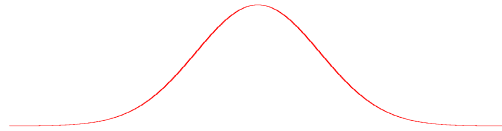
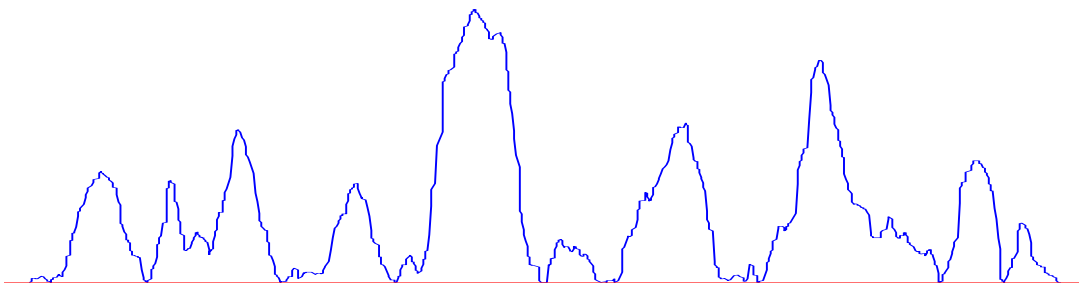
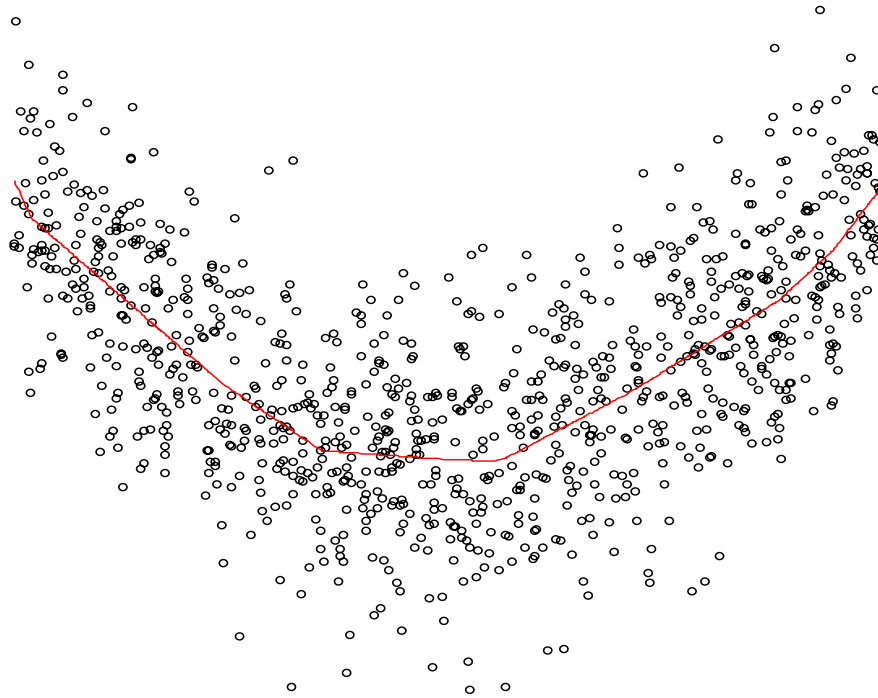


Statistiek in vorm



Intreerede

Prof. dr. ir. G. Jongbloed



9 april 2008

Statistiek in vorm

Prof. dr. ir. G. Jongbloed

INTREEREDE

Uitgesproken ter gelegenheid van de aanvaarding van het ambt van hoogleraar
Mathematische Statistiek aan de faculteit Elektrotechniek, Wiskunde en Informatica
van de Technische Universiteit Delft.

9 april 2008

*Mijnheer de Rector Magnificus,
Leden van het College van Bestuur,
Collegae hoogleraren en
Andere leden van de universitaire gemeenschap,
Zeer gewaardeerde toehoorders,
Dames en heren,*

Wat heeft een onderzoek naar het effect van een nieuwe behandelmethode voor een bepaalde ziekte te maken met het vraagstuk van de opwarming van de aarde? Wat heeft de signalering op de snelweg te maken met het bepalen van een optimale houding van een therapeut ten opzichte van een patiënt? Wat heeft het vroegtijdig indelen van komkommers in kwaliteitsklassen te maken met het inroosteren van personeel in een call center? Sterker nog, wat hebben al deze zaken met elkaar te maken? Ik hoop u vandaag te laten zien dat een antwoord op deze vragen ligt in de officiële naam van mijn leerstoel: mathematische statistiek.

Vanmiddag wil ik u graag een rondleiding geven door mijn vakgebied. Daarbij heb ik me ten doel gesteld dat u na afloop van deze plechtigheid

- weet dat wiskunde iets anders is dan veredeld rekenen,
- een indruk heeft van problemen waar statistiek een prominente rol in speelt,
- weet welke kant we als groep graag op willen met het onderwijs en onderzoek op het gebied van de statistiek aan de TU Delft.

Tijdens deze rondleiding zullen aspecten als historie, elegantie, uitdagendheid en relevantie van mijn vakgebied niet onbesproken blijven.

Wiskunde

Laten we beginnen met de wiskunde. Wanneer men hoort dat je wiskunde studeert of doceert, wordt daar normaal gesproken meteen op gereageerd. Een gebruikelijke reactie, die veel aanwezigen bekend in de oren zal klinken is: "dat vond ik op de middelbare school een vreselijk vak en ik ben blij dat ik er niets meer mee hoeft te doen!". Een andere reactie is: "wiskunde, als je daar maar een knobbel voor hebt, dan kun je het zonder je ervoor in te spannen; heb je die knobbel niet, dan zul je

het nooit leren". Voor luie mensen is het dan een aantrekkelijk vak, want je kunt goede resultaten boeken zonder daarvoor hard te werken. Voor die arme scholier die zijn eerste repetities wiskunde in een onvoldoende ziet resulteren is het doek reeds gevallen: het was niets, het is niets en zal ook nooit meer wat worden... Wanneer in een gesprek naar voren komt dat je onderzoek in de wiskunde doet, wordt de blik geheel glazig en kun je zo maar de vraag krijgen: "Wiskunde, is dat niet allang af?".

Deze reacties illustreren de veelheid aan misvattingen waarmee het vak wiskunde heeft te kampen. Wiskunde is echter een creatief vak, een vak waarin ideeën en intuïties op een heel elegante manier kunnen worden geformaliseerd en uitgewerkt. Het is ook een vak met historie. Ter illustratie wil ik u een heel oud resultaat laten zien, met een wiskundig bewijs ervoor.

Stelling

Stel dat p een getal is waarvoor geldt $p \times p = 2$. Dan is p *irrationaal*. Dat wil zeggen: p is niet te schrijven als het quotiënt van twee gehele getallen.

Voordat we tot het bewijs van deze stelling overgaan, verrichten we eerst wat op het eerste gezicht ongerelateerd voorbereidend werk. Om te beginnen schrijven we de kwadraten van de eerste 5 even getallen op:

$$2 \times 2 = 4 \quad 4 \times 4 = 16 \quad 6 \times 6 = 36 \quad 8 \times 8 = 64 \quad 10 \times 10 = 100.$$

Hetzelfde doen we voor de eerste vijf oneven getallen:

$$1 \times 1 = 1 \quad 3 \times 3 = 9 \quad 5 \times 5 = 25 \quad 7 \times 7 = 49 \quad 9 \times 9 = 81.$$

Misschien valt het u op, misschien ook niet, maar *de kwadraten van de even getallen zijn even, de kwadraten van de oneven getallen zijn oneven*. Dit blijkt iets te zijn wat we heel algemeen mogen stellen, dus niet alleen voor de getallen 1 t/m 10 waarvoor we het gecontroleerd hebben. Inderdaad, beschouw een willekeurig even getal m (neem er maar een in gedachten). Dan kan m geschreven worden als $m = 2 \times n$, voor een geheel getal n . Omdat bij het vermenigvuldigen van getallen de volgorde er niet toe doet, kunnen we stellen dat

$$m \times m = 2 \times n \times 2 \times n = 2 \times (2 \times n \times n).$$

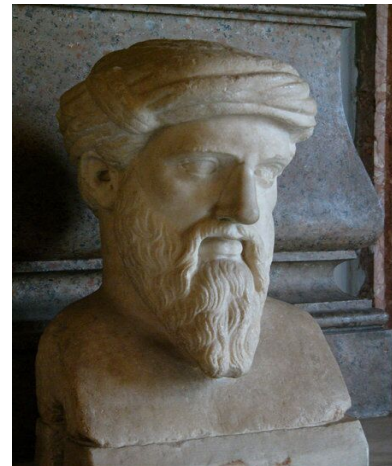
Omdat het getal tussen haakjes een geheel getal is, is $m \times m$ een even getal. Dat het kwadraat van ieder oneven getal oneven is, volgt op een soortgelijke manier. Merk op dat we hier een eigenschap hebben bewezen voor alle even en alle oneven getallen, en dat zijn er oneindig veel. Gelukkig hoeven we de kwadraten niet allemaal uit te rekenen om te zien dat de uitspraak waar is.

