

sector



de week van Tamer Dilaver
Lynxx data inspired by people
complexiteit van optimaliseren René Sitters
tips and tricks scriptieproblematiek
SAS analytics voor de volgende generatie



Voorwoord SECTOR

Het academische jaar loopt alweer ten einde, hopelijk betekent dit dat je alle doelen die je jezelf dit jaar gesteld hebt, hebt behaald, zowel studiegerelateerd als extracurriculair. Om alvast wat inspiratie op te doen voor het komende jaar, zijn in deze editie weer genoeg artikelen te vinden. In de rubriek Extra lees je wat medestudenten het afgelopen jaar naast hun studie gedaan hebben, en in 'de week van' zien we hoe een werkweek van Tamer Dilaver eruit ziet bij AlInvest, waar hij aan de slag is gegaan na het behalen van zijn master.

Voor iedereen die binnenkort begint aan een bachelor- of masterscriptie is de Tips & Tricks geschreven door Ad Ridder een aanrader, hierin vind je handige tips met betrekking tot het schrijven van je scriptie.

In de academische rubriek geeft René Sitters ons een kijkje in de geschiedenis van de optimalisatietechnieken. Sinds zijn periode als PhD-student, waarin hij enkele grondleggers van het vakgebied tegen het lijf is gelopen, is er namelijk

veel vooruitgang geboekt op het gebied van de Operations Research. Ook hebben we weer een bijdrage van Rein Nobel, waarin hij nog eens op kenmerkende wijze zijn ideeën over de kansrekening uiteenzet.

Verder hebben we met twee bedrijven gesproken. Op landgoed Oud-Bussem spraken we met Erwin Huizenga van SAS. SAS is specialist in business analytics software voor klanten wereldwijd. We spreken hem onder andere over het belang van innovatie voor softwareleveranciers. We openen met een interview met Frank Beunder en Peter Nijhuis van Lynxx. Lynxx is een jong en vooruitstrevend adviesbureau op het gebied van data-analyse en het openbaar vervoer en is erg actief in het buitenland. Ze hebben sinds kort een vestiging geopend in Australië, waar zij ons meer over zullen vertellen.

We hopen jullie hiermee van genoeg inspiratie te voorzien om het komende jaar ook weer succesvol te kunnen maken. Veel leesplezier toegewenst!

inhoud

- 1 **voorwoord**
- 3 **bedrijfsinterview Lynxx**
Frank Beunder en Peter Nijhuis
- 8 **column Optimaal Stoppen**
Prof. Henk Tijms
- 11 **academische rubriek**
René Sitters
- 16 **de week van**
Tamer Dilaver van AlInvest
- 19 **bedrijfsinterview SAS**
Erwin Huizenga
- 25 **column Een paradoxaal resultaat in een serie muntworpen?**
Rein Nobel
- 32 **extra**
vier studenten, vier extracurriculaire activiteiten
- 36 **docent vertelt**
Jan Magnus
- 40 **tips and tricks**
scriptieproblematiek door Ad Ridder
- 43 **puzzeltijd**
- 44 **mediarecensies**
- 45 **agenda en publicaties**

colofon

sector
Jaargang 7, nr 2, juni 2017

SECTOR is een uitgave van studievereniging Kraket (Kritische Aktuarissen en Econometristen), de vereniging voor studenten Econometrie & Operationele Research aan de Vrije Universiteit Amsterdam.

redactie:
Maik Dekker
Carlos Groen
Aïcha el Hri
Nienke van 't Klooster
Xerxes Koehoorn

redactieadres:
Studievereniging Kraket - SECTOR
De Boelelaan 1105
Kamer 8A-30
1081 HV Amsterdam
Tel. 020 598 60 15
E-mail: sector@kraket.nl

advertenties:
Informatie over adverteren in de SECTOR kan worden aangevraagd op het bovenstaande redactieadres.

adverteerders:
Aegon
Belastingdienst
Finaps
Lynxx
PwC
SAS

ontwerp:
gedesign
www.gedesign.nl

opmaak / art direction:
gedesign
www.gedesign.nl
info@gedesign.nl

drukwerk:
Flyeralarm

**Frank Beunder**

sloot zich vlak na de oprichting van Lynxx aan als partner van het bedrijf.

**Peter Nijhuis**

studeerde werktuigbouwkunde aan de TU Eindhoven en werkt sinds 2014 als consultant bij Lynxx.

Lynxx Data inspired by people

Als specialist op het gebied van data analyse is Lynxx de afgelopen jaren flink gegroeid. Zowel qua team als qua vestigingen. Sinds kort heeft een nieuwe vestiging namelijk haar deuren geopend in Australië. Onder andere hierover spreken wij Frank Beunder en Peter Nijhuis op het kantoor in Amersfoort.

Hoe zijn jullie hier terecht gekomen?

Peter: Ik heb geen standaard achtergrond als je het vergelijkt met de anderen die hier zitten. Ik heb namelijk werktuigbouwkunde gestudeerd. Met mijn afstuderen ben ik vervolgens met modellen aan het werk gegaan en zo ben ik in contact gekomen met Lynxx. Na een paar bakjes koffie werd er gevraagd of ik eens langs wilde komen. Ik dacht: 'waarom niet?'. Bovendien was het een goede kans om m'n netwerk uit te breiden. Toen ik hier binnenkwam, was ik aangenaam verrast door de mensen en de gezelligheid. Na een aantal mensen gesproken te hebben, begon ik de raakvlakken te zien en vond ik mezelf hier beter passen dan ik in eerste instantie had verwacht. De persoonlijkheid van het kleine team spreekt mij heel erg aan en dat is eigenlijk de reden dat ik hier aan de slag ben gegaan.

Frank: Mijn achtergrond is in tegenstelling tot

de rest die hier werkt niet universitair. Ik heb hbo-bedrijfskunde en informatica gedaan. Mijn achtergrond is in eerste instantie werkplekbeheer van grote bedrijven zoals Philips. Daar ben ik dan ook gestart, waar ik ben doorgesloopt tot consultant. Als consultant ontwierp ik pakketten waar vooral de bazen en managers erg blij mee waren, doordat ze met behulp van deze pakketten alles konden overzien. De medewerkers waren hier echter minder gelukkig mee, omdat ze in hun bewegingsvrijheid beperkt werden. Hierdoor werd het voor mij lastig als consultant om te achterhalen of alles nou liep zoals het zou moeten. Toen ben ik op zoek gegaan naar oplossingen die kunnen helpen met het analyseren en visualiseren van alle data die in je systeem zit. Dat is voor mij het moment geweest dat ik voor het eerst met data aan de slag ben gegaan. Plots had ik iets in de markt gevonden wat iets opleverde. Iets waar mensen gelukkig van

werden. Dat is voor mij de reden geweest om ermee verder te gaan. Vervolgens heb ik Paul en Sanneke (de twee oprichters van Lynxx) ontmoet en samen zijn we Lynxx begonnen.

Kunnen jullie wat meer over de sfeer bij Lynxx vertellen?

We zijn een relatief platte organisatie. We doen veel leuke dingen samen, maar natuurlijk wordt er ook hard gewerkt. Als je start krijg je meteen veel verantwoordelijkheid en dan moet je het ook maar opknappen. Daar leer je tenslotte het meest van. Dat kenmerkt Lynxx voor mij het meest. Wat ik verder heel leuk vind, is dat het bij ons nog weleens voorkomt dat je met z'n tweeën aan een klus werkt. Dan ben je continu met elkaar samen. Op zo'n moment is het zowel voor het werk, maar ook voor jezelf fijn als je het met elkaar kunt vinden. Er moet een bepaalde klik zijn als je met iemand samenwerkt, dat is heel belangrijk. We zijn daarom ook op zoek naar collega's die niet alleen slim, maar ook sociaal en open zijn. Tijdens sollicitaties of oriëntaties van nieuwe mensen denk ik vaak: 'Goh, de persoon die nu staat te presenteren, zou ik daar 's avonds ook een biertje mee willen doen in de kroeg?'

Wat doet Lynxx precies?

We hebben een aantal pijlers opgezet. Aan de ene kant is het dashboarding. Dit houdt in dat we inzicht creëren in wat de bedrijven allemaal aan data hebben. Dit betreft de baselere en eenvoudigere informatie. De kracht zit hem hier heel erg in het presenteren van de data. De informatie is niet moeilijk, maar de uitdaging zit hem hier dus in de creativiteit. Aan de andere kant maken we modellen: tender werk voor aanbestedingen. Hier ga je meer de diepte in en maak je hardcore analyses of voorspellingen. Daarnaast hebben we een stukje machine learning waar we geavanceerde modellen proberen toe te passen op de datasets van de klanten. Onze grootste pijler is het openbaar vervoer, maar we proberen ook andere takken van de markt binnen te dringen, zoals transport- of logistieke organisaties. Dit past bij ons, omdat we ervaren zijn met het openbaar vervoer. Doordat de overgang niet zo groot is, geeft dat de eventueel aanstaande klant meer vertrouwen en bovendien heb je zelf meer verstand van zaken. Verder grenzen de problemen aan elkaar.

Werken jullie met specialisten binnen deze pijlers?

Nee, iedereen komt hier redelijk allround binnen. Formeel hebben we ook geen omschrijving van waar je het beste in bent. Informeel merk je wel dat de één ergens beter in is dan de ander. De ene persoon vindt het leuker om heel dicht bij de klant te zitten en de andere doet bijvoorbeeld liever een project in het buitenland. Op het moment dat er een project wordt binnengehaald, kijken we in eerste instantie wie daar het best bij past, maar uiteindelijk doet iedereen wel alle soorten opdrachten. Daar zit ook onze kracht. We zijn klein en willen flexibel zijn, dus we proberen iedereen zo breed mogelijk in te zetten. Het zal niet zo snel voorkomen dat er voor twee personen géén klus is, terwijl iemand anders voor drie moet werken. Vaak is dat ook voor jezelf leuker. Het kan leuk zijn om je een paar maanden helemaal in een bepaalde richting te verdiepen, maar vervolgens is het ook heel fijn als je weer wat anders kan doen. Een beetje verandering houdt je gelukkig.

“Daar zit ook onze kracht. We zijn klein en willen flexibel zijn.”

Hoe ziet het eruit als je hier binnenkomt?

Aan het begin heb je natuurlijk nog niet echt een idee wat er nou van je verwacht wordt en hoe het kantoor werkt, maar het is hier heerlijk chaotisch. Er wordt gekeken wat er geleerd moet worden en welke taken je al meegegeven kunnen worden. Het ligt er dan natuurlijk aan wat er vrij is, maar je wordt vrij snel in het diepe gegooid. Zelf heb ik eerst twee weken een cursus gedaan. Vervolgens ben ik met een collega meegegaan naar de klant en die werd binnen drie weken overgedragen. Vanaf dat moment was ik verantwoordelijk voor die klant. Zo snel kan het gaan. Natuurlijk wordt er meegekeken of je het daadwerkelijk redt, maar je krijgt gelijk veel verantwoordelijkheid mee. Ze laten je als het ware zwemmen en als je dreigt te verzuipen, word je eruit geholpen. Het is een goede manier om heel snel stappen te maken. Uiteraard kan het ook zijn dat je samen met iemand op een project wordt gezet. Dat

is als nieuweling toch iets veiliger, omdat je dan wat meer vragen kan stellen.

Als je bij een klant zit, is het dan duidelijk wat er van je verwacht wordt?

Het is heel belangrijk om goed te achterhalen wat de échte wens van de klant is. Soms denken ze een idee te hebben van wat ze graag willen, maar als je dan de juiste vragen stelt, komen ze erachter dat ze het probleem heel anders kunnen benaderen. Onze rol als consultant is dan dat wij de klant meenemen naar de plek waar hij eigenlijk naartoe wil.

De eerste opdrachtbeschrijving die je van de klant krijgt, is haast nooit wat je uiteindelijk zult gaan uitvoeren. Zeker met data zie je dit vaak. Het is aan ons om er wat van te maken.

Wordt dit verder helpen op prijs gesteld?

Het is aan je eigen kennis en kunde om zo'n opdracht veel verder te helpen dan dat hij initieel werd ingestoken. Soms zijn klanten ervan overtuigd dat ze weten wat er met de data moet gebeuren. Ze komen dan naar ons toe en zeggen: "Dit is exact wat wij verwachten." Uit ervaring vertellen wij ze dan dat er veel meer mogelijk is dan zij zelf denken. Je gaat dan in eerste instantie met ze meebewegen om ze ervan te overtuigen dat we kunnen wat ze vragen, maar daarna pak je natuurlijk door. Dat wordt zeker gewaardeerd.

Zitten jullie veel in het buitenland?

Jazeker, afgelopen maand hebben we zelfs het eerste buitenlandse kantoor geopend in Australië: The Australian Office. Verder hebben we ook in Zweden en Denemarken projecten gehad, maar Engeland is toch wel de hoofdmoot waar we de grotere klanten hebben. Verder zijn we bijvoorbeeld aan het bieden op een klus in Buenos Aires en in Boston een plek die ons heel aantrekkelijk lijkt, dus ook daar willen we graag naartoe. Onze blik staat absoluut op het buitenland gericht.

Peter: Zelf was ik nog geen jaar in dienst en toen kwamen ze naar me toe met "we hebben een toffe klus in Stockholm voor je". Dat greep ik met beide handen aan, dus reisde ik in de volgende maanden een aantal keer op en neer. Dat is natuurlijk een geweldige ervaring. Je wordt ineens in een team gezet met een hele andere context en een hele andere markt, dus je moet snel je draai zien te

vinden. Je leert hoe je het beste wat kunt bijdragen en natuurlijk hoe zij daar dingen zien, want je hebt echt met cultuurverschillen te maken. Het is een omgeving waar je heel snel, heel veel moet leren en dat is heel gaaf om te kunnen doen.

“Het lijkt misschien alsof zoiets als data science heel technisch is, maar er komt ook veel creativiteit bij kijken.”

Zijn de cultuurverschillen binnen Europa groot?

In Engeland is het, in tegenstelling tot Nederland, niet gebruikelijk om flexibel te werken. Er wordt verwacht dat je op kantoor bent binnen gezette tijden. Je moet gewoon aanwezig zijn en je uren draaien. Of dat nu efficiënt is, is natuurlijk moeilijk te checken, maar ze gaan ervan uit dat als je er lang genoeg zit, je goed hebt gewerkt die dag. Het zit daar in de cultuur dat je eerder dan je baas op het werk bent en later dan je baas vertrekt. Een ander groot verschil is dat Nederlanders vrij direct zijn. In Engeland kan het zo zijn dat je een presentatie hebt gegeven en vervolgens een reactie krijgt als: "Hmm that's interesting". Engelsen zullen, in tegenstelling tot Nederlanders, nooit zeggen dat iets verschrikkelijk (goed) is. Er wordt altijd gematigd gereageerd. In Nederland zijn we niet alleen goed in direct zijn naar onze collega's, maar ook naar onze bazen. Het mooie is dat de cultuur die wij, zij het met mate, meenemen naar het buitenland verfrissend werkt. Doordat Nederlanders zo direct zijn, zijn we veel sneller in het benoemen van wat er niet goed gaat. Hiermee doorbreek je een patroon, wat uiteindelijk gewaardeerd wordt. Een van de projecten die we in Engeland doen, heeft te maken met het openbaar vervoer. Vervoersmaatschappijen mogen dan bieden op een traject voor de komende vijf of tien jaar. Op dat soort projecten zit heel veel druk. De directheid die wij meenemen, werkt dan voordelig, omdat dit het gehele proces een stuk sneller laat gaan.

