensjaren' een zinvoller hoeveelheid kunnen zijn?

Ik hoop U straks duidelijk te maken dat de statistiek niet alleen te gebruiken is om uit een aantal gegevens wat conclusies te trekken, maar, wat veel belangrijker is, dat zij ons in staat stelt door middel van het gebruik van mathematische modellen erachter te komen of de beschikbare gegevens wel relevant zijn voor de beantwoording van de gestelde vragen. Dit is vaak niet het geval tenzij we er allerlei veronderstellingen bij maken. Een statistisch model maakt deze veronderstellingen expliciet en dus bespreekbaar, soms zelfs weerlegbaar.

Ik wil U nu ter illustratie drie voorbeelden geven; de eerste twee heel beknopt vergeleken met de laatste, waarin geavanceerdere wiskunde ter sprake zal komen. Voor het eerste voorbeeld moeten we een blik werpen op Leiden in de zeventiende eeuw. In die tijd werd niet alleen een revolutie in staatkundige zin geconsolideerd, maar ook in de wiskunde. In Nederland had Descartes net de analytische meetkunde ontwikkeld. Hiermee toonde hij aan dat algebra en meetkunde één zijn; men kan de meetkunde van Euclides opbouwen als gewone algebra van gewone getallen. Deze kwantificering van de ruimte bracht zowel nieuwe eenheid in de wiskunde alsook bood zij krachtig gereedschap om haar toe te passen. Zij was een essentieel onderdeel in het ontwikkelen van het Europese mechanistische wereldbeeld dat in de negentiende eeuw onder Laplace zijn hoogtepunt bereikte en waarvan wij de erfgenamen zijn. De nieuwe wiskunde van Descartes was ook slechts onderdeel van een nieuwe en controversiële filosofie en natuurkunde.

In Leiden werd op de klassieke manier de klassieke wiskunde, weinig veranderd sinds Euclides, beoefend. Prins Maurits bracht hier in 1600 verandering in door een leerstoel voor de toegepaste - in het bijzonder, militaire - wiskunde te stichten, die in 1615 toegekend werd aan Frans van Schooten, en na diens overlijden aan zijn zoon, die ook Frans heette, en die een fervente volgeling en populariseerder van Descartes was. Bij deze bijzondere leerstoel diende men onderwijs in het Nederlands te geven in plaats van het toen gangbare Latijn. Descartes verbleef zelfs omstreeks deze tijd enige jaren in Leiden.

De invloed van Van Schooten de Jonge blijkt duidelijk uit het werk van drie van zijn leerlingen: Johan de Witt, Johannes Hudde en Christiaan Huygens. Hoewel de bijdragen van Huygens aan de kansrekening bekender zijn - zijn boek *De Ratiociniis in Ludu Aleae* uit 1657 was de eerste en, mede dankzij de grote verspreiding ervan die Van Schooten bewerkstelligde, vijftig jaar lang de standaard tekst over kansrekening - wil ik hier iets vertellen over het werk van De Witt³. Hier vond een van de eerste toepassingen van de statistiek in de staatkunde plaats. Ook werd hierin kansrekening voor het eerst uit de sfeer van de kansspelen gehaald. (Het voornaamste wiskundige werk van De Witt was trouwens iets heel anders, namelijk een verhandeling over de kegelsneden à la analytische meetkunde.)

In 1671 stond De Witt als Raadpensionaris voor de zware taak snel fondsen bij elkaar te krijgen om de Republiek te verdedigen tegen de op handen zijnde aanval. In die tijden, toen een klein staatsapparaat weinig gelegenheid tot bezuinigen bood, was een gangbaar middel om aan geld te komen het verkopen van lijfrenten: voor een bepaald bedrag kon men een vast jaarinkomen of -rente voor een bepaalde persoon kopen; deze rente werd uitbetaald zolang de aangewezen persoon nog in leven was. De koper benoemde meestal een jong gezinslid. De gangbare koopsom was destijds veertien maal het jaarlijks uit te keren bedrag, ongeacht de leeftijd van de benoemde. De jaarlijkse inflatie bedroeg vier procent. Tegenwoordig kan men als men de kansverdeling kent van de resterende levensduur van kandidaten van een bepaalde leeftijd en uitgaat van een bekend inflatiepeil, zeer gemakkelijk de eerlijke koopsom berekenen. Destijds was dat niet eenvoudig; zelfs de taal ontbrak om in deze termen te denken. Niet alleen de taal maar ook relevante gegevens ontbraken: de eerste en lange tijd enige sterftetabel was pas negen jaar eerder door Graunt in Londen gepubliceerd (maar betrof tevens mortaliteit aan de pest, en dat bij een geschatte totale populatiegrootte) en betekende evenzeer als het werk van De Witt een doorbraak.

De Witt presenteerde zijn rapport over de waarde van lijfrenten aan de Staten Generaal. Dit rapport laat zich lezen als een mathematische verhandeling. Het

begint met de definitie van de verwachting van een toevalsvariabele, overgenomen van Huygens (die ook voor het eerst een mathematische definitie ervan gaf), en vervolgt met een uiteenzetting over hoe men deze moet berekenen. (De definitie verloopt via de prijs van een lot in een als equivalent te beschouwen eerlijke loterij.) Vervolgens construeert De Witt wegens het ontbreken van empirische gegevens de kansverdeling van levensduren door enige aannemelijke veronderstellingen te maken. Hij verdeelt een jong mensenleven in vier perioden. Binnen elk van deze perioden zijn de kansen om te overlijden gelijk verdeeld; verder is het risico in de tweede periode (van 53 tot 63 jaar) een factor anderhalf hoger dan het risico daarvoor (van drie tot 53), in de derde periode (63 tot 73) een factor twee hoger, en in de vierde (73 tot 80) een factor drie.