Welnu, deze eigenschap van kwadraten kunnen we gebruiken om de genoemde stelling te bewijzen. Dat doen we uit het ongerijmde. Dit houdt in dat we veronderstellen dat de p waarvoor geldt $p \times p = 2$ wel rationaal is. Vervolgens zullen we zien dat die veronderstelling leidt tot iets wat niet waar kan zijn. De enige conclusie is dan dat de veronderstelling onjuist was.

Bewijs. Stel dat $p \times p = 2$ en $p = m/n$ voor twee gehele getallen m en n . We mogen aannemen dat deze getallen niet beide even zijn (anders delen we teller en noemer gewoon door twee totdat dit het geval is). Voor het kwadraat van p kunnen we nu schrijven: $2 = p \times p = m \times m / n \times n$, zodat $m \times m = 2 \times n \times n$. In het licht van de zojuist afgeleide eigenschap van kwadraten, mogen we hieruit concluderen dat m even is (inderdaad, want $m \times m$ is even). Maar als m even is, geldt $m = 2 \times k$ voor een geheel getal k . Hieruit volgt weer dat $n \times n = m \times m / 2 = 2 \times k \times k$ even is, dus n ook. Dit is in tegenspraak met het feit dat n en m niet beide even zijn. Q.E.D.

Wat u hier gezien heeft, is niet zomaar iets. Volgens de legende werd de volgeling van Pythagoras van Samos (inderdaad, de man van $a^2 + b^2 = c^2$) die dit inzicht wereldkundig maakte om het leven gebracht, omdat men binnen de stroming der Pythagoreeërs de overtuiging had dat "alles" als verhouding van twee gehele getallen kon worden gezien.

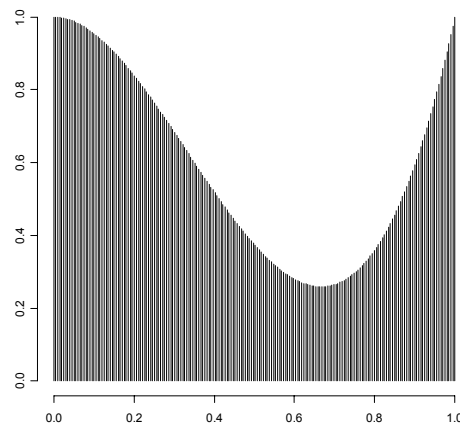
Na deze elegante korte redenering die een fundamentele eigenschap van het getal $\sqrt{2}$ blootlegt, wil ik laten zien hoe in de wiskunde de oppervlakte van een ingewikkeld gebied met behulp van een benaderingsconstructie



*Pythagoras van Samos
(± 600 v. Chr.)*

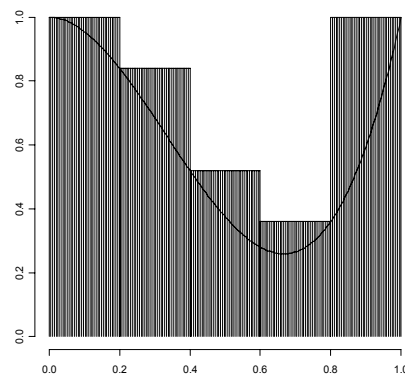
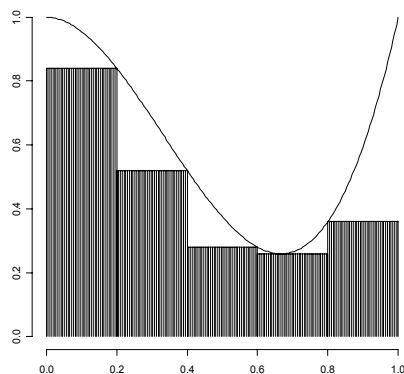
kan worden bepaald.

Zo kunnen we allemaal de oppervlakte berekenen van een rechthoek: lengte keer breedte. Voor een gebied dat kan worden opgesplitst in verschillende rechthoeken, kan deze oppervlakte ook worden berekend. We tellen gewoon de oppervlakten van de afzonderlijke rechthoeken bij elkaar op. Hoe bereken je echter de oppervlakte van een gebied dat er veel ingewikkelder uitziet, als in de onderstaande figuur?



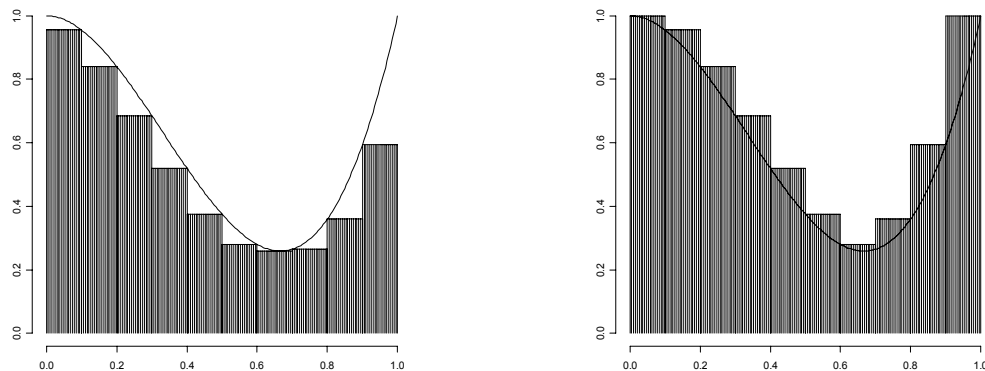
Het gebied waarvan de oppervlakte bepaald moet worden.

Welnu, we benaderen het gebied gewoon steeds nauwkeuriger door gebieden waarvan we de oppervlakte gemakkelijk kunnen bepalen. Heel concreet kunnen we het gebied benaderen door de 'blok vormige gebieden' in de onderstaande figuren.



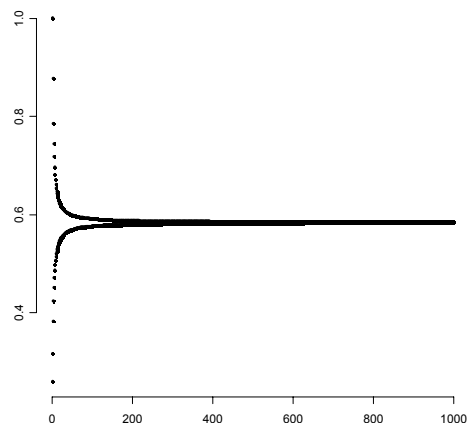
Een onder- en overschatting van de gevraagde oppervlakte op basis van vijf blokken.

Duidelijk is, dat de oppervlakte van het gearceerde gebied in de linkerfiguur kleiner is dan de oppervlakte die we willen bepalen. Evenzo is duidelijk dat de oppervlakte van het gearceerde gebied in de rechterfiguur de te bepalen oppervlakte overschrijdt. Van de gearceerde gebieden kunnen we de oppervlakte eenvoudig bepalen en hieruit mogen we concluderen dat de gezochte oppervlakte tussen de waarden 0.4519 en 0.7440 in ligt. Nauwkeuriger benaderingen worden gegeven in de onderstaande figuur. Daar zijn tien in plaats van vijf blokken gebruikt. De oppervlakten van de gearceerde gebieden in laten zich weer eenvoudig berekenen en we kunnen concluderen dat de gezochte oppervlakte ligt tussen 0.5134 en 0.6610.



Een onder- en overschatting van de gevraagde oppervlakte op basis van tien blokken.

We kunnen ook nagaan wat de oppervlaktes zijn van de benaderingen wanneer het aantal blokken groter en groter wordt. De onderstaande figuur geeft deze waarden afhankelijk van het aantal blokken dat wordt gebruikt.

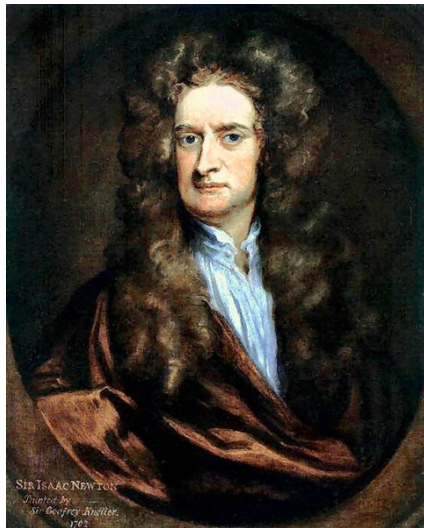


De schattingen van de gevraagde oppervlakte, afhankelijk van het aantal blokken.

