Mount Hermon Moscato, de beroemde Israëlische dessertwijn van de Golanhoogten, alsnog een pacificerende rol hebben kunnen spelen. Als later op de persconferentie premier Bennett verslag doet van de moeizame onderhandelingen en terloops ook vertelt hoe de twee tafelschikkingen van de lunch en het diner tot stand gekomen zijn vraagt onze sterverslaggever van de Ecotribune - het blad voor econometriestudenten aan de Vrije Universiteit - zich af wat de kans is dat tijdens het diner inderdaad niemand naast een lunch-disgenoot gezeten heeft.

Achter deze absurdistische kansopgave gaat een interessant en uitdagend kansprobleem schuil dat in één enkele zin geformuleerd kan worden:

'Stel dat n personen volgens een of ander protocol aan een ronde tafel geplaatst worden en dat vervolgens de tafelschikking veranderd wordt door de personen random een zitplaats toe te wijzen. Wat is de kans dat in de nieuwe tafelschikking geen twee personen naast elkaar zitten die in de oorspronkelijke tafelschikking wel naast elkaar zaten?'

Een combinatorisch kans probleem waarvoor het bijzonder lastig lijkt om tot een analytische oplossing te komen. In een dergelijke situatie brengt Monte Carlo simulatie uitkomst om snel een antwoord te vinden: de gezochte kans is ongeveer 11,4% voor n = 25. Voor n voldoende groot kan een simpele benaderingsoplossing gevonden worden met de Poisson heuristiek. Dit gaat als volgt. Denk in dat een Bernoulli-achtig deelexperiment wordt uitgevoerd voor elk van de n personen. Voor een gegeven persoon wordt dit experiment succesvol genoemd als deze persoon in de nieuwe random tafelschikking iemand rechts van hem/haar treft die in de oorspronkelijke tafelschikking rechts of links van hem/haar zat. De gezochte kans is gelijk aan de kans dat geen enkel deelexperiment succesvol is. De kans dat een gegeven deelexperiment succesvol is wordt gegeven door $2/(n-1)$. De verwachtingswaarde van het aantal succesvolle deelexperimenten is dus $2n/(n-1)$. De deelexperimenten zijn niet onafhankelijk van elkaar, maar voor n voldoende groot zijn de deelexperimenten als 'bijna-onafhankelijk' te beschouwen. Je kunt dan de Poisson heuristiek toepassen. Deze heuristiek geeft dat de kans op geen enkel succesvol deelexperiment bij goede benadering gelijk is aan $e^{-2n/(n-1)}$. Voor n = 25 vind je de benaderingswaarde 0,1245. De benaderingswaarde van 12,5% ligt opmerkelijk dicht bij de simulatiewaarde van 11,4%. Voor n voldoende groot geeft de heuristiek een uitstekende benadering (voor n = 50 is de benaderingswaarde 13,0% waar simulatie de waarde 12,5% geeft). De benaderingsformule is ook bruikbaar voor de situatie van een lange rechte tafel waar de n personen aan één kant van de tafel geplaatst worden. Voor zowel dit model als het model met een ronde tafel kan exact worden aangetoond dat voor n naar oneindig de formule $2^k e^{-2}/k!$ voor $k = 0,1,\ldots$ geldt voor de limietkans dat in de nieuwe random tafelschikking er precies k personen zijn met rechts van zich iemand die ook directe buur was in de oorspronkelijke tafelschikking. Een andere variant van het tafelschikkingsprobleem is een lange rechte tafel waaraan een even aantal personen aan beide kanten van de tafel geplaatst worden en de kans wordt gezocht dat in de nieuwe tafelschikking geen twee personen naast elkaar of direct tegenover elkaar zitten waar dat in de oorspronkelijke tafelschikking wel het geval was. Werk aan de winkel!

Henk Tijms is emeritus-hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Zijn meest recente boeken zijn *Basic probability; What every math student should know* (World Scientific Press, 2021, 2e druk) en *Operations Research; An introduction to models and methods* met de co-auteurs R. Boucherie en A. Braaksma, (World Scientific Press, 2021).
Homepage: https://personal.vu.nl/h.c.tijms/
E-mail: h.c.tijms@xs4all.nl

Oosterscheldekering. Foto: Eddy Westveer | www.beeldbank.zeeland.nl

Vijfenveertig jaar statistiek

Als je met emeritaat gaat dan ligt er meer tijd achter je dan voor je. De afgelopen 45 jaar heb ik als statisticus achtereenvolgens gewerkt bij het CWI, de Universiteit van Maastricht, het CQM (Philips) en de Universiteit van Amsterdam. In twee delen zal ik die 45 jaar beschrijven. Het eerste deel zal besteed worden aan mijn carrière als mathematisch en medisch statisticus.