Even terugkomend op The Australian Office. Hoe gaat het opzetten hiervan in z'n werk?

We hebben dus een kantoor geopend in Australië en een aantal van ons gaat helpen bij het opstarten hiervan. We zijn daar in eerste instantie druk bezig met het binnenhalen van opdrachten. Vervolgens moeten we daar mensen werven, zodat zij hiermee aan de slag kunnen gaan. Het is de bedoeling dat er een aantal van ons voor korte perioden van drie à vier maanden die kant op gaat om daar de boel op te zetten. Na een aantal van deze perioden zullen ze zelfstandig gaan opereren in Australië.

Bevalt de huidige grootte van het team?

Het is nooit onze primaire ambitie geweest om heel groot te worden. We merken allemaal dat een kleine organisatie die lekker knus is en waarin iedereen elkaar kent, het beste werkt. Daar wordt enorm veel plezier uitgehaald. De drive is dan ook niet om extreem groot te worden. We willen een leuk team hebben, leuke dingen met elkaar kunnen doen en leuke opdrachten binnenhalen. Dankzij het kleine team is er een relaxte werkomgeving. Het is natuurlijk ook belangrijk dat er ontspanning is tijdens het werk. Als je de hele dag je hoofd aan het breken bent over een lastig model waar je even niet uitkomt, is het fijn om er even uit te kunnen en dat er andere dingen om je heen gebeuren. Dat brengt ook wel een beetje creativiteit bij jezelf naar boven. Creativiteit die we toch vaak bij onze opdrachten nodig hebben. Het lijkt misschien alsof zoiets als data science enkel heel technisch is, maar er komt ook veel creativiteit bij kijken.

Met de opening van The Australian Office worden jullie wel in een klap groter. Is dat dan niet tegenstrijdig?

Dat klopt, er zullen dan meer mensen zijn, maar tegelijkertijd zullen we in Nederland nooit een team worden van bijvoorbeeld 100 man. We zullen verspreid over de wereld zitten, maar met kleine teams die allemaal zelfstandig werken. We kunnen elkaar helpen, maar het is niet zo dat we heel veel met elkaar te maken zullen hebben. Zo helpen we Australië nu met opstarten, maar zodra daar drie à vier man personeel is, zullen ze het vooral zelf moeten doen. Het is de bedoeling dat daar weer een klein team is die op dezelfde manier in elkaar zit. Uiteindelijk zijn het allemaal zelfstandig opererende

onderdelen. Overkoepelend zou Lynxx dus nog wel groot kunnen worden, maar we verwachten niet dat we per locatie tien kantoren zullen openen.

3 MILLION BIG DATA JOBS

... will need to be filled in the US by 2018.

THE SKILLS GAP

40%
projected growth
in global data
generated each year.

85% +
of Fortune 500 organizations
will be unable to effectively
exploit big data by 2015.



CAREER PERKS

**EARN BIGGER
PAYCHECKS**

17%
Boost your
annual earnings
by 17 percent.

**BE THE ENVY
OF MANY**

15026
948
Data scientist
was labeled the
sexiest job of the
21st century.

**DO WHAT
YOU LOVE**

Organizations in
every industry
need analytics -
you get to decide.

THE SAS® FACTOR



Many paths.
One brilliant journey.

SAS skills are Monster.com's
#1 path to
higher salaries.



SAS Certifications identified
as one of the "Best Big Data
Certifications for 2014."



Learn more:
sas.com/analyticsu



Column Optimaal Stoppen

Elke beslissing is een afweging van risico factoren. Wanneer het weer plotseling verslechtert, moet de meteorologische dienst een beslissing nemen over wanneer het tijd is om te stoppen met het afwachten van hoe het weer zich verder ontwikkelt en het moment gekomen is om een weerwaarschuwing te laten uitgaan. Wanneer Apple een nieuwe versie van de iPhone in voorbereiding heeft, moet het beslissen wanneer het uittesten van het product gestopt moet worden en het juiste moment aangebroken is om het op de markt te brengen. Dit zijn voorbeelden van beslissingen die genomen moeten worden in een omgeving vol van onzekerheden en die grote gevolgen kunnen hebben wanneer de timing slecht is. Niet alleen grote bedrijven hebben te maken met dit soort beslissingen, maar ook in het dagelijks leven krijgt eenieder te maken met soortgelijke beslissingen zoals de bepaling van het juiste moment om een huis te kopen of om al dan niet een aanbod voor een nieuwe baan te accepteren.

In de kansrekening staan beslissingsproblemen met als vraagstelling "doorgaan of stoppen" bekend als optimale stopproblemen. Het meest beroemde optimale stopprobleem is het dating probleem. Voor kranten en tijdschriften een dankbaar onderwerp dat om de zoveel tijd weer volop aandacht krijgt. Stel dat je op een dating site de profielen van een aantal niet-gelijkwaardige potentiële partners één voor één in een willekeurige volgorde te zien krijgt. Je mag met slechts één van hen een afspraak maken. Na een profiel van een kandidaat gezien te hebben, moet je meteen beslissen een afspraak te maken of niet. Op een eerdere afwijzing kun je niet terugkomen. Als je een kandidaat te zien krijgt, kun je alleen maar oordelen of deze kandidaat beter of minder goed is dan de voorgaande kandidaten die je gezien hebt. Anders gezegd, je kunt aan de kandidaten alleen

relatieve waarderingen toekennen. Welke strategie geeft je in deze situatie de grootste kans de beste te vinden? De wiskunde geeft antwoord op deze vraag. Wanneer het aantal kandidaten voldoende groot is, zeg 10 of meer, dan is de optimale beslisseregels dat je ongeveer 37% van de kandidaten voorbij moeten laten gaan om daarna de eerste te kiezen die beter is dan alle vorige kandidaten. Je hebt dan een kans van ongeveer 37% om de beste te vinden. Velen vinden het verrassend dat deze kans zo groot is en vrijwel niet afhangt van het aantal kandidaten. Zo verrassend is dit ook weer niet wanneer je bedenkt dat je al met een kans groter dan 25% de beste vindt wanneer je de eerste helft van de kandidaten laat passeren en daarna de eerste neemt die beter is dan elk van de anderen die je daarvoor gezien hebt. Onder deze simpele strategie vind je in elk geval de beste wanneer de op één na beste in de eerste helft zit en de beste in de tweede helft. De kans op deze volgorde is $10/20 \times 10/19 \approx 0.263$. De ondergrens van 25% vind je met dezelfde redenering ook als je 1 miljoen kandidaten hebt! Een aardige variant van het dating probleem is het Googol spel voor twee spelers. Je laat iemand op een aantal papiertjes verschillende positieve getallen schrijven, waarbij geen beperking opgelegd is aan de grootte van de getallen en de getallen mogen lopen van een miniem klein getal tot een getal van de grootte van een googol (een 1 gevolgd door honderd 0'en) of nog groter. De papiertjes worden grondig door elkaar geschud, omgekeerd op tafel gelegd en daarna één voor één omgedraaid. Velen zullen de kans dat je het grootste getal zal vinden lager inschatten dan deze kans in werkelijkheid is en daarmee zou je een voor jou gunstige weddenschap kunnen afsluiten.

Verschiedende uitbreidingen van het dating probleem zijn onderzocht zoals de bepaling van de strategie

Sources: Big data: The next frontier for innovation, competition, and productivity, McKinsey & Company, May 2011; Gartner Reveals Top Predictions for IT Organizations and Users for 2012 and Beyond, Gartner Research, December 2011; Job Skills That Lead to Bigger Paychecks, Monster; Data Scientist: The Sexiest Job of the 21st Century, Harvard Business Review, October 2012; Best Big Data Certifications for 2014, Tom's IT PRO, March 2014.

om met maximale kans één van de twee beste te vinden. De maximale kans om te slagen stijgt dan tot ongeveer 57% en bij één van de drie beste tot ongeveer 71%, waarbij de optimale strategie door meerdere omschakelpunten wordt bepaald. Iemand die daar onderzoek naar gedaan heeft, was de Russische wiskundige Boris Berezovski die ook een gewiekst zakenman was. In de chaotische tijd in Rusland na de val van de Berlijnse muur vergaarde hij een fortuin en werd de rijkste man van Rusland. Berezovski kon doorgaan met zijn schimmige praktijken totdat hij publiekelijk stelling nam tegen Poetins plannen om de constitutionele macht van de president te vergroten. Deze kritiek leidde ertoe dat Berezovski in 2000 hals over kop naar Engeland moest vluchten waar hij harde kritiek bleef uiten op zijn voormalige protegé Vladimir Poetin. In maart 2013 werd Berezovski onder verdachte omstandigheden dood aangetroffen op zijn landgoed in het Engelse Ascot en tot op de dag van vandaag is de oorzaak van zijn dood onopgehelderd. De wiskundige Berezovski had toch als expert in de theorie van optimaal stoppen moeten weten wanneer te stoppen met zijn kritiek.

Een geheel andere situatie ontstaat wanneer voor elke kandidaat de waardering kan worden uitgedrukt in een getal op een schaal van 0 tot 1. Laten we deze situatie in de volgende context beschouwen. Stel eens dat iemand met een random-generator een van te voren vastgelegd aantal van R getallen tussen 0 en 1 genereert. Deze persoon laat je de getallen één voor één zien en wel in een random volgorde. Je wilt met maximale kans het grootste getal identificeren. Elke keer dat je een getal te zien krijgt, moet je meteen beslissen of je dat getal als grootste getal aanmerkt en dan stopt of dat je het getal laat passeren en vraagt een volgende getal te laten zien. Je kunt niet op een eerdere beslissing terugkomen. Welke strategie moet je volgen? Een optimale strategie is vrij lastig te berekenen, maar je kunt een simpele beslisregel geven die vrijwel optimaal is. Stel dat je een getal, zeg getal a , te zien krijgt terwijl nog k getallen getoond kunnen worden en dit getal a is het grootste getal dat je tot dan gezien hebt. De simpele beslisregel zegt dan dit getal a te nemen alleen dan als $a^k \leq 0.5$, oftewel als de kans dat onder de k resterende getallen een nog groter getal zit niet meer dan 0.5 is. Onder deze heuristiek is de kans

dat je het grootste getal vindt 60.5% als $R = 10$ en 58.9% als $R = 25$, zoals met simulatie kan worden nagegaan. De kans convergeert naar ongeveer 58% als R toeneemt.

Het dating probleem heeft vanzelfsprekend ook de nodige aandacht getrokken van cognitieve psychologen. Zo kwam voor het dating probleem waarin alleen relatieve waarderingen aan de kandidaten kunnen worden toegekend, de Duitse psycholoog Peter Todd met het “magische” getal 12: bij honderd of meer kandidaten, laat de eerste 12 passeren en neem dan de eerste die beter is dan alle voorgaande wanneer je doel is om met maximale kans een partner uit de top 10% te vinden. Voor deze vuistregel hoeft je niet het precieze aantal kandidaten te weten. Voor het dating probleem waarin absolute waarderingen aan de kandidaten kunnen worden toegekend, kwam onlangs de zogenaamde wortelregel volop in het nieuws: bij R kandidaten, laat de eerste $\sqrt{R} - 1$ passeren en neem dan de eerste die beter is dan alle voorgaande. Niet bepaald een regel die aan te bevelen valt wanneer je met maximale kans de allerbeste partner wilt vinden: voor $R = 25$ bijvoorbeeld geeft deze vuistregel een kans van ongeveer 31% om de allerbeste partner te vinden, terwijl de maximale kans ongeveer 59% is. Zou binnenkort een dating bureau met de naam “Square-Root Dating” op de markt verschijnen, houdt u er verre van!

Henk Tijms is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Email: h.c.tijms@xs4all.nl



www.finaps.nl

We're Hiring!

Want to know more about
our company and what we do?

Just stop by for a cup of coffee!

| Business solutions that work |

Over de complexiteit van optimaliseren en de efficiëntie van algoritmes.

Kennis van optimalisatietechnieken is essentieel voor de aanpak van problemen op allerlei gebieden zoals de logistiek, economie, informatica, maar bijvoorbeeld ook de biologie. Mede door een sterk toenemend aanbod van data is de vraag naar efficiënte algoritmes groter dan ooit. In deze rubriek geeft René Sitters een blik op de geschiedenis van de complexiteit van optimalisatie.



René Sitters

is universitair docent aan de afdeling Econometrie en Operations Research aan de Vrije Universiteit.

In de zomer van 2000, ik was nog maar een paar maanden PhD student aan de TU Eindhoven, bezocht ik voor het eerst een groot congres, het 17^{de} International Symposium on Mathematical Programming (ISMP) in Atlanta. Een van de aanwezigen was George Dantzig, toen 85 jaar, wiens naam ik kende als grondlegger van de simplex methode voor lineaire programmeringsproblemen [1]. Het was pas toen dat ik mij realiseerde hoe relatief jong het vakgebied van de Operations Research en met name de Mathematical Optimization is. Veel van de grote namen die ik kon noemen waren hier aanwezig. We verbleven met zes Eindhovense PhD studenten in een dormitory van Georgia Tech. In de avond voetballen we tegen de Duitse PhD studenten en een keer hadden we 's nachts lol met het schijnen met een laserpen (toen nog iets nieuws) voor de voeten van voorbijgangers. Heel irritant natuurlijk en we zijn maar gestopt toen een grote, wat oudere man, wel erg te keer ging richting ons. De volgende dag bleek die persoon Jack Edmonds te zijn, een van de bekende namen en die ik nu ook van gezicht kende.

Edmonds is vooral bekend van zijn werk over het matchingprobleem [3]. Matchings had ik bij mijn studie geleerd in een vak over grafentheorie: Een matching M in een graaf is een deelverzameling van de lijnen zodat elk punt in hooguit één van de lijnen in M voorkomt. Het matchingprobleem is

fundamenteel in de Operations Research, het gaat om het koppelen van vliegtuigen aan gates, van personen aan taken, van patiënten aan bedden, scholieren aan scholen et cetera. Edmonds gaf een algoritme voor het vinden van een grootste matching in een willekeurige graaf en toonde aan dat de running time $O(n^4)$ was (met n het aantal punten). Dit was daarmee het eerst bekende algoritme voor matching met een running time beter dan gewoon alle mogelijkheden nalopen. Natuurlijk, n^4 is al snel te traag voor praktische toepassingen maar het vernieuwende was de nadruk op de efficiëntie. Het algoritme was bewijsbaar efficiënt in de zin van sneller dan gewoon alle mogelijkheden proberen. Tot die tijd hield men zich in de grafentheorie voornamelijk bezig met karakterisaties zoals bijvoorbeeld Berge's Theorem [4].

Berge's theorem (1957):

Een matching M in een graaf $G = (V, E)$ is een grootste matching d.e.s.d. als er geen augmenting path bestaat. (Een augmenting path is een pad met afwisselend lijnen in M en lijnen in $E \setminus M$ beginnend en eindigend met een lijn in $E \setminus M$.)