Na enig rekenwerk becijfert De Witt een forse koopsom van zestien keer de jaarrente bij een benoemde van drie jaar oud. Dit controversiële voorstel onderbouwt hij door vele redenen te noemen waarom de berekening zelfs aan de lage kant zou zijn, daarbij gesteund door de mogelijkheid het effect van variaties op de door hem gekozen factoren door te rekenen. Het is een zeer krachtig betoog dat geen middel onbenut laat (in het bijzonder een kleine statistische analyse van gegevens uit de registers van annuïteiten van Holland en West-Friesland) om zijn voorstel te ondersteunen.

Het voorstel werd geaccepteerd, hoewel De Witt's carrière bijna ten einde was: zoals bekend werd hij een jaar later vermoord. Zijn rapport werd in Europa beroemd en was bijvoorbeeld bekend aan Bernouilli en Leibniz.

Men kan zich afvragen of de leden van de Staten Generaal zich niet hebben laten verblinden door zoveel wiskunde; des te meer daar bij een nauwlettende lezing van het rapport enige vreemde inconsequenties aan het licht komen. De factoren anderhalf, twee en drie worden aan een argument over *voorwaardelijke risico's* ontleend: voor een individu van 58 zou het risico in het komende jaar te sterven anderhalf maal zo groot zijn als dat van iemand van 40. De berekeningen gaan echter over *onvoorwaardelijke* kansen: de kans dat een driejarige in zijn 58^e jaar sterft zou anderhalf keer de kans zijn dat hij in zijn 40^e

sterft. Bovendien blijken de factoren in de berekeningen plotseling twee derde, een half en een derde te zijn geworden; dat wil zeggen de reciproken van de originele! Deze 'schoonheidsfoutjes' zijn lange tijd niet opgevallen. In een correspondentie die hij hierna voerde met Hudde, burgemeester van Amsterdam, die nieuwe empirische gegevens analyseerde, geeft De Witt wel blijk het onderscheid tussen leeftijdsafhankelijk sterfterisico en onvoorwaardelijke kansverdeling van levensduren goed te kennen.

De Witt's berekeningen zijn efficiënt opgesteld en eenvoudig te herhalen met andere waarden van de parameters dan degene die hij koos; het is duidelijk dat deze mogelijkheid zijn modelkeuze beïnvloed heeft. Met de oorspronkelijke factoren in de onvoorwaardelijke kansverdeling komt men vlug tot de politiek onhaalbare prijs van 18 jaarrentes. Het ziet ernaar uit dat hij zijn cijfers heeft gekozen met het oog op het resultaat: na eerst een onacceptabele uitkomst te hebben gekregen heeft hij ze aangepast zonder, inderhaast (laten we veronderstellen dat niet alleen geld maar ook tijd schaars was), de hele redenering terug te volgen en aan te passen. Ook was het ontbreken van een heldere en beknopte wiskundige taal voor het object van zijn studie een handicap; een wiskundig probleem maakte de zaak bovendien nog gecompliceerder: het bepalen van de risicofunctie over de leeftijden heen tot op een constante van evenredigheid na zoals De Witt dat wilde doen, bepaalt niet de hele kansverdeling, terwijl dit wel het geval is bij het tot evenredigheid bepalen van de kansdichtheid.

Dit werk van De Witt heeft uiteindelijk toch weinig invloed gehad op de geschiedenis van de statistiek. De volgende decennia hebben de wiskundigen van Europa zich gestort op de ontwikkeling van de differentiaal- en integraalrekening. De Witt's oplossing voor het lijfrentenvraagstuk raakte in de vergetelheid en werd pas twee eeuwen later herontdekt, nadat eerst vele foutieve oplossingen waren gepropageerd. De kansrekening en de statistiek kwamen pas in onze eeuw tot bloei, toen ze konden wortelen in een veel rijpere wiskunde en gestimuleerd werden door toepassingen van groot wetenschappelijk prestige.

Uit een dergelijke voorgeschiedenis valt dan ook moeilijk lering te trekken

voor de moderne wetenschapsbeoefening. We mogen met waardering zien hoe De Witt met verve een net ontwikkelde abstracte theorie (de theorie van kansspelen) op een urgent politiek vraagstuk toepaste. Dat zijn analyse niet perfect was valt te verdedigen in het licht van een toen nog ontoereikende wiskundige taal en een onvolledig begrippenapparaat. Of hij zelf slachtoffer was van de gelegenheid tot verwarring die deze omstandigheden boden of dat hij deze omstandigheid juist voor zijn politieke doelen gebruikte, laat ik hier in het midden.