Ronald J. M. M. Does

Het was een voorrecht om in de vorige eeuw na de lagere school naar de Hogere BurgerSchool (HBS) te gaan. Een brede vijfjarige vervolgopleiding met vijftien eindexamenvakken, die toegang gaf tot een universitaire opleiding. Ook de school van de oud-premiers Drees, Zijlstra, De Jong, Den Uyl en Kok; de school van de ondernemers Fokker, Philips, Heineken en Heijn; de school van dertien van onze eenentwintig Nobelprijswinnaars; en de school van kunstenaars en schrijvers zoals Vincent van Gogh, Jules Deelder, J. Bernlef, Joris Ivens, Annie M.G. Schmidt, Wim de Bie, Freek de Jonge, Herman Koch en Adriaan van Dis.

Na mijn HBS-b-opleiding begon ik in 1972 aan de studie wiskunde met natuur- en sterrenkunde aan de Universiteit Leiden. Na twee jaar slaagde ik voor mijn kandidaatsexamen en vervolgde ik met een doctoraalstudie wiskunde met bijvak mathematische statistiek. Het bijvak was een van mijn laatste tentamens en dat ging zo goed dat professor Van Zwet na afloop van het tentamen vroeg of ik interesse had om wetenschappelijk medewerker te worden op het Mathematisch Centrum (nu CWI geheten) in Amsterdam.

Kansmodel voor golf- en vervalkrachten

In september 1976 begon ik mijn carrière als 21-jarige op de afdeling Mathematische Statistiek, die onder leiding stond van professor Hemelrijk. Mijn leven als statisticus was begonnen en mijn eerste taak was om samen met Jelke Bethlehem en Richard Gill onder leiding van professor Oosterhoff een werkweek voor te bereiden met als thema stochastische censurering. Een onderwerp waarin de levensduurverdeling van een object centraal staat. De corresponderende overlevingsfunctie wordt gedefinieerd als $S(x) = P(X>x) = 1- F(x)$, waarin F de verdelingsfunctie is van de stochastische variabele X. In woorden: S(x) is de kans dat een object een levensduur groter dan x heeft. Een veelvoorkomend probleem is dat men niet van ieder object uit een aselecte steekproef op tijdstip x de levensduur kan meten omdat het object een langere levensduur heeft of bijvoorbeeld uitgevallen is door een andere oorzaak dan die men aan het onderzoeken is. In dergelijke gevallen spreekt men van censurering. In augustus 1977 werd de werkweek gegeven. Centraal stonden de verdelingsvrije methoden voor het twee-steekproevenprobleem met gecensureerde waarnemingen. Tevens werd tijdens de werkweek een overzicht gegeven van de martingaaltheorie en theorie van de stochastische integralen (zie Bethlehem e.a., 1977). Twee jaar later promoveerde Richard Gill op dit onderwerp met een proefschrift getiteld '*Censoring and Stochastic Integrals*'.

De afdeling Mathematische Statistiek deed in die tijd ook veel consultatie. Met name Rijkswaterstaat, onderzoeksinstituten en accountancykantoren waren grote opdrachtgevers. De aard van de consultaties was uitdrukkelijk mathematisch statistisch. In mijn tweede jaar kreeg ik samen met Jelke Bethlehem en Roelof Helmers de opdracht om de nota getiteld *Beschrijving van de probabilistische methode voor de bepaling van de golf- en vervalkrachten* van de Deltadienst op zijn wiskundig-statistische merites te voorzien. In Bethlehem e.a. (1978) wordt een kansmodel ontwikkeld voor de golf- en vervalkrachten op de stormvloedkering in de Oosterschelde. Omdat de golf- en vervalkrachten kunnen worden samengesteld tot een totaalbelasting is een uitdrukking gevonden voor de kans dat de totaalbelasting een gegeven belastingwaarde tenminste eenmaal overschrijdt. Voor het ontwikkelde kansmodel hadden we over de jaren 1949–1977 de beschikking over hoog- en laagwaterstanden te Vlissingen en Burghsluis; significante golfhoogten, tijdens stormen met het oog geschat op het lichtschip of lichteiland Goeree en gemeten bij een meetpaal met een amplitudeschrijver; en windrichtingen gemeten in Hoek van Holland. De bouw van de stormvloedkering in de Oosterschelde begon in 1979 en werd in 1986 voltooid. Sinds de ingebruikname worden de schuiven ongeveer eenmaal per jaar gesloten vanwege hoog water. Het is goed om te constateren dat de stormvloedkering de totaalbelasting iedere keer goed heeft doorstaan.