Het is duidelijk dat die benaderingen 'ergens heen gaan', de boven- en ondergrens voor de oppervlakte convergeren naar een gezamenlijke *limietwaarde*. Deze kan worden berekend en bedraagt $7/12$. De wiskundige formulering van dit resultaat is

$$\int_0^1 (5x^3 - 5x^2 + 1) dx = 7/12 = 0.58333...$$

Een formule als deze werkt afschrikwekkend op sommige mensen. Deze formule wil echter niets meer of minder zijn dan een efficiënte manier om de intuïtief aansprekende constructie van zojuist uit te drukken. Het is geen abracadabra formule, maar een notatie voor een heel benaderingsproces dat zich als het ware voor je ogen afspeelt. De differentiaal- en integraalrekening, waar de hier gegeven constructie deel van uitmaakt, is in de zeventiende eeuw ontwikkeld door mensen als Isaac Newton en Gottfried Wilhelm von Leibniz en de invloed van deze theorie op de ontwikkeling van de wetenschap kan niet overschat worden.



Isaac Newton (1643-1727) en Gottfried Wilhelm von Leibniz (1646-1716)

De beide klassieke voorbeelden die ik heb gegeven, aangaande de irrationaliteit van $\sqrt{2}$ en het berekenen van een oppervlakte, illustreren dat wiskunde een vak is waarbij je op basis van elementaire uitgangspunten en logische redeneringen een heel bouwwerk van waarheden construeert. Dagelijks wordt de kennis op het gebied van de

wiskunde vergroot. Dagelijks komen er ook nieuwe interessante problemen bij.

Heel veel problemen in wetenschap, technologie en het dagelijks leven, kunnen in essentie worden teruggebracht tot wiskundige problemen. Het is een belangrijke taak van de (toegepast) wiskundige om zulke problemen uit de praktijk terug te brengen tot de essentie en vervolgens door een combinatie van intuïtie, creativiteit en wiskundig vakmanschap op te lossen.

Door allerlei enthousiaste mensen wordt al jaren een grote inspanning geleverd om het 'ware', dat is veelzijdige gezicht van de wiskunde aan scholieren te laten zien, om hen voor een studie wiskunde te interesseren. De vooroordelen ten opzichte van de wiskunde, als zou je als vervolg hierop alleen 'nog maar' wiskundeleraar kunnen worden (een prachtig vak overigens!) of absoluut het contact met de werkelijkheid verliezen, zijn echter hardnekkig. We horen nogal eens dat studenten de keuze voor een vakgebied waar ze later in willen gaan werken graag uitstellen. Aan deze wens wordt dan gehoor gegeven door een 'brede bacheloropleiding' aan te bieden. Ik durf te stellen dat voor zulke studenten de studie wiskunde een heel aantrekkelijke optie is. Afgestudeerde wiskundigen komen terecht in de meest uiteenlopende werkelden omdat wiskunde, zij het soms wat onder de oppervlakte, in zoveel vakgebieden prominent aanwezig is.

Aan het begin heb ik drie doelen geformuleerd voor deze bijeenkomst. Het eerste was u een mijns inziens wat eerlijker indruk te geven van wat wiskunde zoal inhoudt, dan de indruk zoals die in brede kring bestaat. Ik heb geprobeerd te laten zien dat de wiskunde een vak met historie is, waarin praktisch relevante en intellectueel uitdagende problemen creatief en elegant kunnen worden opgelost. Het tweede doel spitst zich meer toe op mijn vakgebied binnen het landschap van de wiskunde: de statistiek. Waar komt de term vandaan? Wat houdt het in? Waar wordt het gebruikt en wat is er leuk aan?

Statistiek

De term statistiek als aanduiding van een vakgebied heeft een opmerkelijke ontstaansgeschiedenis. In zijn collegeserie over de geschiedenis van de statistiek, gehouden in University College London,

beschrijft Karl Pearson deze in drie stappen: eerst de oorsprong van het woord, vervolgens het ontstaan en de teloorgang van het Duitse vakgebied *Statistik* en tot slot de naamsverandering van het Engelse vakgebied *Political Arithmetic* in Statistiek.

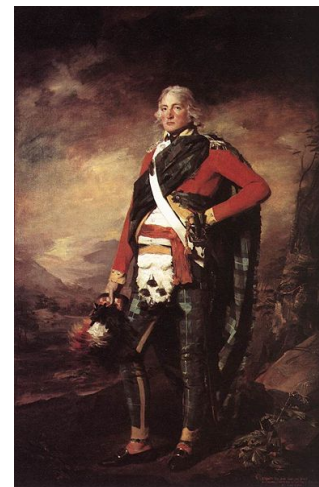
Volgens Pearson was het de Italiaanse Giovanni Botero die met zijn boek *Della ragione di stato* rond 1590 aanleiding gaf tot het gebruik van het bijvoeglijk naamwoord statistisch in de zin van 'met de staat te maken hebbend'. In de zeventiende eeuw ontwikkelde captain John Graunt in Engeland het vakgebied Political Arithmetic, waarin men wiskunde gebruikte om op basis van gegevens antwoorden op problemen van de staat te krijgen. Een eeuw later koos Gottfried Achenwall in Duitsland de naam Statistik voor een wetenschap die zich bezighield met staatkunde, echter in het geheel niet met gegevens. Aan het eind van de achttiende eeuw vond Sir John Sinclair dat statistics een betere naam was voor het vakgebied Political Arithmetic en werd die naam gewijzigd. Ook in Duitsland kwam deze wetenschap van de grond en wel onder de naam Statistik. Gedurende een korte tijd hebben er in Duitsland dus twee heel verschillende vakgebieden bestaan onder de naam Statistik. Uiteindelijk was het de 'oude' Political Arithmetic die internationaal verder ging onder de naam statistiek.



G. Botero (1544-1617)



J. Graunt (1620-1674)



J. Sinclair (1754-1835)

Analyse van gegevens, algemener dan alleen die van de staat, is nog steeds het centrale thema binnen de statistiek, al kunnen vanuit verschillende invalshoeken de accenten wat verschillen. Zodanig beschouwd, zal het u niet verbazen dat statistiek op heel veel plaatsen

aan de orde is. Laat ik enkele voorbeelden geven van situaties waarin statistiek een rol speelt.

In de medische wetenschap wordt onderzoek gedaan naar het effect van een bepaalde behandeling. De vraag kan concreet zijn: "Werkt een nieuw medicijn beter tegen een aandoening dan een tot op heden gebruikt medicijn?". Er wordt een studie gedaan waarbij een bepaalde groep patiënten met het nieuwe medicijn wordt behandeld en een andere, in alle overige opzichten vergelijkbare groep met het oude medicijn. Na verloop van tijd wordt het effect van de beide behandelingen gemeten. Dit levert gegevens op die moeten resulteren in een antwoord op de oorspronkelijke vraag of het nieuwe medicijn beter is dan het oude. Deze zogenaamde klinische studies, waarvoor standaardprotocollen bestaan, zijn niet weg te denken uit de medische wetenschap. Over deze standaardprotocollen is overigens weer een discussie ontstaan naar aanleiding van de resultaten van de PROPATRIA studie naar het effect van probiotica op mensen met een acute alvleesklierinfectie (Besselink *et al.*, 2008).

Een heel ander voorbeeld uit de gezondheidszorg is positron emissie tomografie (PET). In de detectoren van de PET-scanner worden veel gegevens in de vorm van tellingen van fotonen verzameld. Op basis van de verkregen aantallen, geteld *buiten* het lichaam, moeten vragen met betrekking tot de bloedconcentratie *binnen* het lichaam worden beantwoord (Mesina *et al.*, 2003).



Een andere context waar men op basis van bepaalde gegevens een specifieke vraag beantwoord wil zien, is de verkeerskunde. Om stoplichten in zekere zin optimaal te kunnen afstellen, is kennis van het toevalsgedrag van de aankomende verkeersstromen nodig. Op basis van metingen kan deze kennis worden verkregen.

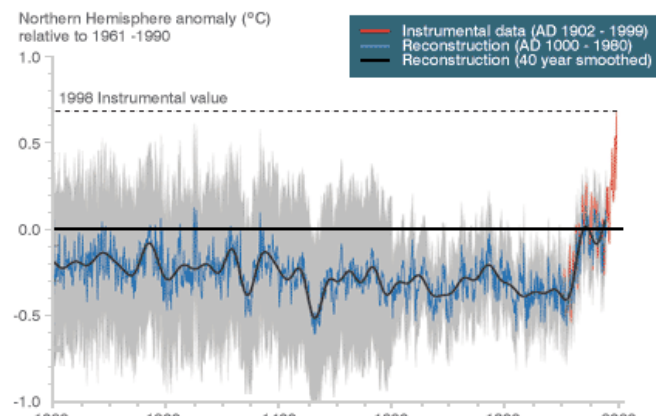
Weer een ander voorbeeld hangt samen met de wereldwijde klimaatdiscussie. Was het laatste decennium van de twintigste eeuw het warmste van het tweede millennium of niet? Vrij nauwkeurige lokale temperatuurmetingen zijn beschikbaar vanaf 1850. Voor 1850 zijn er ook wel gegevens, maar niet op grote schaal, vaak indirect en minder systematisch. We hebben weer een vraag en gegevens, zij het niet de gegevens die we graag zouden willen hebben om de vraag te beantwoorden.