Het bewijs van Berge's stelling is vrij eenvoudig en het reduceert het vinden van een grootste matching tot het vinden van een augmenting path. Echter, ook hiervoor was het niet bekend hoe dat kon anders dan feitelijk alle paden te proberen. Een

andere karakterisatie die ik ooit had geleerd bij het grafentheorie vak is die van Tutte [5].

Tutte's theorem (1947):

Een graaf $G = (V, E)$ heeft een perfecte matching d.e.s.d. als voor elke $X \subseteq V$ het aantal oneven componenten in $G - X$ niet meer is dan $|X|$. (Een matching heet perfect als het alle punten bevat.)

Op het eerste gezicht is dit gewoon een lastige stelling uit de grafentheorie, maar Tutte's stelling heeft een interessante implicatie: Als een graaf een perfecte matching heeft dan is dat uiteraard aan te tonen (te bewijzen) door zo'n matching te geven. Bijzonder is dat het omgekeerde ook geldt: Als er geen perfecte matching bestaat dan is dat aan te tonen door het geven van een $X \subseteq V$ waarvoor het aantal oneven componenten in $G - X$ meer is dan $|X|$. Edmonds vroeg zich af of een dergelijke karakterisatie die beide kanten op werkt het bestaan van een efficiënt algoritme impliceert. Tegenwoordig kennen we dit als de vraag: Is P gelijk aan $NP \cap co-NP$? Deze theorie zou pas tien jaar later gevormd worden en het antwoord op deze vraag is tot op heden niet bekend, met andere woorden, er is geen voorbeeld bekend waarvoor deze implicatie niet geldt.

Een andere grote naam en spreker op het ISMP in Atlanta was Ralph Gomory die bij velen bekend is van de Gomory cuts [6] voor Integer Linear Programming (ILP) problemen. De simplex methode van Dantzig is tot op de dag van vandaag een uiterst efficiënte manier voor het oplossen van lineaire programmeringsproblemen. Gomory realiseerde zich dat veel problemen om een geheeltallige (integer) oplossingen vragen. Het is immers weinig zinvol om een vliegtuig voor de helft aan gate D21 toe te wijzen en voor andere helft aan gate D22. Gomory liet zien hoe het Simplex algoritme voor LPs gebruikt kan worden voor het oplossen van een willekeurig ILP door het iteratief toevoegen van een eindig aantal ongelijkheden aan het LP. Dit was een enorme doorbraak, omdat hiermee elk ILP op te lossen is door een eindig aantal maal Dantzig's simplex algoritme toe te passen. In de praktijk bleek de running time echter erg hoog te zijn en de Gomory cuts werden als onpraktisch gezien. Rond het jaar 2000 werden deze echter weer herontdekt en verwerkt in commerciële optimization software

als Cplex en tegenwoordig Gurobi en open source solvers als CBC. Dit bleek een ongekennde speedup te geven. De Gomory cuts kwamen pas goed tot hun recht in de inmiddels veel grotere ILP problemen.

Een presentatie op ISMP waar ik naar uitzag was die van David Johnson die helaas in 2016 is overleden. Er waren maar liefst 19 parallele sessies en men had Johnson in een veel te kleine zaal ingedeeld (geen perfecte matching dus). In een overvolle zaal sprak hij over de resultaten van de DIMACS TSP-challenge [7] die toen liep. Ik kende zijn naam van het boek Computers and Intractability [8] (1979) over de complexiteitstheorie welke zo'n tien jaar eerder tot stand kwam. Dit boek gaf de state of the art en de auteurs stelden een lijst op met twaalf fundamentele problemen welke nog niet geclassificeerd waren als zijnde makkelijk op te lossen (in P) of moeilijk op te lossen (NP-hard). Inmiddels is dit bekend voor tien problemen op de lijst (5 in P en 5 NP-hard) en staat het nog open voor twee problemen. Interessant op deze lijst van twaalf is bijvoorbeeld het volgende probleem: Gegeven een getal N , is N een priemgetal? Pas in 2002 werd bewezen dat dit efficiënt te bepalen is: 'Primes is in P' [9]. Het meest interessante probleem op de lijst is misschien wel Linear Programming. Het was bekend dat het simplex algoritme goed werkt in de praktijk maar dat de rekentijd in extreme gevallen zeer groot kan zijn. Kort na de publicatie van het boek liet de Russische wiskundige Khachiyan [10] zien dat LPs in theorie efficiënt (in polynomiale tijd) op te lossen zijn met wat nu bekend is als de ellipsoïde methode, welke overigens in de praktijk hopeloos traag is. Simplex rules dus. De nog twee open problemen zijn: Grafenisomorfie (Is graaf 1 gelijk aan graaf 2?) en het volgende elementaire scheduling probleem. Er zijn drie machines en n jobs die elk een lengte van 1 tijdseenheid hebben. Er is een restrictie op de volgorde van jobs welke gegeven is in de vorm van een (acyclische) gerichte graaf $G = (V, A)$. Om precies te zijn, V is de verzameling jobs en als $(j, k) \in A$ dan mag job k niet eerder beginnen dan het moment waarop job j af is. Elke job moet aan een positie (i, t) worden toegewezen wat betekent dat job j start op tijd $t \in \{0, 1, \dots\}$ op machine i . Doel is het minimaliseren van de lengte van het schedule, dwz., het tijdstip waarop alle jobs klaar zijn. Dit is in feite een matching probleem (van jobs aan posities (i, t))

met volgorde restricties. Tot op heden is er geen efficiënt algoritme bekend, noch is er een bewijs dat het NP-hard is. Volgorde restricties komen in de praktijk veel voor. Om bij het eerste voorbeeld te blijven, een vliegtuig wordt niet alleen aan een gate toegewezen (bijv. D21) maar ook aan een tijdslot (bijv. van 13.30-14.00) en daar omheen spelen allerlei processen (inchecken, boarding, tanken, bagage inladen, et cetera) welke niet in een willekeurige volgorde gedaan kunnen worden. Volgorde restricties zijn vaak essentieel maar bemoeilijken het optimaliseringsprobleem.

Inzicht in de complexiteit van optimaliseringsproblemen is fundamenteel voor het geven van efficiënte algoritmes. In 2000 selecteerde het Clay Mathematics Institute of Cambridge zeven Millennium Prize Problems [11] en reserveerde voor elk probleem 1 miljoen dollar voor een oplossing. Inmiddels is er één probleem opgelost en staat het derde probleem op de lijst, P versus NP, nog wijd open. Kort gezegd komt het er hierbij op neer om aan te tonen dat sommige problemen lastiger op te lossen zijn dan anderen. In de praktijk blijken LPs veel eenvoudiger dan ILPs en de theorie zegt dat een oplossing voor een LP in polynomiale tijd gevonden kan worden. Echter, er is geen bewijs dat zegt dat ILPs *niet* in polynomiale tijd zijn op te lossen. Het vermoeden is er maar een bewijs (ter waarde van 1 miljoen dollar) ontbreekt. Prof. Gerhard Woeginger houdt een webpagina [12] bij met een hele lijst van foutieve bewijzen voor $P \neq NP$ of juist $P = NP$. Sommige auteurs zijn klaarblijkelijk onwetend over de verstrekkende implicatie van het geclaimde resultaat.

Efficiëntie staat in de praktijk niet gelijk aan polynomiale running time. Dat gold al niet voor Edmonds $O(n^4)$ matching algoritme en veel nieuwe toepassingen vragen vandaag de dag om een nieuwe aanpak zoals sub-linear time algorithms welke een running time hebben die zelfs lager ligt dan wat nodig is om de hele dataset te lezen. Deze lage running time kan bijvoorbeeld verkregen worden door samples te nemen van de input. Ook sub-linear *space* algoritmes zijn meer en meer interessant. Traditioneel wordt de input gezien als iets wat je in willekeurige volgorde kunt lezen. Dit is niet altijd haalbaar voor moderne problemen en

het streaming model is bedoeld om dit op te vangen. In deze setting spelen samples een cruciale rol. Neem bijvoorbeeld het volgende simpele probleem. Zijn alle n getallen in een gegeven (kolom van) een dataset verschillend? Om dit exact te beantwoorden is lineaire tijd en ruimte nodig. Echter, indien het volstaat om een onderscheid te maken tussen 'alle n getallen zijn verschillend' en 'niet meer dan 0.99n getallen zijn verschillend' dan is het mogelijk om dit door sampling in slechts $O(\sqrt{vn})$ tijd te beantwoorden. Voor het matching probleem is het bijvoorbeeld interessant om in plaats van een optimale matching te vinden, slechts een schatting te geven van de grootst mogelijke matching, waarbij de running time slechts afhangt van de maximale graad d en de error ϵ in plaats van de grootte van het netwerk. Dit kan in $d^{O(1/\epsilon^2)}$ tijd [13], maar onbekend is bijvoorbeeld of dit ook kan in slechts $poly(d/\epsilon)$ tijd. Veel van dit soort interessante ontwikkelingen op het gebied van de computationele aspecten voor *big data* sets worden beschreven in een nieuw te publiceren boek van Blum, Hopcroft, en Kanna [14] waarin beschreven wordt hoe fundamentele theorie over bijvoorbeeld matrix algebra, random graphs, Markov chains, en clustering algorithms toegepast wordt voor het vinden van efficiënte algoritmes voor real-world learning- en optimalisatie problemen.

Mijn eigen onderzoek is gericht op algoritmes voor optimaliseringsproblemen zoals scheduling- en routeringsproblemen, daarbij onder andere gebruik makend van lineaire programmeringstechnieken. Ik zal hier het recente paper *Approximability of average completion time scheduling on unrelated machines* [16] bespreken. Zoals hierboven genoemd zijn ILPs in het algemeen een stuk lastiger op te lossen dan LPs. Veel ILP-algoritmes lossen dan ook herhaaldelijk een LP-relaxatie op om tot een oplossing te komen. In de LP-relaxatie wordt bijvoorbeeld een binaire voorwaarde $x_{ij} \in \{0,1\}$ vervangen door de voorwaarde $0 \leq x_{ij} \leq 1$. Voor schedulingproblemen zien we hetzelfde fenomeen. Een variabele $x_{ij} \in \{0,1\}$ geeft bijvoorbeeld aan of job j wordt toegewezen aan machine i . In de relaxatie is het dan toegestaan om een job over verschillende machines te verdelen wat het probleem over het algemeen eenvoudiger maakt. Neem bijvoorbeeld het volgende load-balancing probleem. We willen n jobs gelijkmatig verdelen over 2 machines. Gelijkmatig

betekent in dit geval dat we het verschil in machine-load willen minimaliseren. Het vinden van de optimale oplossing is lastig maar wordt triviaal als jobs gesplitst mogen worden. Plaats bijvoorbeeld elke job voor 50% op elk van de twee machines ($x_{1j} = x_{2j} = 0.5$ voor alle jobs j). Indien jobs gesplitst mogen worden dan vereenvoudigt dit het vinden van een goede oplossing.

In het artikel [16], laat ik onder andere zien dat voor een elementair scheduling probleem (dat van het minimaliseren van de gemiddelde job completion time) er iets merkwaardigs gebeurt. Ruim 40 jaar geleden had Horn [15] al aangetoond dat dit schedulingprobleem te formuleren is als een matching probleem en dus efficiënt op te lossen is met bijvoorbeeld Edmond's matching algoritme. Echter, de relaxatie waarin jobs gesplitst mogen worden blijkt een heel stuk lastiger op te lossen. Het probleem is zogezegd *APX-hard*. Het is daarmee het enige bekende scheduling probleem met deze merkwaardige eigenschap: De fractionele relaxatie is lastiger dan het niet-fractionele (het integer) probleem!

Het bewijs van deze merkwaardige eigenschap is ook opmerkelijk te noemen. De stelling wordt gereduceerd tot een tweetal ILPs! Er wordt aangetoond dat de stelling correct is indien het volgende geldt: De optimale waarde van het ene ILP is gelijk aan 15 en de optimale waarde van het andere ILP is strict groter dan 15. Vervolgens kan het bewijs dus geverifieerd worden door een willekeurige ILP-solver de twee ILPs te laten oplossen. Bij het indienen van het artikel was het commentaar van de reviewers dat ILP-solvers ook wel eens fouten maken en dit daarmee dus geen geldig bewijs is. Als argument voerde ik aan dat iedereen heel eenvoudig elke solver kan proberen tot in het einde van dagen. Mocht ooit blijken dat alle ILP-solvers jarenlang gefaald hebben op dit punt en dat de optimale waarden toch anders zijn dan geclaimd (wat uiterst onwaarschijnlijk is) dan is het bewijs daarmee aantoonbaar onjuist maar dat is dan ook direct helder. Het alternatief voor het ILP-bewijs is een vele pagina's lange analyse welke door één of twee reviewers één maal wordt nagelopen. Gelukkig heeft dit argument (dat reviewers onbetrouwbaarder zijn dan ILP-solvers op dit punt) de reviewers overtuigd en is het artikel onlangs

gepubliceerd in een toptijdschrift in mathematical optimization. Waar linear programming al niet goed voor is.

Referenties

- [1] G. Dantzig, Linear Programming and Extensions, Princeton University Press (1963).
- [2] Martin Grötschel (Editor), Optimization Stories, Documenta Mathematica, Extra Volume for the 21st ISMP, Berlin (2012).
- [3] J. Edmonds, Paths, trees and flowers, *Canadian J. of Math.* 17 (1965) 449-467.
- [4] J. Berge, Two theorems in graph theory, *Proc. Nat. Academy of Sciences (U.S.A.)* 43 (1957)842-844.
- [5] W.T. Tutte, The factorization of linear graphs, *J. London Math. Soc.* 22 (1947) 107-111.
- [6] R. Gomory, Outline of an Algorithm for Integer Solutions to Linear Programs, *Bulletin of the American Mathematical Society* 64 (1958) 275-278.
- [7] 8th DIMACS Implementation Challenge: The Traveling Salesman Problem, <http://dimacs.rutgers.edu/Challenges/TSP/>
- [8] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the theory of NP-Completeness, Freeman and Company (1979).
- [9] M. Agrawal, N. Kayal, and N. Saxena. PRIMES is in P. *Annals of Mathematics* 160 (2004) 781-793.
- [10] L. G. Khachiyan, A polynomial algorithm in linear programming (in Russian), *Doklady Akademii Nauk SSSR* 244 (1979), 1093-1096.
- [11] *Millennium Problems*, Clay Mathematics Institute, <http://www.claymath.org/millennium-problems>
- [12] The P-versus-NP page, maintained by Gerhard Woeginger, <https://www.win.tue.nl/gwoegi/P-versus-NP.htm>
- [13] Y. Yoshida, M. Yamamoto, and H. Ito. An improved constant-time approximation algorithm for maximum matchings. In *ACM Symposium on Theory of Computing*, pages 225-234, 2009.
- [14] A. Blum, J. Hopcroft, and R. Kannan, Foundations of Data Science, *Draft January 2017*, <https://www.cs.cornell.edu/jeh/book.pdf>
- [15] W. Horn, Minimizing average flow time with parallel machines, *Operations Research* 21 (1973) 846-847.
- [16] René Sitters, Approximability of average completion time scheduling on unrelated machines, *Mathematical Programming* 161 (2017) 135-158.