Naast het uitvoeren van wat kleinere consultatie opdrachten in het tweede jaar kreeg ik ook voldoende ruimte om bij te scholen. Zo werd in het academisch jaar 1978–1979 een werkgroep gestart, bestaande uit statistici van het Mathematisch Centrum en een aantal promovendi in de statistiek uit Nederland en België, om alle opgaven van het boek *Testing Statistical Hypotheses* van E. L. Lehmann op te lossen. Toen we klaar waren met het oplossen van de 192 opgaven, kregen we bericht dat er een tweede editie met extra opgaven zou verschijnen. Die verscheen echter pas in 1986 en toen was de werkgroep al opgeheven.

Hogere orde asymptotische ontwikkelingen voor een klasse rangtoetsen

Een aanstelling bij het Mathematisch Centrum was in principe voor vier jaar met een mogelijkheid tot verlenging van twee jaar. In die periode werd je dan wel geacht te promoveren. In mijn derde jaar werd het dus tijd om een promotieonderwerp te kiezen. Het eerste onderwerp ging over hogere orde asymptotische ontwikkelingen voor functies van uniforme spacings. Uniforme spacings worden gedefinieerd als de verschillen van opeenvolgende order statistics uit een steekproef van n onderling onafhankelijke (0, 1) stochastische variabelen. In wiskundige notatie: noteer met $U_1, U_2, \ldots, U_n$ de steekproef uit een (0, 1)-verdeling en $U_{1:n}, U_{2:n}, \ldots, U_{n:n}$ de order statistics met $U_{0:n} = 0$ en $U_{n+1:n} = 1$. Uniforme spacings zijn dan gedefinieerd als $D_{in} = U_{i:n} – U_{i-1:n}$ voor $i = 1, 2, \ldots, n+1$. Samen met Chris Klaassen en Roelof Helmers werden sommen van functies van deze uniforme spacings bestudeerd. Het onderzoek resulteerde in 5 artikelen (zie ondermeer

Op 3 januari 2018 raasde een storm over Zeeland. De kering werd gesloten.
Foto: Ben Biondina | www.beeldbank.zeeland.nl

Does & Klaassen, 1984 en Does e.a., 1987). Dit onderwerp werd niet mijn promotieonderwerp. Met Willem van Zwet als promotor werden het toch hogere orde asymptotische ontwikkelingen voor een klasse rangtoetsen. Voor het een-steekproefprobleem was dat al gedaan door Wim Albers en het twee-steekproevenprobleem was door Van Zwet en Peter Bickel opgelost. Bleef over de grotere klasse van lineaire rangtoetsen met het twee-steekproevenprobleem en Spearmans rangcorrelatie coëfficiënt als bijzondere gevallen. Het werd een technisch proefschrift met lange en soms ook ingewikkelde formules om Berry-Esseen-stellingen te bewijzen en Edgeworth-ontwikkelingen af te leiden. De complexiteit werd vergroot door onbegrensde score voortbrengende functies toe te laten, zodat de normaalscores, die bijvoorbeeld bij de Van der Waerden-toets worden gebruikt, binnen de resultaten vielen. Het onderzoek leidde tot zes artikelen, waaronder in de *Annals of Probability* en de *Annals of Statistics* (zie Does, 1982 en 1983). In 1982 verdedigde ik mijn proefschrift getiteld *Higher Order Asymptotics for Simple Linear Rank Statistics* aan de Universiteit Leiden.

Medische informatica en statistiek

Nu ik een ruggengraat had als mathematisch statisticus, vond ik het tijd om een carrièreswitch te maken naar de toepassingen. Begin tachtiger jaren lagen de banen echter niet voor het oprapen. Gelukkig was er in het zuiden van het land een nieuwe universiteit opgericht en op 1 mei 1981 ben ik daar begonnen als wetenschappelijk medewerker eerste klasse. De Universiteit Maastricht had een matrixstructuur met capaciteitsgroepen die het onderwijs en onderzoek verzorgden. Ik werd de eerste medewerker van de capaciteitsgroep Medische Informatica en Statistiek. Het onderwijssysteem in Maastricht was probleemgestuurd en het betekende dat je als tutor optrad voor kleine groepen studenten (maximaal 12) bij bijvoorbeeld het blok Wetenschapstheorie en Methodologie. Voor het statistiekonderwijs was er een uitzondering gemaakt. Studenten mochten hoorcolleges volgen en vervolgens praktijklessen volgen in kleine groepen. In mijn periode op de Universiteit van Maastricht steeg het aantal studenten Gezondheidswetenschappen heel snel (van 80 in 1981 naar bijna 1000 in 1989). Ook het aantal collega's groeide mee van 1 naar 40, maar de onderwijslast was ongekend hoog.