Er zijn talloze andere praktische situaties en problemen te noemen waarin statistiek een belangrijke rol speelt. Denk aan kwaliteitsmodellering in de voedselketen, elektronenmicroscopie in de nanotechnologie, oudheidsbepaling van gesteenten en seismisch grondonderzoek in de geologie, prijsbepaling van opties op de financiële markten, vaststellen van pensioenpremies in het actuarieel, het onderscheiden van genetisch- en omgevingsbepaalde aandoeningen in de genetica, enzovoort.

Alle genoemde voorbeelden hebben met elkaar gemeen dat er een concrete vraag is en dat er gegevens beschikbaar zijn of kunnen worden verkregen die informatie bevatten over de gestelde vraag. De brug tussen de vraag en het antwoord wordt geslagen door het centrale thema binnen de mathematische statistiek: het *wiskundig model met daarin een stochastische (kanstheoretische) component*. De gegevens worden opgevat als zijnde gegenereerd door een kansmechanisme waar nog onbekende parameters in voorkomen. De concrete vraag uit de praktijk kan binnen het model worden vertaald in een vraag met betrekking tot de waarden van de onbekende parameters. De gegevens kunnen vervolgens gebruikt worden om uitspraken te doen over de waarden van de onbekende parameters, die weer leiden tot een antwoord op de concrete vraag uit de praktijk.

Om te illustreren dat het modelleren van de gegevens een niet-triviale taak is waar grote gevolgen aan vast kunnen zitten, wil ik terugkomen op de vraag of de jaren negentig van de vorige eeuw het warmste decennium vormden van het vorige millennium. Deze vraag en verschillende antwoorden hierop, hebben op wetenschappelijk en politiek gebied tot veel commotie geleid.

In 1998 verscheen een artikel van Mann, Bradley en Hughes in Nature, waarin gesteld wordt dat het verloop van de gemiddelde temperatuur op het noordelijk halfrond in de loop van het afgelopen millenium een specifieke vorm heeft. Deze vorm wordt aangeduid als hockeystick.



De omstreden hockeystick grafiek.

Het idee is dat de gemiddelde temperatuur op het noordelijk halfrond tussen het jaar 1000 en 1900 min of meer constant is geweest, maar sinds 1900 een duidelijke stijging vertoont. Nu is het niet eenvoudig om voor de periode 1000-2000 een plaatje te maken van deze gemiddelde temperatuur uitgezet tegen het jaar. Dit met name vanwege twee redenen. De eerste reden is dat niet meteen duidelijk is wat moet worden verstaan onder 'gemiddelde temperatuur op het noordelijk halfrond' voor een bepaald jaar. Een andere moeilijkheid is dat de gegevens nogal divers van aard zijn: directe meetgegevens vanaf ongeveer 1850 en veel zogenaamde proxies, indirecte gegevens, gehaald uit schriftelijke bronnen, jaaringen van bomen, gletsjers, korallen en dergelijke. Het modelleren van de beschikbare data en het trekken van conclusies hieruit is een complex probleem.



Verschillende bronnen van gegevens.

Het genoemde Nature artikel heeft nogal wat consequenties gehad. In 2001 nam het Intergovernmental Panel on Climate Control (IPCC) van de Verenigde Naties de hockeystick grafiek op in haar samenvatting, met

alle beleidsmatige gevolgen van dien. Vervolgens ontstond een verhitte internetdiscussie over de grafiek. Deze bracht de U.S. House Committee on Energy and Commerce ertoe een commissie van statistici in te stellen om de validiteit van de hockeystickgrafiek na te gaan. Deze commissie Wegman stelt in haar rapport dat zij de conclusie van Mann, dat het laatste decennium van het vorige millennium het warmste was, niet deelt. Verder komt de commissie met de kritiek dat de onderzoekers in het toepassingsveld van paleoklimatologie verzuimd hebben samenwerking te zoeken met 'mainstream' statistici.

"Voetballer x in vorm is een aanwinst voor ieder team", las ik eens. Hetzelfde geldt voor statistici in een multidisciplinair onderzoeksteam waar op basis van gegevens conclusies moeten worden getrokken. In zo'n veelomvattend project kan een mathematisch statisticus als eenling geen potten breken. Samenwerking met mensen uit andere disciplines is noodzakelijk om goede vorderingen te maken. Wanneer zo'n project echter zonder mathematisch statistici wordt uitgevoerd, laat men kansen liggen. In het hierboven genoemde klimaatvoorbeeld heeft het aan de nodige samenwerking ontbroken.

Nu we in Nederland een onderzoeksprogramma *kennis voor klimaat* hebben, waarin effecten van de klimaatverandering moeten worden ingeschat en vertaald naar beleid, is het belangrijk dat mathematisch statistici vanaf het begin nauw bij de activiteiten van dat programma worden betrokken. Als EURANDOM, het in Eindhoven gevestigde Europese onderzoeksinstituut op het gebied van kansrekening en statistiek waar ik senior fellow ben, zijn we in contact met de Klimaat Kennis Faciliteit en hopen we een fundamentele bijdrage te kunnen leveren aan het welslagen van dit voor ons land zo belangrijke project.

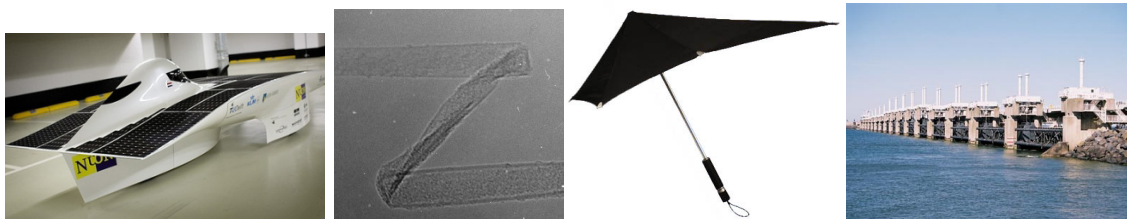
Een wiskundig model met een kanstheoretische component en onbekende parameters is het centrale object binnen de mathematische statistiek. Een object dat de brug moet slaan tussen data en conclusie, waarbij modelaannames expliciet worden geformuleerd. Er zijn ook andere vakgebieden, sommige tamelijk jong, die zich bezighouden met het beantwoorden van vragen op basis van data. Hierbij valt te denken aan gebieden als machine learning, neurale netwerken, bioinformatica, signaalanalyse, econometrie. Onder statistici hoor je nog wel eens dat mensen in die velden bekende concepten uit de mathematische statistiek herontdekken en voorzien van een nieuwe naam om deze vervolgens te gebruiken bij het oplossen van hun problemen, oude wijn in nieuwe

zakken dus. Ten dele mag dit waar zijn. Binnen deze disciplines zijn en worden echter ook nieuwe interessante modellen voor complexe problemen opgesteld, bestudeerd en toegepast. Wat mij betreft kunnen we veel van elkaar leren.

Ik heb in dit deel willen onderstrepen dat statistiek, onder die naam of een andere, op zeer veel plekken te vinden is. Het beantwoorden van vragen op basis van verkregen gegevens komt op heel veel terreinen van de wetenschap, de politiek en het dagelijks leven voor. De behoefte aan wiskundige modellen om van vraag en gegevens tot een antwoord te komen is navenant. Als een bestaand wiskundig model voldoet, is dat mooi. Moet er een nieuw model ontwikkeld worden, dan is dat, voor ons althans, mooier.

Statistiek aan de TU Delft

Is statistiek belangrijk voor de TU Delft? Wanneer u in het nieuws iets hoort over de TU Delft, gaat dat vaak over technologische of natuurwetenschappelijke hoogstandjes. Zo kunt u gelezen hebben over de Nuna die verschillende opeenvolgende jaren het snelst was in de World Solar Challenge in Australië. Het kan ook zijn dat u gehoord heeft van het manipuleren van deeltjes op de zogenaamde nanoschaal (een miljoenste van een millimeter) of van de windbestendige paraplu (de Senz). Wellicht heeft u wat langer geleden gehoord over chirurgische ingrepen op afstand met behulp van robots of nog langer geleden over het bouwen van de Deltawerken in Zeeland. Als het gaat om spectaculaire, voor een breed publiek tot de verbeelding sprekende ontwikkelingen, dan doet de TU Delft het goed. Wat betreft de statistiek vrees ik dat ik u geen in de media breed uitgemeten aansprekende voorbeelden uit de statistiek in herinnering kan roepen.