Laten we nu overgaan op een voorbeeld uit de moderne tijd, en wel de watersnoodramp van 1953. Naar aanleiding van deze ramp is de Delta Commissie gevormd. Aan het Mathematisch Centrum te Amsterdam werd de opdracht gegeven om door middel van een statistische analyse van de hoogwaterstanden van de laatste zeventig jaar te bepalen hoe hoog de zeedijken zouden moeten zijn - het basispeil - om de kans op een overstroming één op tienduizend per jaar (anders gezegd, een honderdste per honderd jaar, plotseling niet meer zo verwaarloosbaar) te laten zijn. Een statistische analyse was toen voornamelijk hoofd- en handenarbeid, zodat men uiterst zuinig en vindingrijk moest zijn met het opstellen en analyseren van verschillende modellen. Door middel van het in de analyse meenemen van één hoogwaterstand voor elk van de in de waarnemingsperiode voorkomende depressies van het kritieke type kon men de onderlinge afhankelijkheid in kort op elkaar volgende hoogwaterstanden uitschakelen. De keuze van deze depressies was echter een tijdrovende en enigszins subjectieve aangelegenheid. Na een zeer grondige analyse heeft men geconcludeerd dat de waarnemingen nagenoeg exponentieel verdeeld waren: dat wil zeggen, als het feit gegeven is dat de waterstand het peil van x meter overschreden heeft, dan is de kans dat er nog een centimeter bijkomt hetzelfde, wat ook de waarde van x mocht zijn. Deze ene kans en de kans op een depressie zijn beide uit de waarnemingen te schatten. Samen bepalen zij het gevraagde basispeil. Het resultaat van deze berekeningen was: een basispeil van 5,1 meter. Dit was veel hoger dan de door politici en ingenieurs bedachte 4,5 meter. Als men rekening hield met de mogelijke schattingsfout werd een

'veilig' basispeil zelfs nog hoger.

In een stormachtige vergadering in de Treveszaal op het Binnenhof hebben de wiskundigen - D. van Dantzig en J. Hemelrijk - de oorspronkelijke viereenhalf meter verhoogd weten te krijgen tot vijf, tot verbazing van vele commissieleden die zich afvroegen wat deze statistici zich wel verbeeldden.

Momenteel wordt het onderzoek herhaald⁴ met een inmiddels dertig jaar langere waarnemingsreeks, veel verfijnder analysemethoden, en ondersteund door moderne computerfaciliteiten. In het bijzonder is - deels door dergelijke toepassingen geïnspireerd - een uitgebreide en elegante theorie ontstaan van extremen in een reeks onderling afhankelijke toevalsgrootheden oftewel een stochastisch proces. In een voorlopig rapport wordt gewerkt met een ingewikkelder model met een extra parameter; het geschatte basispeil komt hierdoor veel lager uit - slechts 4,2 meter (de kansen op verhogingen dalen met het al bereikte niveau, zodat zeer extreme waterstanden onwaarschijnlijker worden). Door het opnemen van deze extra parameter is echter de mogelijke schattingsfout fors toegenomen (en een veilig basispeil onveranderd), zodat de eigenlijke conclusie misschien zal moeten zijn: de gevraagde extrapolatie is in feite niet te maken met de beschikbare gegevens.

Als derde en wat uitgebreider voorbeeld wil ik wat vertellen over een vakgebied in de moderne statistiek waar ik zelf bij betrokken ben geweest, de zogenaamde 'survival analysis'. Deze term - een eufemisme zoals U straks zult begrijpen - is de verzamelnaam voor statistische methoden die onder meer te gebruiken zijn bij de analyse van waargenomen tijdlangtes tussen het begin van een medische behandeling en het falen van die behandeling; men verzamelt zulke gegevens bijvoorbeeld ter vergelijking van een nieuwe behandeling en een oude bij klinische behandeling van bepaalde soorten kanker. Men moet zich hierbij een onderzoek voorstellen in een academisch ziekenhuis waarbij in de loop van een vijftal jaren misschien honderd tot tweehonderd nieuwe patiënten met een bepaalde ziekte worden opgenomen en behandeld. Voor elk van deze patiënten laat men het toeval de keuze tussen nieuwe en oude therapie bepalen om het effect van andere factoren uit te schakelen en een zo onberispelijk

mogelijke vergelijking te kunnen maken. Het gaat vaak om ziekten die nu niet te genezen zijn maar die wel een aantal jaren onder controle te houden zijn. Dit betekent dat de belangrijke te vergelijken patiëntenrespons niet een ja/nee variabele is (genezen/niet genezen), maar de tijdlangte tussen opname of behandeling en het einde van remissie, dat wil zeggen terugkeer van de ziekte. Deze tijdlangten zijn zeer variabel; het gaat er dus om de verdelingen van 'overlevingsduren' in de twee groepen te vergelijken.

In de jaren vijftig en zestig heeft het kankeronderzoek een enorme vlucht genomen. Onder Kennedy bijvoorbeeld is de 'fight against cancer' net zo grondig aangepakt als de wedstrijd om als eerste een man naar de maan te sturen. Niet alleen vele medische onderzoekers zijn er zich mee bezig gaan houden, maar in hun kielzog zijn er ook veel statistici bij betrokken. Deze confrontatie met een nieuw soort statistische problemen heeft geleid tot een nieuwe bloei van de statistische theorie.