Mijn eerste onderzoeksonderwerp in Maastricht kwam van een immunoloog die geïnteresseerd was in de bepaling van de relatieve frequentie van een bepaald type cel in een populatie van cellen. Ook nu nog een heel actueel onderwerp bij de bestrijding van het COVID-19-virus, waar de zogenaamde T-cellen van belang zijn voor de immuunrespons. In een laboratoriumopstelling kan worden vastgesteld of het type cel al dan niet aanwezig is. Door slim een verdunningsreeks te ontwerpen waarbij in het begin een hoge score op aanwezigheid wordt gehaald en naar mate de verdunningen toenemen lagere scores op aanwezigheid worden gehaald, kan deze relatieve frequentie van de type cel bepaald worden. Zowel de proefopzet als de statistische analysemethoden werden door Wim Albers en onze eerste promovendus (Leo Strijbosch) bestudeerd. Het leidde tot artikelen in onder meer *Biometrics* en *Statistics in Medicine* (zie bijvoorbeeld Does e.a., 1988), maar ook in de medische tijdschriften zoals *Transplantation* en *Journal of Immunological Methods*. Een tweede (Tjaart Imbos) en derde (Frans Tan) promovendus deden onderzoek naar een stochastisch groeimodel toegepast bij herhaald toetsen van academische kennis. Bij de Medische Faculteit deden ze dat met de voortgangstoets; een instrument om het cognitieve einddoel van de medische studie te meten. Vier keer per jaar werd een voortgangstoets bestaande uit 250 juist-onjuist vragen afgenomen bij alle studenten uit het eerste tot en met het zesde jaar. Een student, die het onderwijs volledig volgt, neemt dus tenminste 24 keer deel aan een voortgangstoets, zodat vastgesteld kan worden hoe de groei in kennis van individuele studenten verloopt. Een van de artikelen in deze onderzoekslijn verscheen in *Psychometrika* (zie Albers e.a., 1989).

Naast onderwijs en onderzoek heb ik mij bij de Universiteit van Maastricht ook beziggehouden met bestuurlijke taken zoals voorzitter van de examencommissie, lid van de faculteitsraad, hoofd van de afdeling, en mede-organisator van twee internationale congressen (the Satellite Meeting on Mathematical Statistics and Probability to the ISI Session in 1985 en the Tenth International Meeting on Clinical Biostatistics in 1989).

In november 1989 vond ik het tijd om wederom een carrièreswitch te maken. Ik koos voor het bedrijfsleven en Jos de Kroon van het CQM, een adviesbureau binnen Philips, stelde mij in de gelegenheid om een echt beroep te leren: adviseur. In deel 2 van dit persoonlijke verslag zal de periode 1990-2021 besproken worden.

Literatuur

Albers, W., Does, R.J.M.M., Imbos, Tj. & Janssen, M.P.E. (1989). A stochastic model applied to repeated tests of academic knowledge, *Psychometrika* 54, 451-466.

Bethlehem, J.G., Does, R.J.M.M. & Gill, R.D. (1977). *Verdelingsvrije methoden bij censurering*, Rapport SN 6/77, Amsterdam: Mathematisch Centrum.

Bethlehem, J.G., Does, R.J.M.M. & Helmers, R. (1978). *Statistische analyse van verdelingen van golfhoogten en waterstanden bij de Oosterschelde*, Rapport SD 113/78, Amsterdam: Mathematisch Centrum.

Does, R.J.M.M. (1982). Berry-Esseen theorems for simple linear rank statistics under the null hypothesis, *Annals of Probability* 10, 982-991.

Does, R.J.M.M. (1983). An Edgeworth expansion for simple linear rank statistics under the null hypothesis, *Annals of Statistics* 11, 607-624.

Does, R.J.M.M., Helmers, R. & Klaassen, C.A.J. (1987). On the Edgeworth expansion for the sum of a function of uniform spacings, *Journal of Statistical Planning and Inference* 17, 149-157.

Does, R.J.M.M. & Klaassen, C.A.J. (1984). The Berry-Esseen theorem for functions of uniform spacings, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 65, 461-471.