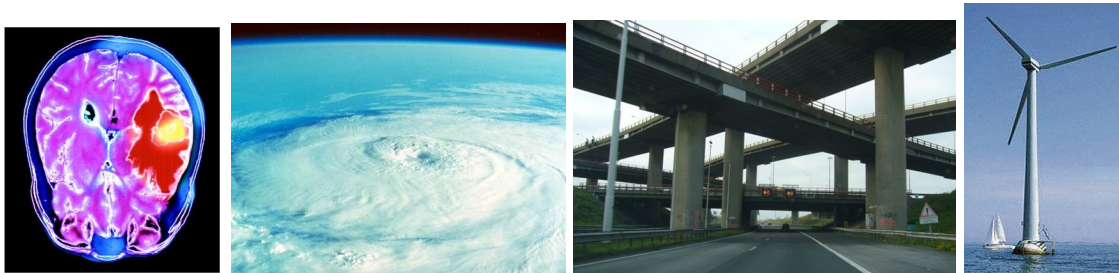
Nuna, nano, senz en delta.

Nogmaals de vraag: is statistiek belangrijk voor de TU Delft? Binnen de TU Delft worden sinds een aantal jaren vanuit het College van Bestuur

zwaartepunten geformuleerd op het gebied van onderzoek. Onlangs zijn de volgende maatschappelijk relevante zwaartepunten (International Research Platforms) geformuleerd voor de jaren 2009-2017:

- Gezondheid
- Milieu
- Infrastructuur
- Energie

Statistiek zit er niet direct bij. Toch stemt dit mij niet somber. Wanneer u zich de voorbeelden van situaties herinnert die ik schetste als toelichting op de breedte van het vakgebied statistiek, dan is duidelijk dat statistiek in deze zwaartepunten prominent aanwezig is. Niet als kernactiviteit, maar als essentieel onderdeel waarmee in feite projecten kunnen staan of vallen.



De vier zwaartepunten in beeld.

Omdat statistiek op zoveel terreinen van de samenleving en de (technische) wetenschappen een rol speelt, is kennis van statistiek onontbeerlijk voor de meeste ingenieurs. Dit inzicht was er ook al bij het College van Bestuur in 1952, toen de leerstoel mathematische statistiek aan deze universiteit werd ingesteld, al voordat er een opleiding tot wiskundig ingenieur bestond. Sinds die tijd heeft mijn vakgebied stevast een positie gehad binnen de TU Delft. Graag wil ik, als het gaat over statistiek aan de TU Delft, afzonderlijk aandacht schenken aan de twee activiteiten die wij op dat gebied ontplooiën: onderzoek en onderwijs.

Onderzoek

Sinds de instelling van de leerstoel mathematische statistiek is er (soms meer en soms wat minder) samenwerking geweest met andere onderzoeksgroepen binnen de TU Delft. Als leerstoelgroep willen we graag de contacten met zulke groepen in stand houden en uitbreiden. Dit willen we o.a. doen door het gericht intensiveren van gezamenlijke masterprojecten, met accent op statistische problemen die om maatwerk vragen en niet meteen passen in een bekend model. Ons inziens kunnen

op die manier voor onze studenten zeer interessante masterprojecten worden geformuleerd. Ook zijn we als leerstoelgroep begonnen met het uitnodigen van onderzoekers op andere terreinen die tegen problemen van statistische aard aanlopen. Wij hopen met hen gezamenlijke projecten te kunnen aangaan waarbij in voor andere disciplines relevante problemen verantwoord statistiek wordt bedreven en waar wij tegen problemen zullen aanlopen die ook vanuit wiskundig oogpunt interessant zijn om te onderzoeken.

Toen professor Hemelrijk¹, de eerste bekleeder van de leerstoel mathematische statistiek in Delft, in 1952 zijn ambt officieel aanvaardde met zijn oratie² "Enkele aspecten der mathematische statistiek", zei hij: "Indien de theoretische kern van een tak der toegepaste wiskunde verwaarloosd wordt, verliest deze tak zijn levenssappen en is gedoemd te verdorren". Nu, 56 jaar na dato, neem ik die opmerking graag voor mijn rekening. Voor de wiskunde in het algemeen en statistiek in het bijzonder. Hoe belangrijk statistiek ook is in allerlei andere disciplines en hoe graag we ook een bloeiende multidisciplinaire samenwerking nastreven met (Delftse) onderzoeksgroepen, blijft het een deelgebied van de wiskunde waar wiskundig uitdagende problemen liggen die erom vragen te worden opgelost. Aan dat deel van het 'kernonderzoek' wil ik ook enkele woorden wijden

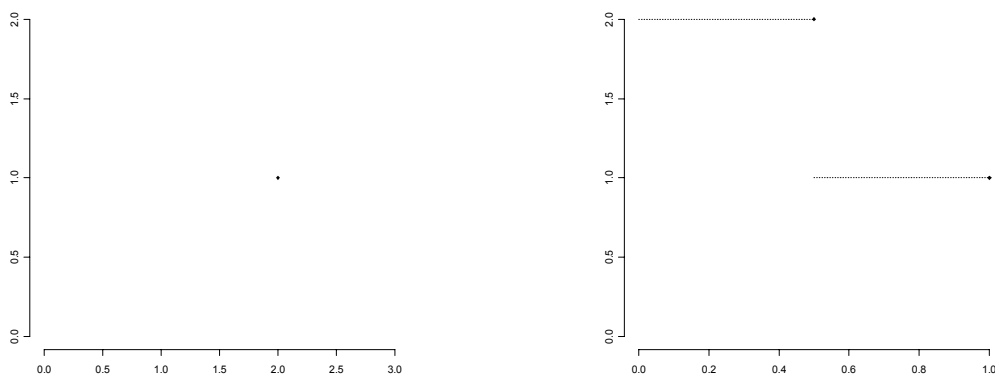
Zoals ik eerder betoogd heb, kenmerken de wiskundige modellen waar in de mathematische statistiek aan wordt gewerkt zich door de aanwezigheid van een kanstheoretische component en zogenaamde 'parameters'. Deze (onbekende) parameters moeten op basis van door het model gegenereerde gegevens worden achterhaald. Een elementair en klassiek voorbeeld is het volgende. Duizend maal wordt een experiment gedaan waarbij twee uitkomsten mogelijk zijn: 1 (succes) en 0 (mislukking). We zijn geïnteresseerd in de kans op succes wanneer het experiment opnieuw wordt uitgevoerd. Die kans, genoteerd als p , is een onbekende parameter met waarde tussen de 0 en 100%, die moet worden geschat op basis van de geobserveerde rij van lengte duizend, bestaande uit nullen en enen. Dit model vormt de basis voor analyses die weer zullen worden gemaakt wanneer zogenaamde exit polls zullen worden gebruikt om in een vroeg stadium te bepalen of de volgende

¹ Prof. dr. J. Hemelrijk, hoogleraar mathematische statistiek aan de TH Delft, 1952-1960.

² Met dank aan ir. T.C.A. Mensch, die me van de teksten van enkele oude Delftse oraties voorzag.

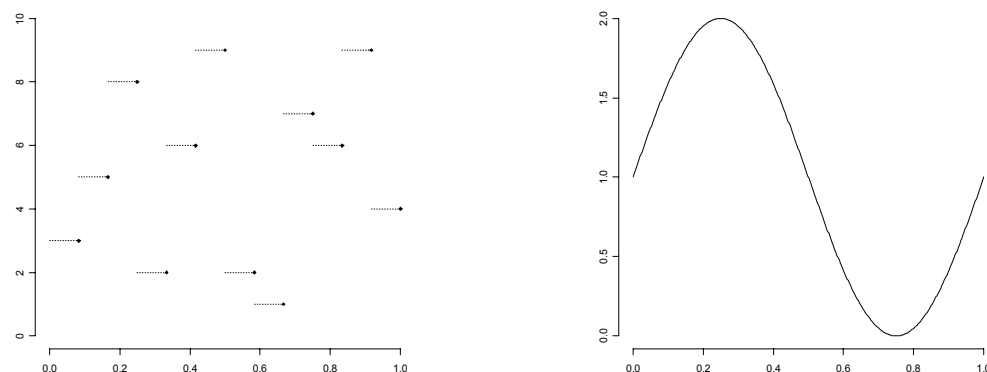
president van de Verenigde Staten een Democraat of een Republikein zal zijn. Ook bij het analyseren van de eerder genoemde data uit het probiotica onderzoek van het UMC in Utrecht, speelt dit model een rol.

Interessanter wordt het wanneer de parameter p hoog- of zelfs oneindig dimensionaal is. Bij twee- of driedimensionale parameters kan worden gedacht aan een onbekend punt in het vlak of in de ruimte. Alternatief kan (in het geval van drie dimensies) worden gedacht aan een grafiek met horizontaal de waarden $1/3$, $2/3$ en 1 en vertikaal de waarden van de drie componenten van de parameter.