Het vergelijken van de verdelingen van een grootheid in twee populaties, waarvan uit elk een steekproef is getrokken, is een standaardopgave voor een toegepast statisticus. Een ieder die wel eens een elementaire cursus statistiek heeft gevolgd, zij het voor natuurkundigen, psychologen of wie ook, zal met de parametrische 'Student's t-toets' en zijn niet-parametrische tegenhanger, de twee-steekproeven toets van Wilcoxon, geconfronteerd zijn. De kwalificaties 'parametrisch' en 'niet-parametrisch' duiden er hierbij op, dat de eerstgenoemde toets op zware veronderstellingen van normaalverdelingen in de twee te vergelijken populaties berust, dat wil zeggen verdelingen van een bepaalde vorm zodat verder twee numerieke parameters (gemiddelde en variantie) voldoende zijn om ze helemaal vast te leggen, terwijl de tweede onder minieme veronderstellingen een correcte procedure oplevert. In het soort medisch onderzoek waar ik het hier over heb, ligt het meer voor de hand niet-parametrische methoden te gebruiken. Men wil een zo overtuigend mogelijke conclusie kunnen trekken dat de nieuwe behandeling beter c.q. slechter is dan de oude.

Dusver lijkt het er niet op dat deze materie een nieuwe uitdaging voor de

statisticus biedt. Wat ik nog niet heb vermeld is het storende verschijnsel dat *censurering* heet. Men wil natuurlijk zo snel mogelijk kunnen beslissen over de relatieve merites van de twee behandelingen. Dit betekent dat op het tijdstip waarop men de gegevens moet analyseren, een behoorlijk aantal patiënten nog in remissie is; hoe beter de nieuwe behandeling, des te meer. Verder kunnen patiënten die in remissie zijn zich uit het onderzoek terugtrekken of onverhoopt aan geheel andere oorzaken overlijden. De waarneming van de overlevingsduur van al deze patiënten noemt men *gecensureerd*. Op een gegeven (wel waargenomen) tijdstip valt een sluier over hun verdere levensgeschiedenis.

Het zo maar weglaten van al deze gevallen in de analyse is op zijn best inefficiënt, en op zijn slechtst totaal misleidend. Het is van groot belang alle gegevens, gecensureerde en niet-gecensureerde, met erkenning van dat onderscheid, mee te nemen.

Aanvankelijk - ik denk hierbij aan de jaren vijftig - zijn voor dit probleem vele ad hoc aanpassingen van de eerder genoemde klassieke statistische analysemethoden bedacht. Het enige positieve van deze methoden was dat ze iets bruikbaar leverden. Een doorbraak kwam pas met essentieel nieuwe methoden die de dynamiek van de situatie volledig erkenden. Ik moet meteen beklemtonen dat het toegepaste statistici waren, in het bijzonder N. Mantel en W. Haenszel⁵, met bijzonder weinig wiskundige bagage maar met een sterke en gezonde intuïtie, die deze nieuwe methoden invoerden.

Het kernidee is het volgende. Het heeft geen zin om aantallen in remissie zijnde (of overlevende) patiënten in de twee groepen op een bepaalde tijdsduur na opname te vergelijken: deze aantallen zijn namelijk niet alleen door het beëindigen van de remissie tot stand gekomen, maar ook door censurering: het beëindigen van het onderzoek en andere onafhankelijke oorzaken. Maar stel nu dat we op een bepaald tijdstip binnen de twee groepen overlevenden kijken. Als de censurering werkelijk 'onafhankelijk' is, heeft elke patiënt in deze groepen dezelfde *voorwaardelijke* kans om in een hierop volgend klein tijdsinterval uit remissie te treden, gegeven dat men tot de groep overlevenden behoort, als het geval zou zijn zonder censurering. Men kan dus per tijdsinterval apart een

eerlijke vergelijking tussen de twee behandelingen maken, tenminste wat betreft de voorwaardelijke kans om uit remissie te treden gegeven dat men tot dan toe in remissie is geweest. Tenslotte moet men deze hele reeks deelvergelijkingen combineren, waarbij men telkens voorwaardelijk werkt op wat tot een bepaald tijdstip is gebeurd, en waarbij men de ontwikkelingen in het volgende infinitesimale tijdsinterval bekijkt.

Ik moet nu wel even ingaan op wat technische bijzonderheden om aan te geven dat de oplossing van Mantel en Haenszel ook een paar zeer ongebruikelijke elementen bevat. Bekijk een tijdsinterval waarbinnen een bepaalde patiënt uit remissie treedt dat zo kort is dat in dit tijdsinterval deze patiënt de enige is die of uit remissie treedt of gecensureerd is. Deel dan alle patiënten die bij het begin van dat interval beschikbaar zijn, in een twee-bij-twee tabel in met als categorieën: per rij, behandelingsgroep 1/2; per kolom, treedt in deze periode wel/niet uit remissie. Bereken de (niet gekwadrateerde) bijdrage van één van de vier cellen aan de bekende chi-kwadraat toetsingsgrootheid voor onafhankelijkheid (waargenomen minus verwachte aantal), tel deze bijdragen op over alle beschikbare tijdstippen, kwadrateer, deel door de som van de verwachte aantallen, en vergelijk met de chi-kwadraat verdeling met één vrijheidsgraad.

Op iemand die bekend is met de statistische theorie zal deze procedure, hoe gezond het kernidee ook is, heel wonderlijk overkomen. In elk van onze twee-bij-twee tabellen heeft één kolom een randtotaal van één - één uit remissie tredende patiënt - terwijl we de praktische vuistregel geleerd hebben dat de asymptotiek - de 'grote-steekproef' benadering - van de chi-kwadraat toetsingsgrootheid pas opgaat als er tenminste vijf individuen per cel onder de hypothese van onafhankelijkheid van behandelingsgroep en uit remissie treden voorkomen. Een verder bezwaar is dat al deze vergelijkingen niet onafhankelijk van elkaar zijn - immers, hoe meer patiënten nú uit remissie treden, hoe minder er overblijven voor de vergelijking op een later tijdstip.