Does, R.J.M.M., Strijbosch, L.W.G. & Albers, W. (1988). Using jackknife methods for estimating the parameter in dilution series, *Biometrics* 44, 1093-1102.

Ronald Does is emeritus hoogleraar Industriële Statistiek aan de Universiteit van Amsterdam.
E-mail: r.j.m.m.does@uva.nl

Beter inzicht in energie-impact

De trage totstandkoming van een boycot op olie en gas uit Rusland vanwege de inval in Oekraïne maakt pijnlijk duidelijk hoe afhankelijk Europa is van fossiele brandstoffen. Naast deze directe afhankelijkheid zijn door te toegenomen economische onzekerheid de olie- en gasprijzen naar ongekende hoogte gestegen met een direct effect op de inflatie. In de media wordt druk gediscussieerd over het effect van deze ontwikkelingen op onze economie en de vraag of we een versnelling van de transitie naar duurzamere energieopwekking gaan zien. Het aandeel van energie in de economische prestaties van ons land is volgens het CBS 3,5% van het bruto binnenlands product (BBP)[1], daarmee lijkt de invloed van wijzigingen in de energie beperkt. In hun laatst verschenen rapport[2] onderzoeken economen van het IPCC onder andere de impact van veranderingen in energiegebruik. Zij komen tot de conclusie dat de impact van maatregelen om de gevolgen van klimaatverandering te beperken, zoals de reductie van energiegebruik, minder dan 0,1% van de jaarlijkse groei van het BBP zal zijn. Daarmee is een incentive om nu te verduurzamen nauwelijks aanwezig. Gezien de nu al merkbare invloed van stijgende energieprijzen heb ik zo mijn twijfels of de inschatting van die beperkte invloed wel betrouwbaar is.

Energie als productiefactor

Om het effect van veranderingen in de beschikbare energie op de economie te bepalen wordt vaak gebruik gemaakt van een productiefunctie, bijvoorbeeld de Cobb-Douglas productiefunctie, bekend uit de economieboeken.

$$Y = A \cdot L^{1-\alpha} \cdot K^{\alpha}$$

Deze functie wordt door neoklassieke economen veel gebruikt om de relatie weer te geven tussen input, in dit geval fysiek kapitaal (K) en arbeid (L), en output (Y of wel BBP). De A-factor vertegenwoordigt technologie in het model en kan worden gebruikt om verbeteringen in de efficiency van kapitaal en arbeid te modelleren. De parameter α meet de verandering in output bij een verandering van kapitaal en is dus als output elasticiteit te interpreteren. Met $\alpha = 0{,}30$ zal de output met 0,3% toenemen als kapitaal met 1% toeneemt.

Wat direct opvalt, en dat geldt voor veel van de neoklassieke IPCC modellen, is dat energie als productiefactor ontbreekt. Een manier om energie alsnog op te nemen is door het analoog aan kapitaal en arbeid aan het model toe te voegen[3]

$$Y = A \cdot L^{1-\alpha-\mu} \cdot K^{\alpha} \cdot E^{\mu}$$

De parameter μ van de energiefactor is voor Nederland gelijk aan het aandeel van energie in het BBP. Een afname van energie met 50% leidt dan volgens dit model tot een afname van het BBP met 2,05%. Niet direct iets om je zorgen om te maken? In het bovenstaande model wordt de aanname gedaan dat veranderingen in energie beschikbaarheid geen invloed hebben op de prestaties van kapitaal en arbeid en dat is een te simpele weergave van de werkelijkheid. Machines hebben nu eenmaal energie nodig om te kunnen produceren. Energie is dus geen productiemiddel maar een voorwaarde voor kapitaal en mensen om te kunnen produceren.

Energie is niet marginaal

Om energie beter in de productiefunctie te verwerken kunnen arbeid en kapitaal als functie van energie worden uitgedrukt[4]

$$K(E) = K \cdot E_K \cdot e_K \text{ en } L(E) = L \cdot E_L \cdot e_L$$

met K en L de hoeveelheid kapitaal en arbeid, E_K en E_L de energie-input voor respectievelijk kapitaal en arbeid en e_K en e_L de energie-efficiëntie van beide productiefactoren. Voor arbeid geldt dat $E_L \cdot e_L = \Lambda$ al eeuwen constant is, ongeveer 100 Watt per uur per persoon. Door substitutie van de functies in de Cobb-Douglas productiefunctie volgt:

$$Y = A \cdot \Lambda^{1-\alpha} \cdot (K \cdot E_K \cdot e_K)^{\alpha} \cdot L^{1-\alpha}$$