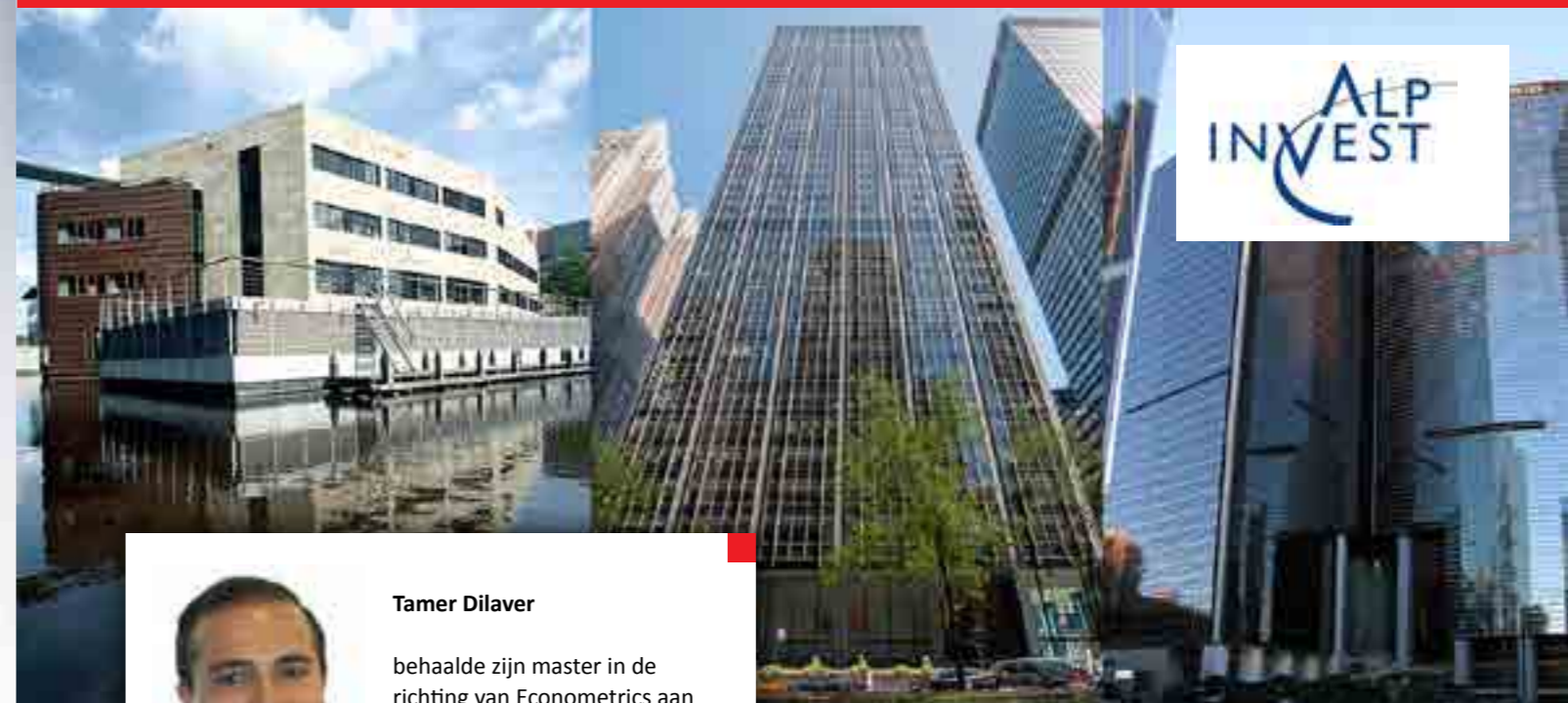
De Belastingdienst wil het mensen steeds makkelijker maken. Dat kan niet zonder data scientists die altijd het overzicht behouden.



Met miljoenen inwoners in Nederland is inzicht in data bij de Belastingdienst onmisbaar. In ons centrale datafundament willen we met informatiedashboards de gewenste gegevens naar boven kunnen halen. Die daar-

na door data scientists met een helikopterview scherp kunnen worden geanalyseerd. Voor zulke data specialisten is er mooi en belangrijk werk te doen bij de Belastingdienst. Meer weten? Kijk op werken.belastingdienst.nl.

Nederland kan niet zonder **de Belastingdienst** kan niet zonder jou.



Tamer Dilaver

behaalde zijn master in de richting van Econometrics aan de Vrije Universiteit en werkt sinds oktober 2016 bij AlInvest Partners als Analyst.

De week van Tamer Dilaver

AlInvest is een toonaangevende wereldwijde private equity investeerder met \$41 miljard assets under management (per 31 december 2016). Het bedrijf heeft sinds haar oprichting in 2000 globaal gezien ruim \$59 miljard aan kapitaal geïnvesteerd, met kantoren in New York, Amsterdam en Hong Kong. De activiteiten van AlInvest omvatten een brede waaier van private equity-investeringen, waaronder primary fund commitments, co-investments, and secondary purchases. Voor SECTOR vertelt Tamer Dilaver over hoe zijn week als Analyst eruit ziet.

Dinsdag

We beginnen de week na Pasen met het lezen van e-mails en het drinken van een kopje koffie. Er zijn grofweg twee richtingen binnen mijn afdeling Portfolio & Risk Analysis. De eerste groep houdt zich voornamelijk bezig met de Audit en de tweede groep bestaat uit Quants. Ik behoor tot de Quant groep en daarom lees en beantwoord ik de e-mail van een Analyst in New York. Ik help met zijn berekeningen van de Internal Rate of Return, i.e. een gecompliceerde performance indicator die rekening houdt met de duur van een investering.

Bij AlInvest is het heel erg belangrijk dat je eigenaar bent van je taken en op tijd aangeeft wanneer je een deadline niet gaat halen. In verband met collega's die op vakantie zijn en door verschuivingen van deadlines herverdelen we prangende taken

onderling. Ondertussen bespreken we hoe onze weekenden waren en bekijken we de to-do lijst. De Limited Partners zijn de investeerders in een fonds en de General Partners zijn de beheerders van een fonds. Ik krijg uitleg over de opzet van nieuwe Limited Partnership Agreements van een nieuw fonds en de allocatieregels die daarvoor gelden van een Associate Vice President. Belangrijk is namelijk dat alle investeerders eerlijk en gelijk behandeld worden. Ondertussen stel ik de wekelijkse agenda op van de team meeting van vanmiddag. Ik bespreek nog een aantal openstaande issues van donderdag en bespreek mogelijke oplossingen met een Associate.

Wij voorzien investeerders van o.a. maand-, kwartaal-, halfjaar- en jaar rapporten. Ik begin met het opstellen van de kwartaalrapporten naar de

investeerders toe en controleer de automatische berekeningen die ons geautomatiseerde systeem heeft gegenereerd. We lunchen aan de overkant van het gebouw waar een grote gezamenlijke kantine is voor verschillende bedrijven.

Het is tijd voor een tweede kopje koffie met een koud glaasje water. Vervolgens werk ik verder aan de kwartaalrapporten. Ontwikkeling en zelfreflectie zijn belangrijke aspecten bij AlInvest. Ik heb met een Associate Vice President maandelijks een update sessie waarin ik feedback krijg over mijn functioneren binnen AlInvest. Ik krijg ook ruimte om aan te geven of ik verbeterpunten zie.

Soms moet ik aan het systeem sleutelen, en dan schrijf ik code in verschillende programmeertalen. Ik bouw bijvoorbeeld een nieuw jaarlaag in Crystal, onze rapportage software van SAP.

We houden onze wekelijkse team meeting met een video conference call met de collega's uit New York. Door onze kantoren in Amsterdam, New York en Hong Kong kunnen we 24 uur per dag onze klanten over de hele wereld te woord staan. Ik werk aan een Capital Call en als die voorbereid is vraag ik een collega Analyst of hij die kan reviewen. Ik haal een peer uit de dagelijkse fruitmand. Ik eindig de dag met het draaien van rapporten die veel rekenkracht kosten, zodat ze morgenochtend gedraaid zijn.

Woensdag

We gaan meteen kijken of het draaien van de rapporten gelukt is. We maken vervolgens wat handmatige aanpassingen. Ik besluit dan een kopje koffie te halen vergezeld met een mini kokos macrou. Investment Accounting is de afdeling die ervoor zorgt dat de gegevens van de kasstromen in ons systeem worden geboekt. Ik verwijder een bug uit de code. Vervolgens komt er een collega van Investment Accounting langs, en hij vraagt ad hoc om een verzoek dat ik meteen even uitstuur naar hem.

Ik rond de eerste review van het kwartaalrapport af. We sturen overigens nooit wat uit naar collega's binnen AlInvest of naar klanten zonder dat er minstens een tweede persoon naar heeft gekeken. Dit is ons zogenaamde 'Four Eyes Principle'. We

bestellen lunch van de Mech voor het team. Ik heralloceer een taak naar een collega uit het team. We eten onze broodjes achter de pc.

Ik bespreek met een collega-Analyst of er mogelijk een aanpassing gemaakt moet worden of niet. Vervolgens ga ik naar Investment Accounting om de aanpassing door te laten voeren. Capital Calls zijn doorgaans momenten aan het begin van een fonds waarin je geld opvraagt van je investeerders om een investering te maken. Distributions zijn doorgaans momenten aan het einde van het fonds waarbij gelden worden uitgekeerd aan de investeerders. Door de aanpassing moet ik opnieuw naar de Capital Call kijken of die aanpassing wel goed is doorgevoerd. Ik werk vervolgens snel mijn e-mails door. Ik informeer een teamgenoot dat hij de kwartaalrapporten kan nakijken.

“Ontwikkeling en zelfreflectie zijn belangrijke aspecten bij AlInvest.”

We passen de geografische regio aan van een investering van Europe naar Non-Traditional Markets in ons systeem door een overname van een onderliggend portfolio bedrijf. Hierdoor moet ik opnieuw de kwartaalrapporten runnen. Ondertussen nemen we bestellingen op van collega's en bestellen we maaltijden voor vanavond. Ik begin met het voorbereiden van maandelijks en driemaandelijks DNB-rapporten. DNB houdt namelijk toezicht op instellingen zoals AlInvest. Onze maaltijden zijn gearriveerd en we hebben avondeten op kantoor. Ik verstuur een aantal e-mails naar collega's. Ik rond de maandelijks DNB-rapportage af. Tot slot update ik de to-do lijst zodat de Principal kan zien wat er allemaal nog openstaat. Wanneer er iets niet op tijd afgerond dreigt te worden dan kan hij middels deze lijst alsnog tijdig ingrijpen.

Donderdag

Ik lees mijn e-mails door en drink een kopje koffie. Als AlInvest moeten wij natuurlijk altijd up-to-date zijn van alle state-of-the-art ontwikkelingen binnen

Private Equity. Ik lees een interessant artikel over creditfaciliteiten voor Private Equity. Vervolgens lees ik commentaar op mijn werk van mijn collega dat ik vervolgens meteen aanpas. Ik moet naar huis, omdat ik met een verbouwing bezig ben moet ik met een aantal aannemers praten. We hebben doorgaans wel standaard kantooruren maar er is altijd ruimte voor flexibiliteit bij AlInvest. Ik haal snel een broodje makreelsalade van een foodtruck en die eet ik op kantoor. Het verschil tussen studeren en werken is dat je opdrachten die vergelijkbaar zijn met een case study of integratiepracticum foutloos moet afronden binnen vijf werkdagen in plaats van een hele maand. Ik pas de DNB-rapporten aan op basis van commentaar van een Associate. Ik lever de rapporten vervolgens aan DNB. We hebben een borrel op kantoor die georganiseerd is door collega's in Amsterdam die in april jarig zijn. AlInvest heeft naast borrels o.a. een jaarlijks kerstfeest, een jaarlijkse ski-trip voor Young Guns en participeert bijvoorbeeld sporadisch aan een voetbaltoernooi met collega's.

Vrijdag

Ik drink snel even een smoothie die mijn collega meebrengt voor ik begin. In verband met ons nieuwe fonds werkt ons allocatie model niet meer voor kosten die gemaakt worden per fonds. Mijn dag staat vooral in het teken van het extraheren van data uit onze database en het maken van Excel-spreadsheets. Het proces om tot geschikte allocatiemodellen te komen is behoorlijk in lijn met een praktische toepassing van het vak wiskundige economie. We eten buiten in de zon een broodje zalm bij de foodtruck om de hoek met het team. Vervolgens drink ik een cappuccino met een glaasje water. Ik rond het nieuwe model af en stuur mijn resultaten op naar een collega in New York die mijn model moet goedkeuren voor maandag. Dit was het einde van mijn werkweek. Stuur me een bericht als je meer wilt weten over AlInvest!



Erwin Huizenga

behaalde zijn master in de toegepaste statistiek en werkt sinds 2014 bij SAS, op dit moment als pre-sales consultant.

SAS, The Power to Know

SAS is specialist in business analytics software en dienstverlening en de grootste onafhankelijke business intelligence leverancier. Hiermee helpt SAS klanten op meer dan 80.000 locaties hun prestaties te verbeteren en waarde te creëren door sneller, betere beslissingen te nemen. Het Nederlandse kantoor is gevestigd in Huizen op landgoed 'Oud-Bussem', waar ruim 170 medewerkers werken. Hier spreken wij met Erwin Huizenga.

Kunt u iets over uzelf vertellen?

Mijn naam is Erwin Huizenga en ik ben inmiddels zo'n 2,5 jaar werkzaam bij SAS, waarvan inmiddels bijna een jaar als pre-sales consultant. Tussendoor ben ik een periode werkzaam geweest bij IBM in een development team dat onder andere met het Watson team samenwerkte. Toch ben ik teruggekeerd naar SAS, omdat ik meer energie krijg van de innovatieve projecten bij SAS en een wat jonger team waar wij mee werken. Ook het feit dat we ons bij SAS volledig richten op de data-analyse spreekt mij erg aan.

Ik ben na mijn studie echter niet direct bij SAS aan de slag gegaan. Ik ben bij een startup begonnen die webapplicaties bouwt voor IBM. Deze software was gericht op bedrijven die veel met calamiteiten te

maken hebben, zoals bedrijven die grote netwerken in de lucht moeten houden, of bedrijven waarbij security erg belangrijk is. Omdat er bij die startup nog weinig mensen werkten, had ik een hybride rol waarbij ik van alles deed; van development tot bij klanten zitten.

Wat zijn uw werkzaamheden als pre-sales consultant?

Als pre-sales consultant adviseer je de klant in het voortraject, wanneer de klant met een bepaald probleem of een uitdaging zit. Wij adviseren ze dan hoe ze dat het beste kunnen aanpakken. Voor problemen waar nog geen bestaande oplossing voor is, maken wij dan een prototype om aan de klant te kunnen bewijzen dat wij een goede oplossing voor

hun probleem hebben. Hiervoor stelt de klant ons meestal data ter beschikking. Het is echter geen eenrichtingsverkeer, je moet het traject van vraag naar oplossing echt zien als een samenwerking. Je gaat samen met de klant bedenken welke oplossing toegevoegde waarde kan leveren. De samenwerking met de mensen die de applicatie zullen gaan gebruiken is daarbij erg belangrijk. Na deze fase haken er andere consultants aan die de software operationeel maken en daadwerkelijk gaan installeren bij de klant.