Hoewel men dus niet tevreden was met de rechtvaardiging van deze methoden, heeft men ze toch voor lief genomen, vooral nadat uit heel andere

maar even bizarre redeneringen (Fisher score toets gebaseerd op de marginale dichtheid van de rangnummers van de waarnemingen in een denkbeeldige maar intuïtief gezien vergelijkbare experimentele opzet) dezelfde toetsingsgrootheid gelijk een wonder naar voren kwam⁶. Met steeds gedurfter heuristiek en intuïtie heeft men telkens verfijndere analysemethoden ontworpen om allerlei variaties op het hier beschreven basisprobleem aan te pakken, met als hoogtepunt het regressiemodel⁷ van D.R. Cox, geïntroduceerd in 1972. Een en ander voltrok zich voornamelijk in de medisch-statistische en toegepast-statistische vakliteratuur, in het algemeen genegeerd of met argwaan bekeken door mathematisch statistici, hoewel enkelen van hen deze nieuwe statistische methoden met behulp van klassieke mathematische technieken analyseerden. In een typerend geval⁸ ontdekte men na gigantische rekenpartijen waarin grote formules als door toverkunst tegen elkaar wegvielen, dat de uiteindelijke formules compact en aantrekkelijk waren én hetzelfde als degene die door de gedurfd heuristiek naar voren waren gekomen.

In deze informele wiskunde is de aandacht verschoven van het traditionele tellen over individuen naar het tellen over tijdstippen. De formele wiskunde - in dit geval de waarschijnlijkheidsrekening - die men nodig heeft om dit te preciseren en te begrijpen zal deze overgang ook moeten maken. Tegelijkertijd is voor de medische praktijk gebleken dat we makkelijker statistisch te werk kunnen gaan als we bereid zijn in eerste instantie het tijdsvariërende voorwaardelijke risico van uit remissie treden te onderzoeken, dan wanneer we ons willen richten op de onvoorwaardelijke kansen om op verschillende tijden uit remissie te gaan.

Pas in 1975 is voor het eerst de mathematisch-statistische theorie opgesteld die voor een bevredigende, elegante en volledige verantwoording van deze methodieken nodig was, en wel in een proefschrift⁹ uit Berkeley van de Noorse statisticus Odd Olai Aalen. De uitwerking hiervan heeft nog ruim tien jaar geduurd en het is mij een genoegen geweest daaraan deel te nemen. Aalen kwam uit Oslo met enige praktische ervaring in het analyseren van overlevingsduren: zijn doctoraal scriptie ging over een onderzoek naar de

overlevingsduur (in de baarmoeder) van het voorbehoedsmiddel het spiraaltje. Bovendien had hij een degelijke achtergrond in de theorie van stochastische processen ('one bloody thing after another', zoals de beroemde Engelse statisticus R.A. Fisher het eens aan een journalist uitlegde), vooral Markov processen, zoals gebruikt bij statistische toepassingen in de demografie. In Berkeley kwam hij in contact met een groep wiskundigen waaronder de Fransen Brémaud en Jacod die bezig waren een splinternieuwe theorie van stochastische integralen - een stochastische infinitesimaalrekening - toe te passen bij problemen van controle en filteren van telprocessen. Het was namelijk gebleken dat het idee van de tijdsafhankelijke, voorwaardelijke intensiteit van nieuwe gebeurtenissen in een toevallige reeks nauw verband hield met een kernbegrip uit deze theorie: de Doob-Meyer ontbinding van een net stochastisch proces in een systematisch ('voorspelbaar') gedeelte en een zogenaamde *martingaal*. Aalen herkende deze zelfde elementen in de 'survival analysis' en ontdekte voorts dat allerlei statistisch interessante grootheden op eenvoudige wijze beschreven konden worden in termen van deze theorie, namelijk als stochastische integralen van voorspelbare processen ten opzichte van martingalen.

Ik zal proberen deze vreemde termen enigszins te verduidelijken. De term martingaal is afkomstig uit Monte Carlo: het is namelijk dat gokstelsel, waarbij men ervan uitgaat dat als bij het spel roulette het balletje een tijd lang in minder dan de helft van de ronden op rood is gevallen, de kans dat dit wel gebeurt in volgende ronden groter zal zijn. In de kanstheorie echter is een martingaal (de abstractie van) de cumulatieve winst (als functie van de tijd) in een eerlijk gokspel, en het gemiddelde daarvan is nul, ongeacht gokstelsel of inzetregel. In ons voorbeeld geldt dat, als de twee behandelingen even goed zijn en als censurering op elk tijdstip onafhankelijk optreedt, de uitkomst op elk tijdstip van wie uit remissie treedt, zo iemand dat doet, net als in het roulettospel puur toeval is. De toetsingsgrootheid van Mantel en Haenszel is de eindwinst in dit spel bij gebruik van een bepaalde inzetregel. Als de ene behandeling beter is dan de andere is het spel niet eerlijk en treedt er (gemiddeld) een cumulatieve

winst (of verlies) op. Uit de martingaaltheorie is af te lezen hoe groot een winst moet zijn om niet aan toeval te wijten te zijn, dat wil zeggen, die een systematisch verschil aannemelijk maakt. De theorie is beslist niet triviaal, zeker in de 'continue-tijd' versie die wij hier nodig hebben.