Twee representaties van het punt $(2,1)$ in de tweedimensionale ruimte.

De eerste representatie ligt meer voor de hand, maar is moeilijker te realiseren in dimensies hoger dan drie (alhoewel de geometrische intuïtie in hoogdimensionale ruimtes van het grootste belang is). Dit is anders voor de tweede, zogenaamde functierepresentatie. Zo zien we hieronder de functierepresentaties van een twaalfdimensionale- en een oneindigdimensionale parameter.



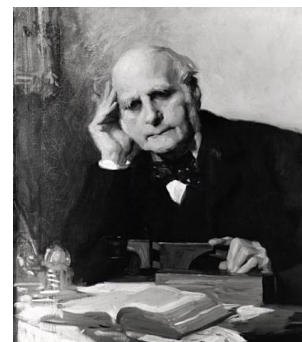
Representatie van het punt $(3,5,8,2,6,9,2,1,7,6,9,4)$ in de twaalfdimensionale ruimte en van een punt in een oneindigdimensionale ruimte.

In veel gevallen is zo'n functie niet maar zo een willekeurige functie, maar is bekend dat de grafiek bepaalde eigenschappen heeft. Vaak zijn dit eigenschappen die kwalitatief van aard zijn, eigenschappen die zich manifesteren in de *vorm* van de grafiek van de functie. Een grote klasse van problemen waar zulke kwalitatieve informatie beschikbaar is, wordt gevormd door de zogenaamde inverse problemen. Deze klasse bevat veel (ook meerdimensionale) censureringsproblemen, waarbij bepaalde aspecten van de gegevens niet of slechts ten dele beschikbaar zijn. Interessante aspecten van zulke problemen zijn

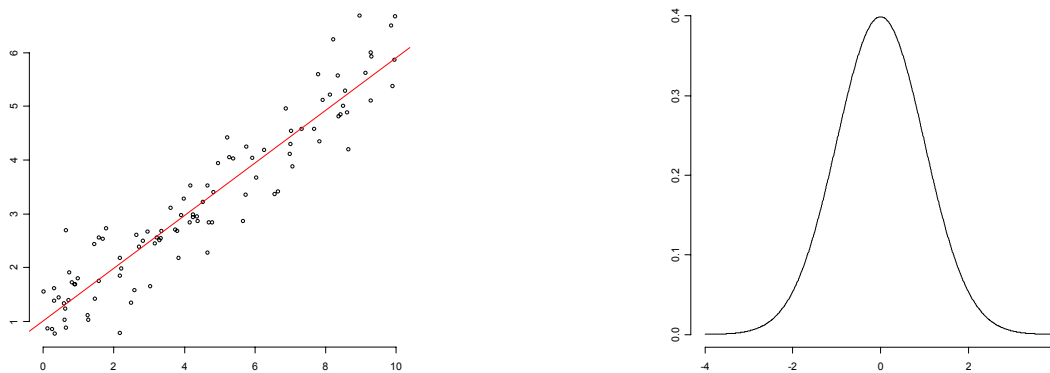
- de maximaal haalbare nauwkeurigheid waarmee aspecten van de parameters kunnen worden geschat,
- de definitie van een optimale schattingsmethode,
- computationele aspecten van schatters die vaak als oplossing van een optimalisatieprobleem over een ruimte van functies zijn gedefinieerd,
- eigenschappen van (optimale of niet optimale) schatters, met name de nauwkeurigheid als de steekproefomvang naar oneindig gaat.

Deze vragen leggen een link vanuit de mathematische statistiek naar andere deelgebieden van de wiskunde, zoals kansrekening, optimalisatietheorie en approximatietheorie.

Als voorbeeld ter illustratie van problemen waarbij vormrestricties een rol spelen, wil ik de aandacht vestigen op enkele *regressiemodellen*, die deze naam danken aan een artikel van Sir Francis Galton in 1886. Uitgangspunt is een scatterplot of puntenwolk, waarin punten worden weergegeven die corresponderen met objecten of individuen. Op de horizontale as van zo'n scatterplot staat bijvoorbeeld de lengte van een persoon en op de verticale as het gewicht van de betreffende persoon. In de figuur links bovenaan de volgende pagina is een scatterplot te zien. De puntenwolk lijkt zich rond een denkbeeldige rechte lijn te concentreren. Deze lijn, de regressielijn, kan op basis van de gegevens worden geschat. Doen we dat met behulp van de klassieke methode der kleinste kwadraten, dan krijgen we de lijn die in de linker figuur is weergegeven.



Sir F. Galton
(1822-1911)

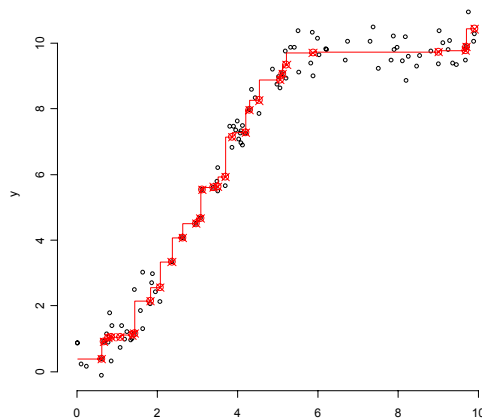


Scatterplot met lineaire kleinste kwadraten schatting en de Gauss curve.



*C.F. Gauss
(1777-1855)*

Wanneer we ons afvragen hoe nauwkeurig de benadering is voor de onbekende onderliggende lijn, is het een klassiek resultaat dat de gemaakte fout van de orde $1/\sqrt{n}$ is, waar n het aantal punten in de plot is. Dit wil ruwweg zeggen dat we, om de benadering twee keer zo nauwkeurig te maken, het aantal metingen met een factor vier dienen te vermenigvuldigen. Kijken we naar het kanstheoretische gedrag van de herschaalde afwijking, dan kan deze worden beschreven door de zogenaamde Gauss verdeling, de verdeling die een dichtheid heeft met de karakteristieke klok vorm als weergegeven in de bovenstaande figuur rechts.

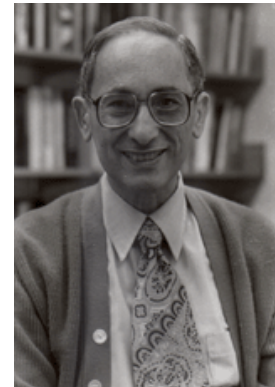


Scatterplot met stijgende kleinste kwadraten schatting.

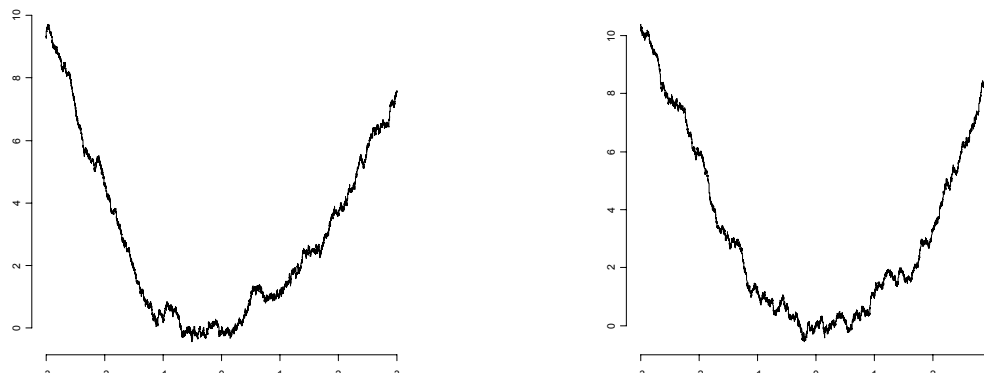
Interessanter wordt het, wanneer we niet aannemen dat de denkbeeldige lijn een rechte lijn is, maar alleen dat het een kromme lijn

is die stijgt. De voorgaande figuur laat een puntenwolk zien die door zo'n model kan worden beschreven. De stijgende curve in deze figuur is de kleinste kwadraten benadering voor de denkbeeldige curve waarin we geïnteresseerd zijn.

De theorie van isotone regressie stelt ons in staat deze benadering te berekenen. Ook voor deze benadering kan worden nagegaan hoe nauwkeurig deze is. De fout die gemaakt wordt, is in dit geval van de orde $1/\sqrt[3]{n}$. Concreet betekent dit dat om de fout twee keer zo klein te krijgen, er acht keer zoveel gegevens moeten worden verzameld. In statistische zin is het dan ook een moeilijker probleem dan het voorgaande. Het kanstheoretisch gedrag van de herschaalde afwijking wordt bepaald door de zogenaamde Chernoff verdeling, genoemd naar Herman Chernoff die deze verdeling in zijn artikel uit 1964 introduceerde. Deze verdeling is gecompliceerder dan de Gauss verdeling. Zo is bijvoorbeeld geen uitdrukking voor de kansdichtheid van deze verdeling bekend. Wel kan de Chernoff verdeling heel elegant constructief worden gekarakteriseerd in termen van de zogenaamde klassieke tweezijdige Brownse beweging met een parabolische drift. De onderstaande figuur laat twee realisaties van dit grillige stochastische proces op het interval $[-3,3]$ zien.