“Als softwarebedrijf ben je genoodzaakt om grote vernieuwingen door te voeren om bij te blijven met de ontwikkelingen in de technologie.”

Om de klant te overtuigen dat wij de dingen kunnen doen die zij vragen, moet je veel inhoudelijke kennis betreffende lopende projecten hebben. Verder moet je communicatief vaardig zijn en innovatief kunnen denken. Wat ik ook regelmatig doe, is op conferenties spreken over waar wij mee bezig zijn. Het gave van deze functie als consultant is dat je altijd aan het voortraject zit en je daardoor veel verschillende klanten ziet en veel verschillende dingen doet. Je leert daar ontzettend veel van.

Kunt u iets over het bedrijf SAS vertellen?

SAS bestaat inmiddels 41 jaar, maar is ooit begonnen met een stel studenten die statistiek studeerden aan de North Carolina University. Zij deden statistisch onderzoek op het gebied van landbouw. De middelen die zij daarvoor tot hun beschikking hadden vonden zij niet toereikend. Ze hebben vervolgens hun eigen statistische taal ontwikkeld, die ze konden gebruiken voor hun onderzoek. Dat was zo'n groot succes, dat de software de universiteit ontgroeide en zo zijn ze hun eigen bedrijfje gestart. Dat bedrijfje is uitgegroeid tot SAS zoals we het nu kennen,

met wereldwijd ruim 14.000 werknemers. Sinds 1987 zijn wij actief in Nederland.

Wij doen vrijwel alles op het gebied van data: data management, data quality, het integreren van verschillende databronnen, et cetera. We hebben echter zelf geen databases, en dat is meteen ook onze kracht. We kunnen koppelen met alle verschillende databases en platformen die er zijn, van Adobe tot SQL. Na het beschikbaar maken van de data is de eerste stap het visualiseren van de data. Hiervoor is Business Intelligence nodig om de eerste inzichten uit de data te halen. Daarna komt het bouwen van modellen, zoals machine learning, forecasting, optimalisatie of operations research technieken. Uiteindelijk moet het ook tot iets leiden: we noemen dit van data tot deployment. Wij doen dus alles wat met analytics te maken heeft.

Kunt u iets vertellen over de SAS software?

De basis is de SAS programmeertaal. Het is een macrotaal waarmee je zowel zelf dingen kunt programmeren, als ook standaard snippets gebruiken zoals een regressiemodel. Hiermee maken we tools voor bijvoorbeeld het verwerken van data en het implementeren van machine learning en forecasting modellen. Daarbovenop zijn GUI's gemaakt voor de verschillende doeleinden.

Op dit moment zijn we bezig met een nieuw platform, dat meer gericht is op de cloud-technologie. Dit platform kan in de cloud runnen en is daarnaast makkelijk op- of terug te schalen. Daarin zie je de visie terugkomen dat je één GUI hebt, maar afhankelijk van je rol bepaalde functionaliteiten krijgt. Dit platform is tweeledig, we hebben een code variant en een GUI variant. Sommige mensen willen graag zelf coderen, anderen werken liever visueel zodat je snel dingen kunt maken. Het platform wordt ook meer open, zodat je vanuit je R-studio, Java of Python bij de engine kan.

Deze verandering is noodzakelijk. Eigenlijk ben je als softwarebedrijf genoodzaakt om circa om de 10 jaar grote vernieuwingen door te voeren, zodat je bijblijft met de ontwikkelingen in de technologie. Wanneer de technologie verandert, verandert ook de vraag in de markt. Als je hier niet op inspeelt, denk ik dat je uiteindelijk de boot zal gaan missen.

Hoe wordt dit nieuwe platform uitgerold?

Met dit soort ontwikkelingen gaan altijd grote migratietrajecten gepaard. Bedrijven die al wat langer klant zijn, hebben veel applicaties die kunnen op het oude platform. Hierbij gaat de migratie stap voor stap. Bij nieuwe klanten, zoals bijvoorbeeld startups, kun je direct met het nieuwe platform aan de slag.

Wij zijn als softwarebedrijf echter niet de enige die met deze ontwikkelingen te maken hebben. Veel andere bedrijfstakken moeten zich om de zoveel jaar ook opnieuw uitvinden, om zo mee te kunnen blijven doen met de ontwikkelingen in de technologie. Zij worden genoodzaakt om te innoveren en te veranderen, om te voorkomen dat zij in worden gehaald door de nieuwe startups. Dit is ook goed, het houdt iedereen scherp. Wij adviseren hen in hoe ze die stappen kunnen zetten.

Verder geloof ik echt dat data-science/analytics een echte driver is voor die innovatie. Je ziet namelijk ook heel sterk dat succesvolle bedrijven, zoals Airbnb, Uber of Netflix, analytics in hun DNA hebben. Zij denken in data. Daar is waar het denk ik naartoe gaat, en waar iedereen uiteindelijk ook wel naartoe moet. De een gaat hierbij sneller dan de ander.

Waar wordt jullie software gemaakt?

De producten worden allemaal in Amerika gemaakt. Onze R&D, marketing en sales afdeling bevindt zich op de campus in North Carolina. Het is na Silicon Valley de grootste research-triangle van Amerika met zo'n 4 à 5000 mensen, er zitten grote bedrijven als Redhead en IBM. Je ziet tegenwoordig dat de grote relatief jonge bedrijven als Google campussen met zwembaden, een tandarts, kapper et cetera hebben, maar eigenlijk is SAS daar ooit mee begonnen. Tegenwoordig zitten er ook R&D afdelingen in China.

We hebben meer dan 200 producten, in veel verschillende domeinen. Zo houden wij ons bijvoorbeeld bezig met het domein van fraude, risk, marketing en finance, maar ook met industrieën als telecomebedrijven, banken, ziekenhuizen, universiteiten en de overheid.

Waarom kiezen jullie klanten voor de software van SAS?

Wij hebben erg veel ervaring op het gebied van data-analyse, en hebben voor veel problemen een bestaande oplossing. Software-development is echt specialistisch werk, waar vaak de nodige research & development bij komt kijken. Je zou er als bedrijf natuurlijk voor kunnen kiezen om toch zelf de tool te bouwen, maar de ervaring leert dat dit veel meer tijd kost. Met onze ervaring kunnen wij dit veel sneller en daarnaast kunnen wij een zekere kwaliteit garanderen. Als bedrijf wil je dat de data-scientist zich bezig kan houden met de dingen waar hij waarde kan toevoegen, namelijk data-analyse, zonder dat hij daarvoor 80% van zijn tijd kwijt is aan het knutselen aan data in een Excelbestand. Dat is zonde van de energie, tijd en kennis. Wij verkorten de zogenaamde time-to-value, en bovendien zijn onze tools schaalbaar, governbaar en herhaalbaar. Verder zijn we in staat om heel snel van een idee naar prototype naar productie te gaan.

“Je onderscheidt je echt als je op het snijvlak van IT en business kan lopen: dat je weet wat er op IT gebied gebeurt en tegelijkertijd de vertaalslag naar de business kunt maken.”

Alles wat wij bouwen is ‘integrated by design’. Met onze software kun je erg makkelijk en natuurlijk schalen en bouwen aan je systeem. Er zijn zelfs startups die de kennis in huis hebben om het zelf te doen, maar toch voor een samenwerking met ons kiezen omdat ze dan veel sneller een product op de markt kunnen brengen. Zij kunnen zo veel sneller waarde genereren.

Waarom onderscheidt SAS zich van de concurrenten?

Wij onderscheiden ons voornamelijk van de concurrent met onze focus, kennis en ervaring. Wij

focussen ons volledig op analytics, hier ligt dus echt ons specialisme. Omdat wij al ruim 40 jaar actief zijn op dit gebied, is er voor vrijwel ieder probleem al wel een integrale of gedeeltelijke oplossing beschikbaar. We zijn bescheiden op het gebied van marketing maar dat past ook bij onze oorsprong, we gaan uit van onze eigen kracht. Verder hebben we geen aandeelhouders, we zijn dus niet kwartaal gedreven. Dat verschil merkte ik wel toen ik bij IBM werkte, waar ze wel aandeelhouders hebben. De werkwijze is daar meer sales-gedreven. De competitive advantage van een softwarebedrijf ligt echter juist bij de innovatie. Bij SAS wordt ongeveer 25% van de omzet direct in R&D gestoken. Wij geloven heel erg in het blijven ontwikkelen van nieuwe dingen, en het nadenken over hoe nieuwe technologieën benut kunnen worden.

Met welke projecten zijn jullie momenteel bezig?

Dat zijn er aardig wat. Een groot pensioenfonds in Nederland wil haar data-science capabilities opbouwen. Ze zijn bijvoorbeeld druk bezig met het werven van data-scientists. Wij helpen hen om deze capabilities te ontwikkelen door mee te denken over wat ze nodig hebben qua technologie. We helpen ze om volwassen te worden op het gebied van data-gedreven werken en denken. Die reis vind ik heel gaaf.

Daarnaast heeft een telecomeaanbieder een aantal uitdagingen aan de marketingkant, waarbij we ze helpen met onze technologie. Ze willen graag kunnen voorspellen of een klant potentieel weggaat, zodat ze daar extra aandacht aan kunnen besteden. En voor een groot ziekenhuis zijn we bezig met het ontwikkelen van CT-scan recognition, voor het opsporen van kankercellen. Omdat deze case inhoudelijk erg technisch en innovatief is, werken we hiervoor samen met onze R&D afdeling.

Wat is de verhouding technische kennis vs. softskills die nodig zijn wanneer je werkt bij SAS?

Tijdens de studie, zeker bij bèta-studies, ligt de focus voornamelijk op de theoretische inhoud. Er wordt erg weinig aandacht besteed aan het maken van de vertaalslag naar de business. Het is natuurlijk leuk en gaaf dat je een geavanceerd model uit kunt werken, maar het gaat er hier uiteindelijk om de business te ondersteunen. Het is dus essentieel

dat je ook begrijpt wat de business wil. Om met een idee waarde te kunnen toevoegen voor de eindgebruiker is het belangrijk dat je de vertaalslag kunt maken, maar ook dat je vooruit kunt kijken, over verschillende business units heen kunt denken en communicatieve vaardigheden zijn ook erg belangrijk.

Je onderscheidt je echt als je op het snijvlak van IT en business kan lopen, dus dat je weet wat er op IT gebied gebeurt, en daarnaast de vertaalslag kan maken naar de business. Naast het technische aspect kun je je dan ook richten op de sturing, strategie en visie. Je bent dan enorm waardevol. Dit is een tip voor de studenten; ga op dit snijvlak zitten. Zorg dat je stageloopt en bij bedrijven rondloopt als het je ambitie is om deze richting op te gaan.

Ten tweede is creativiteit ook heel belangrijk. Dit wordt wel eens onderschat. Bij het bedenken van business oplossingen komt heel veel creativiteit kijken. Hoe ga je er nieuwe data bij trekken als je model niet doet wat je wilt? Hoe ga je je resultaten visualiseren? Hoe ga je het eindresultaat implementeren en ervoor zorgen dat de business er waarde uit kan halen? Bij al deze vragen komt veel out-of-the-box denken kijken. Dat maakt het werk ook zo uitdagend.

Gebruik je nog veel van de kennis die je hebt opgedaan tijdens je studie?

De kennis die ik vanuit mijn studie nog veel gebruik is voornamelijk op theorie gebied. Toch is dit niet een-op-een. In de praktijk is het vaak trial-and-error en laat je veel theoretische afwegingen achterwege. Er komt vaak de eerste keer niet uit wat je wilt, de tweede keer kan je er dan bijvoorbeeld nieuwe data bij doen. Het is altijd tweaken en proberen. Ook een dataset bij een bedrijf is nooit zoals je hem tijdens je studie krijgt. Tijdens een discussie met een ervaren data-scientist bleek ook dat in de praktijk niet alle theoretische checks structureel worden nagelopen zoals je dat tijdens je studie doet. Veel promovendi of theoretisch wiskundigen zijn vaak op zoek naar het perfecte model met de perfecte parameters. Bij sommige cases kun je je echter afvragen of het de moeite waard is dat je model in plaats van 90%, voor 91% goed kan voorspellen, gegeven de tijd die deze verbetering kost.

Universiteiten zouden meer aandacht kunnen besteden aan de link tussen theorie en business vaardigheden, want op dit moment merken wij dat deze brug voor pas afgestudeerden vrij groot is. Ook aan creativiteit zou meer aandacht besteed mogen worden, omdat dit een niet te overschatten skill is.

“Waar bij sollicitaties vaak naar gekeken wordt, is in hoeverre iemand in staat is om nieuwe dingen te leren en op te pakken.”

Wat is de achtergrond van de mensen die bij SAS werken?

Bij SAS werken zeker niet alleen mensen die econometrie of business analytics hebben gestudeerd. Zo werken er mensen met een achtergrond in de wiskunde, informatica, maar ook bedrijfskunde, geschiedenis of politicologie. Econometristen zijn natuurlijk niet de enigen die een regressiemodel kunnen bouwen. Het belangrijkste is dat je een passie hebt voor data-science en research en er ook affiniteit mee hebt. Daarnaast is het belangrijk dat je weet hoe het proces werkt, en dat je open staat voor nieuwe dingen. Iedereen die bij ons met klanten in aanraking komt heeft wel een dergelijke data achtergrond, hetzij met deskundigheid voor een verschillende richting. Hierdoor is de totale set aan vaardigheden groter dan wanneer er alleen maar econometrie studenten zouden werken. Naarmate je vaker met klanten in aanraking bent geweest en meer situaties als consultant hebt meegemaakt, zul je je gaan specialiseren in een bepaalde richting, bijvoorbeeld banking, insurance of de commerciële/publieke sector.

Nemen jullie veel nieuwe studenten aan, voor stageplekken of startersbanen?

Omdat wij in Nederland niet een hele grote organisatie hebben, met circa 180 mensen, zijn we niet continu op zoek naar nieuwe studenten.

Veel van onze klanten en partners zoeken echter wel altijd studenten. Wat wij kunnen aanbieden is dat we studenten ter voorbereiding op hun loopbaan SAS leren programmeren, door tijdens het academische programma bijvoorbeeld cursussen en hackatons te organiseren. Omdat onze software op heel veel plekken gebruikt wordt, is de kans groot dat je het tegen gaat komen. Als je er dan al eens mee in aanraking bent geweest, is dat zeker een voordeel. Daarnaast krijg je voor onze cursussen een speciaal certificaat, waarbij je een extra stukje programmeerkennis aan je CV kunt toevoegen. Dit certificaat heeft een bepaalde waarde bij bedrijven. Als iemand echt goed is zullen wij zeker kijken of we diegene kunnen aantrekken, maar dat is niet onze eerste insteek. De markt schreeuwt echter wel om mensen met een goede opleiding, waar econometrie zeker onder valt.

Heeft u nog tips voor de studenten?

Je hebt een voorsprong als je leergierig bent en jezelf dingen aanleert. We hebben bijvoorbeeld een online University Edition, download die een keer, ga er mee aan de slag om zo SAS te leren programmeren. Als je die instelling hebt, dan ben je denk ik perfect voor dit werkveld. Dan val je ook op tussen de andere studenten, want bedrijven zien dat je geïnteresseerd bent. Je zou bijvoorbeeld ook mee kunnen doen aan een hackaton. Zo heb ik bijvoorbeeld veel van mijn kennis opgedaan door zelf aan apps te bouwen en te programmeren.