Waar het op neer kwam was dat, bijna pasklaar, precies de goede mathematische theorie voorhanden kwam die de intuïtie van de eerste toepassers vertaalde in harde, exacte, wiskundige stellingen, en waar nodig deze intuïtie bijstelde: bepaalde aspecten ervan hadden hen op dwaalsporen gebracht. Deze mathematische theorie was net tot stand gekomen door de activiteiten van Paul-André Meyer en zijn school in Straatsburg. Lange tijd is zij gezien als totaal ontoegankelijk, daar zij met zoveel Franse abstractie en detaillering omgeven is. Zij was bovendien louter als zuivere wiskunde ontworpen - *l'art pour l'art* - en berustte op diepgaande en abstracte resultaten uit de potentiaaltheorie van onder meer Choquet. Slechts langzamerhand, en daarbij geholpen door de toepassingen, is men het essentiële van deze theorie gaan herkennen.

In de loop der jaren is er nog meer theorie bijgekomen die telkens direct voor de toepassingen gebruikt kon worden; hierbij denk ik in het bijzonder aan de martingaal centraal limiet stellingen van leden van de Russische school rond A.N. Shiryaev. Hiermee kan men voorwaarden aangeven, waaronder allerlei statistische grootheden bij benadering normaal verdeeld zijn. Het gebruik van deze stellingen behoort tot de meest onmisbare onderdelen van de dagelijkse statistische praktijk (in het bijzonder bij het gebruik van de methode van Mantel en Haenszel).

Achteraf gezien kost het geen moeite allerlei voorlopers van en aanwijzingen voor deze theorie in de toegepaste literatuur te vinden. Een interessant en zeer expliciet voorbeeld verschaft een artikel¹⁰ in een beroemd Engels biostatistisch tijdschrift waarin de auteurs zelfs op last van de redactie iedere uitwerking met behulp van martingaaltheorie *weglieten* omdat deze te moeilijk zou zijn voor de lezers, die immers meest uit toegepaste statistici bestonden. Dit is inderdaad een gemiste kans. Overigens levert ditzelfde artikel ook een voorbeeld van het

ontsporen van de gezonde praktische intuïtie - ik zeg dit om te beklemtonen dat de fraaie maar moeilijke wiskunde niet louter een spelletje is, maar juist hard nodig om helderheid, precisie, en de vaste grond voor de volgende vlucht van intuïtie te leveren.

De toepassing van deze theorie van martingalen, telprocessen en stochastische integralen heeft, via een aantal schitterende successen, geleid tot een uniforme behandeling van het hele scala van methoden uit de 'survival analysis', tot een duidelijke omlijning van hun toepasbaarheid, en tot een besnoeiing in de wildgroei van deels nuttige, deels vruchteloze concepten en theorieën. Men kan nu met uiterst gemak nieuwe methoden bestuderen in een bestaand theoretisch kader; bovendien kan men de ideeën uit de survival analysis overbrengen van hun eerste beperkte toepassingsgebied op allerlei andere situaties waarin de intensiteit van gebeurtenissen in de tijd bestudeerd wordt: ik zou voorbeelden kunnen noemen uit de demografie, de epidemiologie, de ethologie, de ontwikkelingspsychologie, ... Tevens is dat wat eerst een barokke en ontoegankelijke abstracte theorie was nu een uiterst krachtige en intuïtief goed te hanteren calculus geworden.

Aan dit laatste voorbeeld wil ik nog enige opmerkingen toevoegen. In de eerste plaats hebben al deze activiteiten andere gebieden van de statistiek bijzonder gestimuleerd. Zo leverde de praktijk een aantal voorbeelden van *semiparametrische modellen*: statistische modellen waarin oneindig dimensionale parameters voorkomen. Wat dat betekent kan ik in dit bestek niet uitleggen, maar het essentiële punt is dat precies op het juiste tijdstip, de 'survival analysis' een aantal specifieke en interessante voorbeelden bood en daarbij een sterke impuls gaf aan een nu tot stand komende synthese van parametrische en niet-parametrische methoden in het hart van de mathematische statistiek: er deden zich verschijnselen voor die vroegen om inpassing in een algemene theorie; experimentele situaties waren voorhanden waarvan elke zichzelf respecterende theorie de eigenschappen moest kunnen voorspellen en verklaren.

Zoals ik al memoreerde zijn er ook belangrijke gevolgen voor andere

toepassingsgebieden. In het bijzonder begint het mogelijk te worden de interactie tussen waarnemingskader en bestudeerd object (dat wil zeggen levensgeschiedenissen) in de demografie en epidemiologie op degelijke wijze te onderzoeken. Ik doel hierbij op het feit dat men in deze vakgebieden veelal noodgedwongen werkt met waarnemingsmateriaal dat retrospectief van aard is: achteraf verzameld, en afhankelijk van de toevallige ontwikkelingen die men bestudeert al dan niet in de steekproef opgenomen, met alle gevolgen van selectiviteit vandien.

Deze ontwikkelingen hebben ook gevolgen gehad voor de statistische traditie in bepaalde landen, met name Frankrijk en Rusland, waar sterke probabilistische scholen bestaan, maar die wat toegepaste statistiek betreft relatief zwak staan. Door deze nieuwe mogelijkheden hun 'eigen' zuivere wiskunde toe te passen is daar een hernieuwde belangstelling voor toegepaste statistiek ontstaan, en zijn wiskundigen van formaat er zich mee bezig gaan houden.