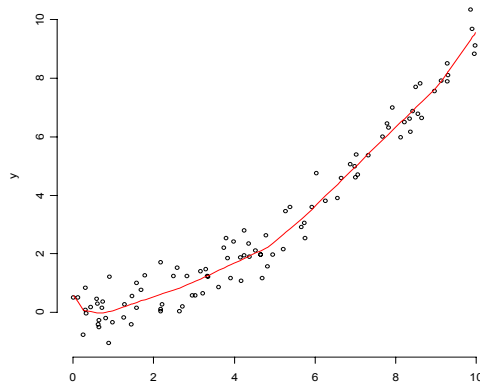


H. Chernoff



Twee realisaties van tweezijdige Brownse beweging met parabolische drift.

De scatterplot op de volgende pagina laat weer een andere vorm zien. De denkbeeldige lijn door deze puntenwolk is *convex*. Dit wil zeggen dat de curve verder naar rechts steeds 'steiler omhoog loopt'. We kunnen ook een kleinste kwadraten benadering berekenen voor die curve, enkel gebruikmakend van de vormrestrictie van convexiteit.



Scatterplot met convexe kleinste kwadraten schatting.

Voor deze benadering van de curve kunnen we laten zien dat het aantal gegevens met een factor $4\sqrt{2} \approx 5.66$ moet worden vermenigvuldigd om een twee keer zo nauwkeurige benadering te krijgen. Dit laat zien dat het probleem van het schatten van een convexe regressiefunctie vanuit statistisch oogpunt gemakkelijker is dan het schatten van een monotone regressiefunctie. Vanuit wiskundig oogpunt is dit convexe probleem echter ingewikkelder dan het monotone probleem. Dat geldt voor de

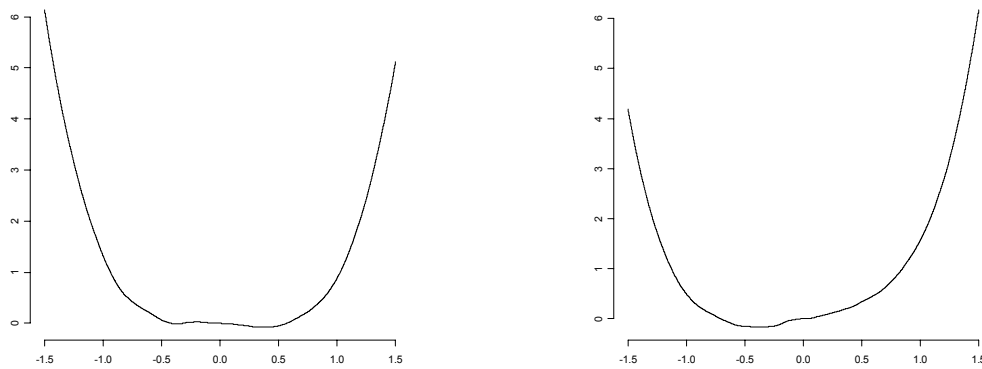


P. Groeneboom

computationele aspecten (waar zich spannende zoektochten in een hoogdimensionale ruimte ontspinnen), maar meer nog voor de asymptotische. Zo wordt het kanstheoretisch gedrag van de herschaalde afwijking beschreven door de zogenaamde Groeneboomverdeling (Groeneboom *et al.*, 2001). Deze verdeling kan ook worden gekarakteriseerd in termen van een 'inhullende' van een stochastisch proces dat kan worden uitgedrukt in de eerder genoemde Brownse

beweging: de geïntegreerde tweezijdige Brownse beweging met een vierdemachts drift. Deze karakterisering is impliciet dan die van de Chernoff verdeling. De figuren op de volgende bladzijde laten twee realisaties zien van het genoemde stochastische proces op het interval $[-1.5, 1.5]$.

In recent onderzoek is deze Groeneboomverdeling ook in andere problemen met zogenaamde $n^{2/5}$ -asymptotiek naar voren gekomen (Jongbloed en Van der Meulen, 2008, Balabdaoui *et al.*, 2008).



Twee realisaties van geïntegreerde Brownse beweging met vierdemachts drift.

Met deze drie regressieproblemen heb ik wat willen overbrengen van de ontwikkeling van theorie voor laagdimensionale statistische modellen (het lineaire regressiemodel) naar theorie voor hoogdimensionale modellen waarbij vormrestricties een rol spelen (monotone en convexe regressie). Er zijn ook zeer interessante één- en meerdimensionale censureringsproblemen, inverse problemen en problemen in het financiële toepassingsveld waarbij vormrestricties een rol spelen. Als statistiegroep zullen we dat soort problemen aanpakken en oplossen, waarbij ik er van overtuigd ben dat we nog op zeer uitdagende wiskundige problemen zullen stuiten.

In de statistiek worden problemen opgelost, maar er komen ook steeds meer problemen bij. Voor mensen die bezig zijn met gegevens is het belangrijk om een degelijke basiskennis van het vak te hebben. Ook is het nodig dat studenten zich puur in het vakgebied gaan verdiepen om de vele complexe problemen op het gebied het hoofd te bieden. Dit brengt me op het volgende, in mijn ogen zeer belangrijke aandachtspunt als het gaat om 'statistiek aan de TU Delft': het onderwijs.

Onderwijs

Zoals eerder opgemerkt, maakt statistiek al sinds een lange tijd deel uit van het basiscurriculum van vrijwel alle opleidingen aan de TU Delft. Die statistiek wordt voor het grootste gedeelte gegeven vanuit de leerstoelgroepen Kansrekening en Statistiek.

Een natuurlijke vraag die zich aandient is: "Wat moet een Delftse ingenieur weten van statistiek?". We kunnen dan denken aan een

zekere basiskennis, bestaande uit begrippen en inzichten in het vakgebied, die nodig is om kritisch statistische studies te kunnen beoordelen en zelf conclusies uit data te kunnen trekken³. Anderzijds kan gedacht worden aan kennis die primair samenhangt met de statistische problemen die frequent binnen het betreffende vakgebied voorkomen. Ik pleit voor een gedegen basisvak en een pallet van toepassingsgerichte vervolgvakken voor de algemene TU student. Voor de wiskunde masterstudenten wil ik daarnaast nog verdiepingsvakken en vakken met een onderzoeksgelateerde inhoud. Specifiek voor die laatste categorie studenten, zie ik de waarde van het landelijk masterprogramma *mastermath*. Binnen mastermath wordt centraal een breed pakket geavanceerde specialistische cursussen verzorgd op het gebied van de wiskunde, waaronder statistiek. Mijn ervaring in dit samenwerkingsverband is dat studenten eerst wat afhoudend reageren vanwege de reistijd, maar na afloop zeer positief terugzien op de gevolgde vakken en de ontmoeting met studenten van andere universiteiten.

Naast de vraag naar de inhoud van het onderwijs, is er ook de kwestie van de vorm waarin het moet worden aangeboden. Ik hoor nogal eens dat de tijd van de hoorcolleges op haar einde loopt omdat de huidige generatie studenten zich op die manier niet meer zou laten boeien.



Persoonlijk zou ik dat jammer vinden. Een deel van het onderwijs in de wiskunde is kennisoverdracht, praten over intuïtie achter begrippen, in woorden toelichten wat de formules zeggen, motiveren waarom een structuur van een bewijs voor de hand ligt. Daarbij is het voor de docent belangrijk om voeling met het publiek te hebben, de mogelijkheid te hebben studenten 'erbij te trekken'. Een hoorcollege is een efficiënte onderwijsvorm die zich daar heel goed voor leent. Wel is het zo dat in ieder geval bij het vervolgonderwijs in de statistiek deze vorm van kennisoverdracht niet de enige methode van onderwijs moet zijn. Ik zie een begeleidend computerpracticum als een absolute noodzaak. Enerzijds om te leren zelf

³ Prof. ir. J.W. Sieben, hoogleraar mathematische statistiek aan de TH Delft, 1961-1983, stelde getuige zijn oratie groot belang in het deelgebied van de statistiek dat gaat over het opzetten van experimenten.

statistische analyses uit te voeren. Anderzijds om met behulp van stochastische simulatie gevoel te krijgen voor de gebruikte procedures.