Waar bij sollicitaties tegenwoordig vaak naar gekeken wordt, is in hoeverre iemand in staat is om nieuwe dingen te leren en op te pakken. Deze wereld verandert zo ontzettend snel, dat dit key is. Zo weet ik via een contact bij Google dat wanneer zij op zoek zijn naar developers of data-scientists, ze het niet per se belangrijk vinden dat je alles uit je hoofd kent, maar dat het belangrijker is dat je in staat bent om snel de juiste informatie te vinden en om nieuwe dingen te proberen.

Nog een tip voor de mensen die net beginnen in de data-science wereld: het is belangrijk dat je zoveel mogelijk in je rugzak hebt, daarmee verhoog je je waarde. Daar hoort naast veel open-source kennis wat mij betreft ook zeker SAS bij.



AEGON

Aandacht voor je toekomst

Eeuwige student... én data talent?

Wil jij de studieboeken achter je laten en de stap zetten naar een eerste baan? Óf blijf je liever verder leren? Misschien wil je nog niet kiezen. En dat snappen we bij Aegon heel goed! Daarom bieden wij analytisch talent een traineeship aan. Dan breng je kennis en kunde meteen in de praktijk. En leer je de fijne kneepjes van het vak met trainingen en opleidingen. Zo blijf je als analytisch talent – nog héél even – die eeuwige student.

Meer weten? Ga naar:

werkenbijaegon.nl

Column

Een paradoxaal resultaat in een serie muntworpen?

Kansen bestaan niet, of zoals de Italiaanse wiskundige Bruno de Finetti (1906-1985) schreef, *La probabilità non esiste*. Met deze provocerende uitspraak begon ik vaak het college kansrekening om onze eerstejaars van meet af aan in te prenten dat kansen geen onderdeel uitmaken van de (empirische) werkelijkheid, maar louter binnen het *wiskundevak* kansrekening 'het levenslicht zien' door geïntroduceerd te worden als een maat [wiskundig begrip!] die voldoet aan de axioma's van Kolmogorov. Kansen worden via deze axioma's *gedefinieerd* en alle stellingen uit de kansrekening worden afgeleid [bewezen] met alleen deze axioma's als uitgangspunt. Voor het beoefenen van de kansrekening is een nadere interpretatie van het kansbegrip overbodig! Ter verduidelijking, ook rechte lijnen komen we niet in de fysische werkelijkheid tegen, ook zij 'bestaan' alleen binnen de wiskunde, in casu de meetkunde, waar rechte lijnen worden *gedefinieerd* via een stelsel axioma's. Dit kunnen de Euclidische axioma's zijn, maar ook alternatieve niet-Euclidische axiomastelsels zijn mogelijk, en verschillende axiomastelsels definiëren verschillende soorten 'rechte lijnen'. Bij het toepassen van de meetkunde zal men bepaalde entiteiten, bijvoorbeeld een lichtstraal of een strak gespannen touw, *modelleren* als een rechte lijn, maar een lichtstraal en een strak gespannen touw zijn zelf geen rechte lijnen. Het identificeren van objecten uit de werkelijkheid met bepaalde meetkundige vormen [rechte lijnen, cirkels, et cetera] is bijna nooit problematisch en de resultaten van een meetkundige toepassing vinden we via de uit het gekozen axiomastelsel af te leiden stellingen door die simpelweg terug te vertalen in termen van de toepassing.

Bij het toepassen van de kansrekening ligt dit

allemaal een stuk moeilijker, omdat het niet zo eenvoudig is entiteiten in de werkelijkheid aan te wijzen die als kansen gemodelleerd kunnen worden. Dat verklaart ook de uitvoerige 'filosofische' discussies over het kansbegrip die nooit definitief beslecht zijn, zoals dat gebruikelijk is met filosofische discussies. Voor *frequentisten* zijn kansen *langetermijnfracties* en voor *Bayesianen* zijn kansen een *degree of belief*. Beide opvattingen zijn problematisch om de doodeenvoudige reden dat noch een 'langetermijnfractie', noch een 'degree of belief' in de werkelijkheid is te lokaliseren. De langetermijnfractie niet omdat deze notie per definitie gebaseerd is op het *wiskundeconcept* 'limiet', dus dat brengt ons weinig verder, en ook de kansopvatting van een 'degree of belief' brengt geen soelaas omdat een 'mate van geloof' een vage notie is die verdere definiëring vereist wil men er kwantitatief mee uit de voeten kunnen. Vele voorstellen zijn in dezen gedaan, o.a. door genoemde Bruno de Finetti, maar zoals gezegd, de discussie over het kansbegrip is nooit definitief beslecht ten gunste van de ene of andere stroming, en duurt voort tot op de dag van vandaag. Gelukkig kan de kansrekenaar los van deze discussies zijn vak verder ontwikkelen en zelfs bovengenoemd uitgangspunt dat kansen niet bestaan anders dan binnen het vak zelf met een gerust hart volhouden, zoals het ook geen obstakel is voor de meetkundige zijn vak te beoefenen in de wetenschap dat rechte lijnen niet bestaan. Dit sluit overigens niet uit dat kansrekenaars en meetkundigen door hun werkzaamheden een grote vertrouwdheid met en een scherpe intuïtie voor kansen en rechte lijnen ontwikkelen die een belangrijke, zo niet noodzakelijke leidraad en inspiratiebron zijn bij hun onderzoek. Zonder een goede intuïtie voor de abstracte begrippen uit zijn vak staat de wiskundige met lege handen! Je zou

kunnen concluderen dat deze abstracties tot leven gekomen zijn binnen zijn brein, zoals het paard en de looper concrete wezens lijken te zijn geworden in het brein van de schaakspeler, die mede dankzij zijn door het veelvuldig spelen gevormde intuïtie de goede zetten weet te vinden.

Toch zal evenals de meetkundige ook de kansrekenaar zijn vak toegepast willen zien, en daarmee ontkomt hij er niet aan een verbinding te leggen tussen zijn wiskundige begrippenarsenaal [kansen, verwachtingen, onafhankelijkheid, et cetera] en de empirische werkelijkheid. Hij doet dit door het veelvuldig bezigen van metaforen die suggereren dat hij *fysische* experimenten beschrijft. Het spreekwoordelijke voorbeeld is natuurlijk de alomtegenwoordige 'muntworp' of meer algemeen de 'serie van n onafhankelijke muntworpen'. Dit woordgebruik suggereert weliswaar een beschrijving van een eenieder vertrouwd fysisch experiment, maar voor de kansrekenaar dient dit frivole woordgebruik slechts voor het aanduiden van een tweetal *toevalsexperimenten* die specifieke kansmodellen *dicteren*, in casu een *Bernoulli-proef* resp. een *serie van n onafhankelijke Bernoulli-proeven* die los van welk fysisch experiment dan ook gedefinieerd worden. Het wordt eentong, 'muntworpen', i.e. Bernoulli-proeven, bestaan net zo min in de werkelijkheid als rechte lijnen, en als iemand echt een munt opgooit die even later op de grond valt dan *lijkt* deze handeling misschien op een muntworp zoals de kansrekenaar die bedoelt, maar hij is er net zo min mee te identificeren als een lichtstraal met een rechte lijn. We stellen opnieuw vast dat de fysische worp met de munt alleen *gemodelleerd* wordt als een muntworp, dus als een Bernoulli-proef, maar er niet mee *geïdentificeerd* kan worden, omdat wiskunde en werkelijkheid nooit samenvallen. Was het bij het modelleren van lichtstralen als rechte lijnen ook voor de leek niet al te moeilijk het verschil tussen model en werkelijkheid te onderkennen, bij het modelleren van een fysisch uitgevoerde worp met een munt [realiseer je dat deze beschrijving op zich al vaag is; wat is precies een worp met een munt?] als een muntworp [exact vastgelegd als een Bernoulli-proef] zien we buiten de wereld van de kansrekenaars [maar ook daarbinnen!] velen die het fysische experiment en het toevalsexperiment [kansmodel]

wel met elkaar identificeren, waarschijnlijk om via deze sluipteg aan het kansbegrip in de fysische werkelijkheid alsnog een betekenis te kunnen toekennen. Men beseft dan nauwelijks dat men zo leentjebuurt speelt bij de wiskunde en een begrip als kans [of verwachting] gedachteloos de werkelijkheid binnensmokkelt zonder zich te bekommeren over een precieze interpretatie van het kansbegrip. Anders gezegd, de kansrekening laat zijn licht schijnen op het werpen van munten, presenteert zijn resultaten in een terminologie die slechts binnen dat vakgebied precies is vastgelegd, en het 'volk' neemt deze terminologie zonder enige aarzeling over bij het interpreteren van de resultaten. Nog anders geformuleerd, het terugvertalen dat bij meetkundige toepassingen zo eenvoudig was [een rechte lijn wordt weer een lichtstraal] is nu geblokkeerd omdat er geen terugvertaling mogelijk is van de kans theoretische termen: een kans blijft bij het terugvertalen een kans! En zo zijn we terug bij af, kansen bestaan niet, maar de leek gebruikt aan de kansrekening ontleende begrippen alsof die buiten de kansrekening ook in de fysische werkelijkheid een betekenis hebben, anders gezegd alsof kansen los van de kansrekening bestaan. Daarbij komt nog dat men zich dit in het geheel niet realiseert, omdat men dagelijks over kansen hoort spreken. Iedereen praat over kansen alsof hij/zij begrijpt wat ermee bedoeld wordt, en daardoor is de leek bijna gedwongen het verschil tussen model en werkelijkheid te negeren.

Dit verschil tussen werkelijkheid en model wordt ook in de [populair-wetenschappelijke] media te vaak veronachtzaamd, en mede als gevolg hiervan door leken zelfs volledig uit het oog verloren, met kwalijke gevolgen voor het publieke debat en de politieke besluitvorming. De verhitte discussies over immigratie en klimaatverandering zijn twee duidelijke illustraties: de media publiceren al of niet alarmerende voorspellingen, vaak geformuleerd als kansuitspraken, en de argeloze lezer denkt dat de 'wetenschap' hem een blik op de toekomstige werkelijkheid verschaft, terwijl de wetenschap niet meer *kan* doen dan *model* resultaten presenteren, en hoeveel data ook de basis vormen van de gebruikte modellen, het blijven modellen die nooit met de werkelijkheid vereenzelvigd mogen worden, en zeker niet in geval er flink met kansen wordt gestrooid.

Bovenstaande bespiegelingen speelden weer eens door mijn hoofd toen ik onlangs het artikel *Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers* van Joshua Miller en Adam Sanjurjo kreeg doorgestuurd. Zoals de titel aangeeft komen twee zaken aan bod: de *Gambler's Fallacy* en de *Hot Hand Fallacy*. Het eerste 'misverstand' betreft het geloof van veel casino-bezoekers dat bij het spelen van roulette de kans (*sic!*) op 'zwart' na een serie ononderbroken uitslagen 'rood' steeds groter wordt, en het tweede 'misverstand' verwijst naar de overtuiging van bijvoorbeeld basketbalspelers dat na een serie treffers de kans (!) op een volgende treffer toeneemt evenals de kans op een volgende misser na een serie missers, dus precies het tegenovergestelde van wat de gokker in het casino gelooft. De gokker acht het *afbreken* van een serie gelijke uitkomsten steeds waarschijnlijker en de basketbalspeler gelooft juist dat het *voortzetten* van een serie gelijke uitkomsten steeds waarschijnlijker wordt. Dit is op zich geen tegenstrijdigheid, want het betreft twee totaal verschillende omgevingen, namelijk het casino, waar *Vrouwe Fortuna* de scepter zwaait, resp. de sporthal waar het *talent van de speler* de resultaten bepaalt, een gegeven dat niet bepaald uitnodigt tot een modellering als een toevalsexperiment.

Om de Gambler's Fallacy te weerleggen modelleert de kansrekenaar een serie roulette-uitkomsten ['rood' of 'zwart'] als een serie *onafhankelijke* Bernoulli-proeven met succeskans $p=18/36$ [de nul laten we voor het gemak buiten beschouwing], en daarmee lijkt de kous af: de 'weerlegging' van de Gambler's Fallacy krijgen we cadeau via de modellering! Waarom zou je er verder nog woorden aan vuil maken? Een serie uitkomsten met een roulette komt volledig overeen met wat de kansrekenaar modelleert als een serie onafhankelijke muntworpen, dus Bernoulli-proeven, en daartegen zal ook de leek geen protest aantekenen. Natuurlijk kan men twijfelen aan de kwaliteit van de roulette, maar dat is hier niet het punt van discussie, ook bij een 'zuivere roulette' [ook een wiskundig begrip; er bestaan geen zuivere roulettes, ze zijn zelfs niet te definiëren buiten de kansrekening!] gelooft de naïeve gokker dat na 'rood' een uitkomst 'zwart' *waarschijnlijker* is, wat naar de werkelijkheid vertaald weer niets anders wil zeggen dan dat hij

denkt geld te kunnen verdienen door een op dit geloof gebaseerde inzet-strategie te volgen [merk op hoe onvermijdelijk vaag zo'n interpretatie is]. Dus als de gokker accepteert dat de kansrekening een legitiem middel is om over zijn geloof een uitspraak te doen, dan zal hij geen bezwaar maken tegen de voorgestelde modellering, met als resultaat een weerlegging van zijn dwaling! Merk op dat we hiervoor geen realisaties van roulette-uitkomsten [data!] nodig hebben, omdat de Gambler's Fallacy niets anders is dan een onware uitspraak over een *kansmodel*, in casu een serie van onafhankelijke Bernoulli-proeven. Als een gokker dit kansmodel accepteert dan zal hij voorgoed genezen zijn.