Het past natuurlijk ook iets te zeggen over de gevolgen voor het kankeronderzoek, waarover dit voorbeeld in eerste instantie ging. We mogen best erkennen dat de statistiek daar zo'n onmisbare rol speelt omdat men er juist niet in slaagde een werkelijke medische doorbraak te maken. (In feite is de statistiek in het algemeen onder dergelijke omstandigheden op zijn best.) Voorlopig moeten telkens met veel moeite kleine verbeteringen in 'gemiddelde' overlevingsduur worden opgespoord en bewezen. Op die manier werden verschillen in optimale behandeling voor verschillende patiënten aangetoond. Wel zijn er telkens andere factoren zoals 'levenskwaliteit' die deze kleine winsten soms weer teniet doen.

Ik wil graag een tweede relativerende opmerking maken. Laten we ons de vraag stellen: waaruit bestond eigenlijk de toepassing van de martingaaltheorie in dit geval? Wat werd er toegepast? Afgezien van het *modelleren* van bepaalde verschijnselen, en het *motiveren* van bepaalde methodieken van gegevensanalyse (allebei wel van bijzonder belang), leverde deze fraaie wiskunde voornamelijk benaderingen waarvan de nauwkeurigheid niet (althans,

nog niet) uit de theorie te bepalen is: bijvoorbeeld dat een toetsingsgrootheid *asymptotisch* normaal verdeeld is, en in een bepaalde zin *asymptotisch* optimaal. Ik wil niet suggeren dat dergelijke resultaten onbruikbaar zijn, integendeel; maar we moeten ons realiseren dat deze toepassing eigenlijk niet voltooid is. In alle situaties van enige gecompliceerdheid, zullen we altijd genoeg moeten nemen met benaderde oplossingen van de werkelijk gestelde problemen. De theorie moet worden aangevuld met empirisch onderzoek en praktische ervaring.

Dit voorbeeld illustreert het feit dat de wiskundige stellingen van de mathematische statistiek, fraai en diepgaand als ze kunnen zijn, vrijwel nooit in enge zin worden toegepast. (Ik ben van mening dat hetzelfde geldt voor elke tak van de wiskunde als het om meer dan triviale toepassingen gaat.) Wat de mathematische statistiek biedt is gereedschap om mee te denken, gereedschap om complexe problemen te analyseren en door benaderingen te reduceren tot goed te begrijpen oplosbare deelproblemen. Door de wisselende eisen van de praktijk moet een telkens wisselend compromis worden gevonden tussen analyseerbaarheid en realiteitsgehalte van de wiskundige modellering. Omdat er nooit één juiste oplossing is, heeft de wiskundige de vrijheid de mooiste, de meest verhelderende oplossing te vinden. De inferentiële statistiek waar ik het over heb gehad, komt tot zijn recht op de grens tussen determinisme en chaos, waar patronen nog net te onderscheiden zijn. Statistiek is pas echt interessant als er zoveel kans in het spel is, dat een zekere conclusie niet mogelijk is.

Geachte toehoorders, laat ik tot slot Dirichlet citeren¹¹, die in 1852 meende dat de tendens van de toen moderne analyse was: 'het vervangen van berekeningen door ideeën'. Dit moet mijns inziens ook nu nog de taak van de mathematische statistiek zijn. Het onderscheid tussen zuivere en toegepaste wiskunde waar ik mijn betoog mee begon, laat zich gevoelsmatig karakteriseren door schone ideeën enerzijds en wellicht nuttige maar saaie en eindeloze berekeningen anderzijds. Ik hoop U te hebben aangetoond dat deze tegenstelling misplaatst is. Het gaat er juist om in de toegepaste wiskunde de ideeën te vinden die de berekeningen vanzelfsprekend en helder maken en die de

grens van het kunnen (en kennen) zo ver mogelijk verleggen. De beste 'applied mathematics' is de mooiste wiskunde en tevens de moeilijkste wiskunde die er te produceren is.

Aan het einde van mijn rede gekomen, wil ik allereerst mijn dank betuigen aan Hare Majesteit de Koningin voor het besluit de Stichting Leids Universiteits-Fonds bevoegd te verklaren tot het instellen van deze leerstoel. Het algemeen bestuur van het Leids Universiteits-Fonds ben ik zeer erkentelijk dat het mij op deze leerstoel heeft willen benoemen, en in het bijzonder voor de eer in mij een waardig opvolger van de hooggeleerde Zoutendijk te zien. Voorts wil ik allen bedanken die zich voor mijn benoeming hebben ingezet.

Ook wil ik mijn dank uitspreken aan de leden van de Subfaculteit Wiskunde en Informatica, in het bijzonder de Vakgroep Toegepaste Wiskunde, voor het enthousiasme waarmee ik in hun midden verwelkomd ben en voor de plezierige en stimulerende samenwerking die ik al een jaar van hen ondervonden heb.

Graag wil ik nu een aantal mensen persoonlijk bedanken die een grote invloed hebben gehad op mijn wetenschappelijke vorming in dit land.

Hooggeleerde Hemelrijk, ik werd, toen ik pas in Cambridge was afgestudeerd, medewerker op Uw afdeling Mathematische Statistiek van het toenmalige Mathematisch Centrum, inmiddels Centrum voor Wiskunde en Informatica, te Amsterdam. Bij U stond centraal het op een onberispelijke wijze toepassen van de statistiek in en voor de maatschappij. Nog steeds tracht ik in de statistische consultatie Uw doortastendheid en instinct voor het vinden van het werkelijke probleem na te volgen.