Goed onderwijs staat of valt met de docenten die het geven. Een goede docent aan de universiteit is een docent die enthousiast is over zijn vakgebied. Niet alleen over het deel dat grenst aan de frontlinie van onderzoek, maar ook de basis ervan. Zo'n docent moet zich in de kennissituatie van de studenten kunnen inleven en in staat zijn de studenten vanuit die kennissituatie mee te nemen naar voor hun nog onontgonnen terrein. Van studenten mag verwacht worden dat ze zich vervolgens ook actief gaan inzetten om dat terrein persoonlijk te ontginnen. Al doende leert men. Wiskunde is werken, niet het koesteren van een knobbel.

Aan het begin van deze rede heb ik me drie doelen gesteld. In de eerste plaats wilde ik u een beeld geven van de wiskunde dat verschilt van het hardnekkige beeld dat in brede kring leeft. Ik hoop dat u inziet dat wiskunde een creatief, uitdagend, nuttig en (mag ik zeggen vooral) leuk vak is, anders dan een per definitie onbegrijpelijk goochelen met cijfers en letters. Het tweede doel was u te schetsen waar mijn specialisme binnen de wiskunde, de statistiek, zoal over gaat. Daar waar vragen moeten worden beantwoord op basis van gegevens, speelt statistiek een rol en wiskundige modellen vormen de sleutel om vanuit de gegevens tot een antwoord op de gestelde vragen te komen. Het belang van het ontwikkelen en bestuderen van deze wiskundige modellen is dan ook evident. Het derde doel was het uiteenzetten van de ideeën die we hebben over het onderwijs en onderzoek in de statistiek specifiek hier aan de TU Delft. Wij willen vanuit onze expertise graag werken aan wiskundig uitdagende problemen in de statistiek en op die manier meewerken aan het verder uitbreiden van de kennis die er op het gebied van de mathematische statistiek is. Tevens willen we met mensen uit andere disciplines werken aan statistische aspecten van technologisch- en maatschappelijk relevante problemen, aspecten waarbij behoefte is aan statistiek op maat. We willen verder vanuit een degelijke basis motiverend onderwijs aanbieden dat het hele spectrum van toegepast tot theoretisch bestrijkt. Op die manier willen we studenten voorbereiden op werk in het onderzoek (in de statistiek of in een van de vele toepassingsgebieden hiervan) en in het bedrijfsleven, waar ook veel belangrijke beslissingen op basis van gegevens moeten worden genomen.

Tot slot

Het is mij een genoegen vanaf deze plaats enkele woorden van dank uit te spreken. Het college van bestuur van deze universiteit wil ik bedanken voor het in mij gestelde vertrouwen.

Ook wil ik hooggeleerde Groeneboom⁴ bedanken. Piet, in 1989 wist jij me te boeien met het vak mathematische statistiek II. Dat lag niet alleen aan het vak, gegeven op basis van het door jouw voorganger⁵ geschreven diktaat, maar zeker ook aan de manier waarop jij het gaf. Statistiek werd voor mij een onderdeel van de wiskunde waarin begrippen die intuïtief aanspreken formeel worden gedefinieerd en stellingen worden bewezen. Ik mocht in 1991 bij je afstuderen en in 1995 bij je promoveren. Gelukkig is onze samenwerking daarna doorgegaan. Ik bewaar goede herinneringen aan onze samenwerking en hoop dat die, al ben je nu 'in staat van rust' (wat dat bij jou ook moge betekenen), zal blijven bestaan.

Een andere grootheid in de statistiek waarmee ik lange tijd heb mogen samenwerken is hooggeleerde Van der Vaart van de Vrije Universiteit. Aad, de wijze waarop jij je wiskundig inzicht, je snelheid en organisatorische kwaliteiten verenigt, is een voorbeeld voor mij.

Deze gelegenheid wil ik ook te baat nemen om de collega's en studenten te bedanken waarmee ik aan de VU gewerkt heb op o.a. het gebied van onderwijs, onderzoek en internationalisering. De samenwerking is altijd uitermate plezierig geweest.

De vaste staf verbonden aan de leerstoel Mathematische Statistiek hier in Delft, Rik Lopuhaä, Eric Cator en Frank van der Meulen wil ik bedanken voor hun collegialiteit en de prettige werksfeer waarin we samen invulling geven aan onderwijs en onderzoek in ons vak. We delen de visie dat we zowel aan de fundamentele als aan de toegepaste kant van ons vakgebied relevante en uitdagende problemen willen oplossen en ik heb alle vertrouwen in een prettige en vruchtbare voortzetting van onze samenwerking. Tevens wil ik de promovendi bedanken voor de samenwerking en opwekken hun werk met grote inzet voort te zetten.

⁴ Prof. dr. P. Groeneboom, hoogleraar mathematische statistiek aan de TU Delft, 1988-2006.

⁵ Prof. dr. P.J. Rousseeuw, hoogleraar mathematische statistiek aan de TH / TU Delft, 1984-1987.

Ook de wiskundigen binnen het Instituut voor Toegepaste Wiskunde alsmede het ondersteunend personeel wil ik bedanken voor de goede verstandhouding die er onderling is. Die heeft ertoe bijgedragen dat ik me meteen vanaf mijn komst welkom heb gevoeld.

Ik beschouw het als een groot voorrecht dat ik mijn beide ouders bij deze gelegenheid kan bedanken voor alles wat ze voor me hebben betekend en nog steeds betekenen. Mijn vrouw Simone, onze kinderen Henk, Willem, Heidi en Brenda: bedankt voor alles. Ook mijn schoonouders, die hier niet kunnen zijn omdat hun aanwezigheid elders⁶ noodzakelijk is, wil ik bedanken voor hun nauwe betrokkenheid bij ons gezin.

Voor zover u zich nog niet aangesproken voelt in de dankbetuigingen die ik tot nu toe heb uitgesproken, wil ik u allemaal bedanken voor uw aanwezigheid en aandacht.

Ik wil eindigen met een formulering van een dankwoord die velen in deze zaal bekend zal voorkomen. Het is een zin die op de VU door de rector magnificus wordt uitgesproken na afloop van een bijeenkomst als deze. Voor mij persoonlijk betekenen die woorden veel meer dan een slotformule die volgens een protocol moet worden uitgesproken. Deze lofverheffing wijst op de kern, de zin en het doel van het hele bestaan.

De Naam des HEEREN zij geprezen, van nu aan tot in der eeuwigheid.
[Ps. 113:2]

Ik heb gezegd.

Referenties

Balabdaoui, F., Rufibach, K. en Wellner, J.A. (2008). Limit distribution theory for maximum likelihood estimation of a log-concave density. arXiv:0708.3400.

Besselink, M.G. *et al.* (2008). Probiotic prophylaxis in predicted severe acute pancreatitis: a randomized, double-blind, placebo-controlled trial. LANCET **371**, 651-659.

⁶ Op de avond voor het uitspreken van deze rede werden we opgeschrikt door het plotselinge overlijden van mijn zeer gewaardeerde zwager Hans van der Waaij.

- Botero, G. (1589). *Della ragione di stato*. Gioliti, Venetië.
- Chernoff, H. (1964). Estimation of the mode. *ANNALS OF THE INSTITUTE OF STATISTICAL MATHEMATICS* **16**, 31-41.
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *JOURNAL OF THE ANTHROPOLOGICAL INSTITUTE* **15**, 246–263.
- Groeneboom P., Jongbloed, G. en Wellner, J.A. (2001). Estimation of a convex function, characterizations and asymptotic theory. *THE ANNALS OF STATISTICS* **29**, 1653-1698.
- Hemelrijk, J. (1952). *Enkele aspecten der mathematische statistiek*. Intreerede TH Delft, uitgegeven door de VVS, Rotterdam.
- Jongbloed, G. en Van der Meulen, F.H. (2008). Estimating a concave distribution function from data corrupted with additive noise. Te verschijnen in *THE ANNALS OF STATISTICS*.
- Mann, M.E., Bradley, R.S. en Hughes, M.K. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *NATURE* **392**, 779-787.
- Mesina, C.T., Boellaard, R., Van den Heuvel, O.A., Veldman, D.J., Jongbloed, G., Van der Vaart, A.W. and Lammertsma, A.A. (2003). Effects of attenuation correction and reconstruction method on PET activation studies. *NEUROIMAGE* **20**, 898-908.
- Pearson, E.S. (1978). *The history of statistics in the 17th & 18th centuries*. Lectures given by Karl Pearson during the academic sessions 1921-1933. Griffin & Co. Ltd.
- Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *SANKYĀ A* **31**, 23-36.
- Sieben, J.W. (1961). *Experimenten*. Intreerede TH Delft, uitgegeven door de VVS, Rotterdam.
- Vellinga, P. (2007). Eerste bericht van het nationaal onderzoeksprogramma 'Kennis voor Klimaat'.
- Wegman E.J., Scott, D.W. en Said, Y.H. (2006). Ad hoc committee report on the 'hockey stick' global climate reconstruction.