Heel anders ligt dit bij de Hot Hand Fallacy, omdat nu niet ondubbelzinnig gerefereerd wordt aan een kansmodel, zelfs in de formulering van de Hot Hand Fallacy is vaagheid troef vanwege het niet nader gespecificeerde gebruik van het woord kans in een reële context, in casu de wereld van het basketbal. Een modellering van een serie schoten op een basket als een serie van Bernoulli-proeven met een vaste succeskans [waarvan de onderlinge (on)afhankelijkheid dan nog nader moet worden onderzocht om de 'hot hand' al of niet te kunnen vaststellen] staat juist haaks op het feit dat het succes van een basketbalspeler gebaseerd is op zijn talent. Daarom is het postuleren van een kans dat een basketbalspeler scoort bij zijn volgende worp al problematisch. Hierboven is uitvoerig betoogd dat kansen niet bestaan, dus iemand die zo'n kans postuleert laadt de intellectuele plicht op zich nader te omschrijven wat hij bedoelt. De frequentist zal opmerken dat daarmee natuurlijk (*sic!*) bedoeld wordt het succespercentage van de speler genomen over een groot aantal *eerder* uitgevoerde worpen. Dat kun je natuurlijk voorstellen, maar de voor de modellering noodzakelijke vervolgstap dat deze geponeerde kans voor de komende serie worpen [de serie die onderzocht gaat worden om de Hot Hand Fallacy te toetsen] relevant is, anders gezegd dat hij in de komende serie worpen weer opnieuw een 'vergelijkbaar' (?) succespercentage zal halen, moet ook worden aangenomen. Met de gepostuleerde succeskans wordt de kwaliteit van de speler kwantitatief gewaardeerd, maar dit blijft natte-vinger-werk. Elk volgend schot wordt door de speler met grote concentratie uitgevoerd, en zo'n

handeling staat wel erg ver af van een draai aan een roulette in een casino, waar de vakbekwaamheid van de croupier juist gebaseerd is op het genereren van series voor de casino-bezoeker [en hemzelf!] *onvoorspelbare* uitkomsten: een croupier met een 'hot hand' zal ontslagen worden, terwijl een basketbalspeler die opereert als een random generator niet snel zal worden opgesteld! De conclusie moge duidelijk zijn: een serie basketbalschoten is niet met goed fatsoen te modelleren als een serie al of niet onafhankelijke Bernoulli-proeven, omdat het postuleren van een succeskans al problematisch is en, nog belangrijker, de speler zich juist zal inspannen de *afhankelijkheid* tussen de opeenvolgende schoten te maximaliseren, hij streeft er tenslotte naar dat *al* zijn schoten treffers zullen zijn! Als kansrekenaar zou je idealiter een model willen construeren dat het gedrag van een [individuele] basketbalspeler bijna perfect nabootst, zoals ook een natuurkundige een bijna perfect model maakt voor bijvoorbeeld de planeetbanen om de zon. Misschien kom je dan uit op een complexe Markov-keten, zodoende toch enige ruimte latend voor stochastische aspecten, maar daarvoor is een goed inzicht in het functioneren [brein plus lichaamscoördinatie] van de basketbalspeler een *conditio sine qua non*, en dat inzicht ontbreekt juist tot nu toe. Als dit inzicht er wel zou zijn, was er ook geen probleem, dan zouden we evenals bij de Gambler's Fallacy het antwoord op de vraag of de 'hot hand' bestaat cadeau krijgen via de modellering die het functioneren van een basketbalspeler modelmatig zou beschrijven. Helaas is dit nu nog niet mogelijk, maar de modellering van een serie basketbalschoten als een serie Bernoulli-proeven duidt erop dat men blijkbaar niet veel hoop heeft ooit enig inzicht in de wisselwerking tussen eerder succes en toekomstig succes te zullen krijgen. De belangrijkste motivering voor het voorgestelde model lijkt het onbegrip van het fenomeen te zijn dat men wil onderzoeken. Bovendien kiest men voor een model waarvan statistische toetsing min of meer standaard is, wat bij een modellering als Markov-keten al een stuk lastiger zou zijn.

De vraag is daarom of een kansrekenaar, gegeven de huidige staat van onze kennis, wel zijn licht kan laten schijnen op de vraag of de 'hot hand' een reële of een ingebeeld verschijnsel is. Mijn antwoord is

nee, maar daar denkt de rest van de wereld duidelijk anders over, getuige het grote aantal artikelen dat over deze materie verschenen is. In al deze artikelen worden reeksen uitkomsten van treffers/missers aan statistische toetsen onderworpen en, natuurlijk geen verrassing, de verschillende artikelen komen tot tegengestelde conclusies, zelfs na gebruik van exact dezelfde 'data reeksen'. In genoemd artikel wordt beweerd dat de 'hot hand' bestaat en de schrijvers attaqueren het artikel *The Hot Hand in Basketball: On the Misconception of Random Sequences* van Thomas Gilovich, *et al.* uit 1985 waarin juist het tegendeel geconcludeerd wordt. In al deze artikelen wordt de nulhypothese geformuleerd dat de data-reeks van schoten een serie realisaties is van onafhankelijke Bernoulli-proeven, met een *op deze reeks zelf gebaseerde succeskans* [op zijn minst merkwaardig]. Vervolgens wordt deze nulhypothese dan wel of niet weerlegd door de reeks treffers/missers te onderzoeken *na* een reeks van een drietal treffers of een reeks van een drietal missers. Voor de liefhebbers verwijs ik naar 'google'. Ik acht me als eenvoudige kansrekenaar niet bevoegd over deze literatuur een oordeel te vellen, maar door het al of niet bestaan van de 'hot hand' louter te onderzoeken via het analyseren van reeksen treffers/missers worden de complexiteit en de diepe krochten van de menselijke psyche naar mijn bescheiden mening wel schromelijk onderschat. Het brein wordt bij deze statistische aanpak benaderd als een 'black box' en dat voor veel cognitief-psychologen deze methodologie dominant is geeft te denken. Deze reductie tot statistisch toetsen van uitkomstenreeksen komt op mij over als het krassen op een rots: wat het resultaat van statistische toetsing ook zal opleveren, een dieper inzicht in de relatie tussen eerder succes en later succes levert het niet op.

Het aardige van het artikel van Joshua Miller en Adam Sanjurjo is dat zij wijzen op een valkuil waar de dames en heren psychologen ook nog in kunnen vallen bij het toetsen van de genoemde nulhypothese. Zij wijzen op het verrassende resultaat dat in een serie van n onafhankelijke Bernoulli-proeven met succeskans p na een succes met een kans kleiner dan p direct weer een succes optreedt, en zelfs dat deze kans kleiner wordt naarmate er een langere reeks van ononderbroken successen is opgetreden. Dit klinkt paradoxaal, maar is bij nader inzien louter

een gevolg van het feit dat zij een conditionele kans uitrekenen, namelijk de kans op succes direct na een succesreeks van een gegeven lengte k , *gegeven dat in de reeks van n proeven in de eerste $n-1$ proeven minstens één succesreeks van lengte k is opgetreden*. Toch is dit een aardig resultaat dat volledig binnen de kansrekening netjes te formuleren is en daarom valt er ook echt iets te bewijzen!

Hieronder zullen we vragen een formule af te leiden voor de kans op een succes onmiddellijk na een eerder succes in een serie Bernoulli-proeven van *gegeven vaste lengte* [dit is essentieel!]. Voor de kans op succes direct na een ononderbroken reeks van k successen [$k = 2, 3, \dots$] hebben we geen formule kunnen vinden. Uit het bewijs [dus het geval $k = 1$] zal wel blijken waarom dat ook niet zo eenvoudig zal zijn. Na de opgave zullen we kort terugkomen op mogelijke consequenties voor de Gambler's Fallacy. De gevolgen voor het onderzoek naar De Hot Hand Fallacy laten zich raden, maar die liggen op het bordje van de cognitief-psychologen en daaraan wil ik als niet-statisticus mijn handen niet verder branden.

Bovengenoemde formule kun je via enige simpele combinatorische kansrekening afleiden [Miller en Sanjurjo geven een ander, minder doorzichtig bewijs]. We herschrijven de vraagstelling voor de duidelijkheid als een helder geformuleerde tentamenopgave, zodat over de vraag zelf geen enkel misverstand kan optreden:

Zij gegeven een p -munt, dus een munt met kans op 'kruis' gelijk aan p ($0 < p < 1$). We vragen Arjen met deze munt een serie worpen uit te voeren. Telkens als hij 'kruis' gegooit heeft vragen we hem de uitkomst van de direct daarop volgende worp te noteren op een stuk papier. In geval die worp weer uitkomst 'kruis' geeft, dan noteert hij ook de uitkomst van de daarop volgende worp, et cetera. Op een gegeven moment heeft Arjen er genoeg van en hij stopt met werpen. Hij vult nu een vaas met zoveel rode ballen als hij uitkomst 'kruis' genoteerd heeft en zoveel witte ballen als hij uitkomst 'munt' genoteerd heeft. Vervolgens trekt Henk, die de uitkomsten van Arjen niet gezien heeft, blindelings een bal uit de vaas.

De vraag is de kans te berekenen dat Henk een rode bal trekt, of anders gezegd de kans dat Arjen na een

uitkomst 'kruis' direct een nieuwe uitkomst 'kruis' gegooit heeft. Merk op dat voor Arjen de kans dat Henk een rode bal trekt een andere zal zijn dan voor Henk, omdat Arjen weet hoeveel rode en witte ballen hij in de vaas gestopt heeft! Kansen worden (mede) bepaald door de kennis die je al verworven hebt over de uitkomsten van een toevalsexperiment, en in dit geval weet Arjen veel meer dan Henk, die voordat hij een bal trekt nog helemaal niets weet, zelfs niet op grond waarvan Arjen besloten heeft met werpen te stoppen! Henk kent uiteraard wel alle details van de *opzet* van het toevalsexperiment, dus hij kent de kans op 'kruis' van de munt en hij weet precies hoe Arjen de vaas met ballen heeft gevuld.

De vraag is zoals hierboven geformuleerd niet te beantwoorden omdat we niet hebben aangegeven op grond van welk criterium Arjen gestopt is met zijn serie worpen. Met andere woorden, de beschrijving van het *toevalsexperiment* is nog onvolledig, omdat er nog geen kansmodel gedicteerd wordt. Het is mogelijk dat Arjen stopt zodra hij bijvoorbeeld n ballen in de vaas gestopt heeft, maar het is ook mogelijk dat hij stopt na *in totaal* n worpen te hebben uitgevoerd!

Daarom formuleren we nu twee opgaven.

1. Bereken de kans dat Henk een rode bal trekt uit de vaas in geval Arjen doorgaat met werpen totdat hij n uitkomsten genoteerd heeft en dus n ballen in de vaas kan stoppen.
2. Bereken de kans dat Henk een rode bal trekt in geval Arjen besluit te stoppen nadat hij in totaal $n \geq 3$ worpen heeft uitgevoerd *gegeven dat de vaas na n worpen niet leeg is* [dat is de situatie die Miller en Sanjurjo bespreken]. In deze opgave wordt je gevraagd een conditionele kans te berekenen. Om je enigszins op weg te helpen geven we het antwoord:

$IP(\text{Henk trekt een rode bal} \mid \text{Vaas is niet leeg}) =$

$$p + (1-p) \left[\frac{1}{n-1} - \frac{p(1-p)^{n-2}}{1-(1-p)^{n-1}} \right].$$

Toon ook aan dat deze kans kleiner is dan p .

Aanwijzing: Splits uit naar het aantal keren 'kruis' én het aantal combinaties 'kruis-munt' in de serie uitkomsten van Arjen.

Het antwoord van de tweede opgave lijkt erop te duiden dat de kans dat Arjen na een uitkomst 'kruis' opnieuw 'kruis' gooit kleiner is dan men zou verwachten. Dit paradoxale resultaat is volledig toe te schrijven aan het stopcriterium dat Arjen in totaal n worpen uitvoert, en aan het feit dat we de conditionele kans hebben uitgerekend op een rode bal *gegeven dat de vaas niet leeg is*. Merk op dat voor $n \rightarrow \infty$ de kans dat Henk een witte bal trekt convergeert naar $1-p$, gelukkig maar!

Welk licht laat deze formule nu schijnen op de Gambler's Fallacy? Een naïeve gokker zou namelijk met dit resultaat voorhanden overtuigd kunnen raken dat zijn geloof een kern van waarheid bevat, tenslotte is de kans op 'zwart' na 'rood' groter dan $1/2$ zal hij concluderen. Uit het bovenstaande moge duidelijk geworden zijn dat dit alleen het geval is als de croupier van te voren een vast aantal keer het balletje laat rollen, bijvoorbeeld 4 keer, dit procedé vele malen herhaalt, totdat hij zeg 1000 keer een serie van vier rood/zwart resultaten heeft gerealiseerd waarbij in de eerste drie rondes minstens één keer het resultaat 'rood' is opgetreden, en vervolgens de spelers, bijvoorbeeld via een van de resultaten gemaakte video, confronteert met deze door hem gerealiseerde reeks van 4000 rood/zwart resultaten. Als een speler nu consequent op 'zwart' inzet na elke ronde met resultaat 'rood' [behalve als het rangnummer van de ronde een viervoud is!] dan zullen zij ongeveer 25 van de 42 keer winnen bij inzetten op 'zwart' direct na een uitkomst 'rood', zoals uit de formule blijkt voor $n = 4$ en $p = 1/2$ [we verwaarlozen weer de nul]. Deze winst-strategie is louter een gevolg van het weglaten van de viertallen die beginnen met drie keer 'zwart'. Er is dus sprake van manipulatie, ook al zal een gokker die met deze reeks van 4000 zwart/rood uitkomsten geconfronteerd zal worden dit niet meteen opmerken.

Samenvattend kunnen we concluderen dat de afgeleide formule op het eerste gezicht enige verwarring scheidt, omdat de Bernoulli-proeven onafhankelijk zijn, maar bij nader inzien realiseren we ons met een conditionele kans van doen te hebben: series met louter uitkomsten 'munt' in de eerste $n-1$ worpen worden genegeerd, en dat daarmee ook de kans op 'kruis' direct na een eerdere

uitkomst 'kruis' niet gelijk aan p is, is niet verrassend omdat je iets anders berekent dan simpelweg de kans dat de volgende worp 'kruis' oplevert. Dat deze conditionele kans bovendien kleiner is dan p is interessant, maar voor de kansrekenaar met een goed ontwikkelde intuïtie misschien wel triviaal, zoals veel zaken triviaal zijn voor getalenteerde wiskundigen, die voor gewone stervelingen niet onmiddellijk duidelijk zijn. Gewone stervelingen rest niets anders dan stug aan het rekenen te slaan om de valkuilen te vermijden die zelfs in zo'n eenvoudig toevalsexperiment als een serie onafhankelijke Bernoulli-proeven verborgen blijken. En als zo'n elementair toevalsexperiment al vol valkuilen zit, wat moeten we dan denken van alle ons door de media gebrachte wetenschappelijke resultaten? Het minste wat we kunnen zeggen is dat enige scepsis geen kwaad kan, zeker niet bij het vernemen van [sociaal-]wetenschappelijke resultaten met politieke implicaties.

Iedereen wordt weer uitgenodigd de oplossing van de twee opgaven naar me toe te sturen via e-mail [r.d.nobel@vu.nl]. De deadline voor het inzenden ligt op 30 september. Traditiegetrouw worden correcte inzendingen weer beloond met een boek!

Rein Nobel



Alle ruimte om het beste uit jezelf te halen

Ook dat is #werkenbijpwc

PwC is een netwerk van firma's in 157 landen met meer dan 223.000 mensen, waarvan ruim 4.700 in Nederland. Wij zien het als onze taak om kwaliteit te leveren op het gebied van assurance-, belasting- en adviesdiensten.

Je wilt consultant worden en bent weliswaar bijna klaar met je studie, maar nog lang niet uitgeleerd. Dat komt goed uit. PwC heeft volop dynamische startersfuncties binnen Advisory. Je hebt toegang tot een wereldwijd netwerk en deelt je kennis, ideeën en vragen met de beste professionals uit je vakgebied. Je werkt in multidisciplinaire teams met collega's die elkaar inspireren. Ook daarom vind je bij PwC veel mensen met verschillende opleidingen en culturele achtergronden.

Een wereld vol kansen

Jouw carrière is wat het is: van jou. Jij bepaalt, wij bieden je de kansen. Op uitdagend werk dat er toe doet. Om het maximale uit jezelf te halen. Om mee te bouwen aan vertrouwen in de maatschappij en een bijdrage te leveren aan het oplossen van belangrijke problemen. Je wordt intensief begeleid en er is volop ruimte voor jouw ambitie en persoonlijke keuzes. We bieden afwisselende werkzaamheden bij vooraanstaande klanten.