Hooggeleerde Oosterhof, beste Cobus, als adviseur van de afdeling was het jouw lot mij in mijn promotieonderzoek te begeleiden, in een vakgebied waarvan we in het begin geen van beiden iets wisten. Jouw zorgvuldigheid en inzet zijn voor mij van enorme betekenis geweest. Van jou heb ik geleerd hoe men een promovendus hoort te begeleiden.

Hooggeleerde Groeneboom, beste Piet, jij was slechts een half jaar mijn senior op het MC maar van jou heb ik zeer veel geleerd. Jouw

doorzettingsvermogen om tot de wortel van elk probleem dat jou greep door te dringen, hoe ver dat je ook leidde in onbekende terreinen van de wiskunde, inspireert mij nog steeds.

Hooggeleerde Van Zwet, beste Willem, jouw scherpzinnigheid, eruditie en snelheid van geest zijn voor mij en wellicht voor de rest van je omgeving verbluffend. Het is een voorrecht en een plezier met jou hier in Leiden samen te mogen werken.

Geachte dames en heren curatoren, directie en medewerkers van het Centrum voor Wiskunde en Informatica, beste medewerkers van de afdeling Mathematische Statistiek, ik ben bijzonder blij dat ik deze leeropdracht mag combineren met mijn werkzaamheden op het CWI. Om werkelijk een Centrum te zijn is intensief contact met het universitaire leven van het land van centraal belang; en voor een niet in de eerste plaats educatieve instelling is het goed opleiden van onderzoekers elders een zaak van levensbelang. Het instituut heeft zijn vooraanstaande positie vooral te danken aan het feit dat er met een langetermijnvisie aan fundamentele problemen wordt gewerkt. Bij de statistiek, die op zo'n breed fundament stoelt, is diepgang alleen met lang en geconcentreerd onderzoek te bereiken, en na een werkelijk universitair onderwijs. In onze nogal pragmatische tijden, waarbij het oog veelal niet verder dan morgen kijkt, hoop ik dat deze langetermijnvisie toch ook bij de huidige overheid weerklink zal vinden.

Dames en heren studenten, U hebt gehoord dat de mathematische statistiek een moeilijk vak is dat voor de maatschappij, zij het weleens ter ontnuchtering, van het grootste belang is. In het vak moet men een verfijnde smaak en een grote kennis ontwikkelen, om het juiste pad te kiezen in het woud van mogelijkheden tot abstractie en dus analyse van een bepaald toegepast probleem. Om de statistiek toe te passen, maar ook om aan de theorie ervan bij te dragen, moet men een geschoold wiskundige zijn met inlevingsvermogen voor de doelstellingen van elk toepassingsgebied. Ik hoop dat U de kans zult grijpen en deze uitdaging met toewijding zult aanvaarden.

Ik heb gezegd.

Verwijzingen

1. P.R. Halmos (1981), Applied mathematics is bad mathematics, *Mathematics tomorrow* (L.A. Steen, ed.), 9-20, Springer, New York.
2. W. Albers (1985), *De ongrijpbare zekerheid*, Inaugurele rede, Rijksuniversiteit Limburg.
3. J. de Witt (1671), *Waerdye van Lyf-Renten naer Proportie van Los-Renten*, Den Haag; zie ook: A. Hald (in voorbereiding), *A history of probability, statistics and insurance mathematics before 1980*.
4. Samenwerkingsverband basispeilen (1986), *Onderzoek Basispeilen. Tussentijds verslag, juni 1986*, nota GWIO-86.008, Rijkswaterstaat, Dienst getijdewateren; zie ook: L. de Haan (1986), De ervvijand wiskundig bestreden, *Mathematics and Computer Science II* (M. Hazewinkel, J.K. Lenstra & L.G.L.T. Meertens, eds), CWI Monograph 4, 51-60, North-Holland.
5. N. Mantel & W. Haenszel (1959), Statistical aspects of the analysis of data from retrospective studies of disease, *J. Nat. Cancer Inst.* 22, 719-748; zie ook: N. Mantel (1966), Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Reports* 50, 163-170.
6. J. Crowley & D.R. Thomas (1975), *Large sample theory for the log rank test*, Tech. Report 415, Dept. of Statistics, Univ. of Wisconsin-Madison.
7. R. Peto & J. Peto (1972), Asymptotically efficient rank invariant test procedures (with discussion), *J. Roy. Statist. Soc. (A)* 135, 185-206.
8. D. R. Cox (1972), Regression models and life tables (with discussion), *J. Roy. Statist. Soc. (B)* 34, 187-200.
9. O. O. Aalen (1976), *Statistical theory for a family of counting processes* (reprint), Institute of Mathematical Statistics, University of Copenhagen.
10. R.E. Tarone & J. Ware (1977), On distribution free tests for the equality of survival distributions, *Biometrika* 64, 156-160.
11. G.P. Lejeune Dirichlet (1852), Gedächtnisrede auf C.G.J. Jacobi [bl. 19], *Abh. Preuss. Akad. Wiss.* 1-26; zie ook: E.E. Kummer (1860), Gedächtnisrede auf G.P. Lejeune Dirichlet [bl. 33], *Abh. Königl. Akad. Wiss. Berlin* 1-36; en: N. Bourbaki (1950), The architecture of mathematics [bl. 231], *Amer. Math. Monthly* 57, 221-231.