Verras jezelf en je klanten

Om onze klanten te verrassen, blijven we zoeken naar nieuwe invalshoeken. Want alleen door onszelf uit te dagen, komen we tot de beste oplossingen op het gebied van Deals en Consulting. Daarom helpen we je graag verder bij het ontdekken van je sterke punten. Met coaching, inspirerende collega's, opleidingen en de mogelijkheid te switchen tussen sectoren en vestigingen. Of je gaat een korte of langere periode naar het buitenland. Als het om jouw ontwikkeling gaat, kent PwC geen grenzen. Kijk voor evenementen, stages en vacatures op www.werkenbijpwc.nl.



The opportunity of a lifetime

Extra

Het studentenleven bestaat uit meer dan alleen studeren. Extracurriculaire activiteiten leveren een grote bijdrage aan de vorming van een student. Denk aan de ontwikkeling van je communicatieve vaardigheden, organisatorische skills en andere praktische tools die je niet bij de studie meekrijgt. In deze rubriek vertellen enkele medestudenten wat zij naast de studie Econometrie & OR doen en wat hun ervaringen zijn. Misschien is hun verhaal jouw nieuwe motivatie om ook zelf eens verder te kijken?

Marc Stougie



Marc werkt naast zijn studie econometrie bij ABN Amro.

Mijn naam is Marc Stougie en ik ben student Econometrie en Operationele Research. Dit is alweer mijn derde, en als het goed is laatste, bachelor jaar. Door het nieuwe curriculum binnen de opleiding is er meer ruimte om je minor zelf samen te stellen. Een optie hierbij was om een paar vakken te laten vallen en in plaats hiervan een stage bij een bedrijf te doen.

Afgelopen zomer was ik gelijk geïnteresseerd in deze optie. Het leek mij nuttig om in contact te komen met bedrijven en het sollicitatieproces mee te maken. Met name door de formele activiteiten van Kraket heb ik gedurende mijn tweede jaar al een beetje kunnen oriënteren wat voor een soort bedrijven leuk leken en welke gelijk afvielen. Ik ben aan het einde van de zomervakantie steeds actiever op zoek geweest naar een werkplek. Ik wilde namelijk het liefst wel een vaste werkplek en niet alleen een stageplek. Mijn interessegebied is breed en ik vind dan ook veel dingen leuk. Binnen een grote bank gebeurt er van alles en daarom dacht ik dat ABN Amro een mooie eerste werkgever voor mij zou zijn. Ook de verbinding universiteit - ABN Amro, speelde een rol in de keuze van bank (ongeveer zeven minuten lopen).

“De gemiddelde leeftijd is ongeveer 33 en het opleidingsniveau ligt vrij hoog: 75% van de afdeling heeft een PhD op zak.”

Inmiddels werk ik al bijna een jaar bij ABN Amro en het bevalt mij nog steeds erg goed. Ik functioneer binnen het Model Validation team. MV bestaat uit ongeveer twintig collega's die zich bezighouden met alle modellen binnen ABN Amro. Alle modellen die gebruikt worden of gebruikt moeten gaan worden komen langs. Wij zijn een soort van onafhankelijk team dat de modellen beoordeelt. De gemiddelde leeftijd is ongeveer 33 en het opleidingsniveau ligt vrij hoog: 75% van de afdeling heeft een PhD op zak.

ABN Amro is een hele fijne werkgever voor mij die er ook rekening mee houdt dat ik nog studeer. Ik heb een soort nul-urencontract, maar de afspraak is dat ik gemiddeld genomen ongeveer twintig uur per week kom werken. Dit geeft mij de mogelijkheid om minder te werken in de week voor de tentamens en weer meer te werken na de tentamenweek. Ik heb een contract tot aan de zomer en ik houd alle opties nog open. Het afgelopen jaar heb ik veel geleerd, maar daarnaast was deze combinatie van werken en studie ook erg leuk. Werken naast je studie; zeker een aanrader!

Gillis Kruisselbrink



Gillis is derdejaars econometrist en zit daarnaast in het bestuur van Extensus.

Wanneer ik mensen vertel dat ik Treasurer (penningmeester) ben bij Extensus, moet ik altijd eerst uitleggen wat Extensus precies is. Het wordt bijna een standaard riedeltje: Extensus is de studievereniging voor alle honoursstudenten in Amsterdam. Dus niet alleen honoursstudenten van de VU, maar ook van de UvA en het AUC. En dus ook studenten van allerlei verschillende studies; van wiskunde of natuurkunde tot rechten of culturele antropologie. En natuurlijk (inmiddels een best grote groep) econometristen. Dit vind ik een van de grote charmes van de studievereniging en een groot verschil met Kraket. Je merkt dat studenten met een andere achtergrond soms toch heel anders tegen dingen aan kunnen kijken. Extensus organiseert onder andere gastcolleges, inhousedagen en elk jaar een studytrip, dit jaar zijn we naar Dublin geweest.

“Ik wilde nog graag ergens een keer bestuursjaar doen en toen Extensus op zoek ging naar een nieuw bestuur, was één en één al snel twee.”

Om lid te kunnen zijn van Extensus moet je dus het Honoursprogramma volgen. Ik ben daar tijdens mijn eerste jaar Econometrie aan begonnen. Het leek me leuk mijn horizon te verbreden voorbij de vele modellen en cijfers van Econometrie. Dit kon door het Honoursprogramma: 5 extra vakken, waarvan er drie eigenlijk niets met je studie te maken hoeven hebben. Zo heb ik onder andere een vak over de rechtelijke gevolgen van robotisering gevolgd:

Wie is er verantwoordelijk als een zelfrijdende auto een ongeluk veroorzaakt? Dat is ingewikkelder dan het lijkt!

Op het moment dat ik daar dus mee startte had ik nog nooit van Extensus gehoord. Mijn eerste contact met Extensus was toen ik met vier andere econometristen een mailtje kreeg over een studytrip met Extensus naar Boedapest. Daar zijn we meegegaan, en dat was erg leuk, en nog vele leuke activiteiten volgden. Ik wilde nog graag ergens een keer bestuursjaar doen en toen Extensus dus op zoek ging naar een nieuw bestuur, was één en één al snel twee. Ook kon ik nog gewoon al mijn vakken blijven volgen, aangezien Extensus geen grote vereniging is, en dit dus een parttime bestuur is.

Tot nu toe bevalt het bestuursjaar me erg goed. Nauw samenwerken in een bestuur is erg leerzaam, en het is erg leuk activiteiten zelf te (helpen) organiseren. Dan heb je opeens heel andere uitdagingen dan die tijdens je studie. Het is soms wat stressvoller, maar het geeft een erg goed gevoel als het resultaat er dan staat. Bovendien houd ik ook nog genoeg tijd over voor andere dingen. Zo zullen sommigen mij eerder kennen als student-assistent, en ik werk nog zo'n 8 uur per week bij de Albert Heijn. Ook kan ik mij voorbereiden voor volgend jaar, als ik een jaar in Manchester ga studeren. Al met al is een bestuursjaar, bij welke studievereniging dan ook, een ervaring die ik iedereen kan aanraden!

Rachid Saaliti



Rachid heeft in zijn tussenjaar een kijkje genomen in het bedrijfsleven om praktijkervaring op te doen.

Na 4 jaar studeren, 207 ECTS, een aantal commissies en een

bestuursjaar bij Kraket had ik even geen behoefte aan mijn dagen spenderen op de universiteit. Aangezien er in het oude curriculum geen ruimte was om werk-/stage ervaring op te doen en ik daar wel behoefte aan had, besloot ik een tussenjaar te nemen. Door een aantal maanden fulltime te werken, hoopte ik een beter beeld te krijgen van wat we nou kunnen na het afronden van een bachelor econometrie en welke richting mij interessant lijkt. Door mijn functie als extern coördinator bij Kraket had ik al met een groot aantal bedrijven kennis gemaakt.

Eén van die bedrijven is RiskQuest, een relatief klein consultancy bureau in Amsterdam, met een kantoor op de Herengracht, gespecialiseerd op het gebied van risico management in de financiële sector. Er werken ongeveer 20 mensen bij RiskQuest met een kwantitatieve achtergrond. De opdrachten die RiskQuest uitvoert zijn onder andere het bouwen (in talen als Matlab en R) en valideren van kwantitatieve modellen, zoals modellen die gebruikt worden bij het meten en managen van risico's door hypotheekverstrekkers. Daarmee kan bepaald worden hoeveel kapitaal er gereserveerd moet worden om het risico op wanbetalingen bij een hypotheek te dekken. Dit soort modellen hebben een econometrische basis en het leek mij dan ook interessant om bezig te blijven met econometrie en dit in praktijk te kunnen brengen.

Van september tot en met maart heb ik fulltime als analist/werkstudent bij RiskQuest gewerkt. De werkzaamheden van een werkstudent bestaan onder andere uit het ondersteunen van de consultants, zoals het analyseren van data en de analyses verwerken in een presentatie. Ook zijn er werkstudenten die 2-3 dagen per week bij de klant zitten en daar meewerken aan het bouwen van de modellen. Ik begon toevallig tegelijkertijd met een

nieuw initiatief (onder de naam DataQuest) om (big) data gerelateerde vraagstukken buiten de financiële wereld te beantwoorden. Data en econometrie vormen zowel binnen als buiten de financiële wereld de basis voor het oplossen van kwantitatieve vraagstukken. De methoden en modellen die hiervoor gebruikt worden lijken dan ook veel op elkaar en met de kwantitatieve achtergrond van de werknemers was de kennis om (big) data gerelateerde opdrachten uit te voeren ook in huis.

Nadat ik aan had gegeven dat ik daar geïnteresseerd in was, mocht ik meewerken aan een opdracht voor een groot Nederlands kledingmerk met filialen over de hele wereld. Dit bedrijf had de beschikking over een enorme hoeveelheid historische verkoopdata. Hiermee wilden ze tot een betere prijsstrategie komen voor de sale perioden. Simpel gezegd wilden ze voor alle producten en weken weten hoeveel procent korting er gegeven moest worden om de winst te maximaliseren. Samen met een collega consultant heb ik bekeken hoe we dit vraagstuk kunnen oplossen. Dit hebben we gedaan door een script te schrijven waarmee we het verband tussen de vraag en de prijs kunnen beschrijven. Naast het programmeren bestond het project uit tweewekelijkse overleggen met de klant. Ik mocht hierdoor al meteen mee naar de klant, wat een leerzame ervaring was. Hierin presenteerden wij onze bevindingen en discussieerden we over onze uitkomsten. Naast de programmeer- en modelleer ervaring leer je hierdoor ook hoe je de resultaten het best kan presenteren. Als consultant heb je namelijk ook vaak te maken met mensen die geen verstand van- of geen behoefte aan de technische details van een wiskundig model hebben. Daarom hebben we bijvoorbeeld een dashboard gebouwd waarmee ze zelf een indruk konden krijgen wat de invloed van bepaalde variabelen op het optimale kortingspercentage is.

Kortom, een hele leerzame ervaring. Op inhoudelijk gebied, maar ook wat soft skills betreft. Daarnaast heb ik ook een beter beeld gekregen van wat er in de praktijk speelt. Hierdoor heb ik een beter idee gekregen welke keuzevakken tijdens de master interessant zijn. Ik zou iedereen dan ook aanraden om voor het eind van de studie alvast te proeven aan het bedrijfsleven.

Matthias Versfeld



Naast zijn studie econometrie heeft Matthias stage gelopen bij de KNVB.

Hoi allemaal! Voor diegenen die me niet kennen: Ik ben Matthias, 20 jaar en zit momenteel in het derde jaar van mijn bachelor (EOR uiteraard). Mijn hele leven heb ik al een passie voor voetbal; om het spelletje te spelen, te bekijken en allemaal ontwikkelingen omtrent de grote jongens te volgen. Jammergenoeg ben ik er niet goed genoeg in om prof te worden, dus heb ik een andere manier gevonden om toch bezig te zijn met voetbal en studie: stagelopen bij de KNVB (Koninklijke Nederlandse Voetbal Bond)!

Rond februari zag ik op LinkedIn een berichtje over een Hackaton op het gebied van data analyse van de KNVB in samenwerking met SAS, en zonder te twijfelen schreef ik me in (de twijfel was helemaal weg omdat we na de Hackaton met z'n allen naar de bekerfinale zouden gaan in De Kuip, toch wel een leuke bijkomstigheid!). Bij de uiteindelijke Hackaton won mijn groepje, door in de ogen van de jury (Directieleden van SAS, de KNVB en Pieter Zwart) de beste presentatie neer te zetten met de beste bevindingen. Als prijs mochten alle vier de groepsleden als VIP naar de oefeninterland Nederland – België. Geloof me, dat was zeker weten lachen.

“Begin juli zat ik al bij SAS om verscheidene cursussen te volgen voor de software die gebruikt kan worden.”

Na afloop van de Hackaton kon je aangeven of je geïnteresseerd was in een stageplek bij de KNVB. De hoofdreden van het evenement was

namelijk studenten 'scouten' voor het invullen van stageplekken. Mij leek het wel wat, want in de minor heb je de mogelijkheid om 12 ECTS te verdienen met een stage. Zo is het balletje gaan rollen. Ik werd gebeld en uitgenodigd om er stage te lopen. Begin juli zat ik al bij SAS om verscheidene cursussen te volgen voor de software die gebruikt kan worden. Bij de KNVB heb ik me beziggehouden met het doen van analyses op hun database. Hier staat alles in, van de wedstrijden van de meisjes onder 9 tot en met het aantal verkochte tickets bij de Primera voor interlands van het Nederlands elftal. Zo heb ik bijvoorbeeld dashboards gemaakt voor verenigingsadviseurs (amateurclubs) en heb ik geprobeerd de uitslagen van wedstrijden te voorspellen, zodat we de competitie indelingen zo leuk mogelijk kunnen maken. We werkten in een team van 4: De manager, de Data Analist en twee stagiair Data Analisten. Omdat het zo'n klein team was, kreeg je veel vrijheid om te doen wat je leuk vindt. Bovendien bestond het team nog niet zo lang, wat betekende dat er in het halve jaartje dat ik er was gigantische stappen werden gezet!

Ik zou iedereen aanraden om iets werkgerelateerd naast zijn of haar studie te doen, je leert zo ongelofelijk veel op het gebied van soft skills, plus je ziet eindelijk hoe je opgedane kennis al van pas kan komen in de praktijk.



Jan Magnus

is buitengewoon hoogleraar aan de VU en doceert een deel van het tweedejaarsvak 'Econometrics I'.

Docent Vertelt Jan Magnus

In deze editie van de SECTOR is voor de rubriek “Docent Vertelt” Prof. Jan Magnus geïnterviewd. Hoe is hij bij de VU terecht gekomen en hoe ziet zijn academische loopbaan eruit?

Kunt u uzelf introduceren?

Op mijn achttiende ben ik begonnen met een studie muziek en literatuur aan het Albert-Schweitzercollege in Zwitserland. Daar heb ik ook cello-les gehad van Vilmos Palotai, oud-lid van het Hongaars Strijkkwartet. Omdat ik iets met sociale wetenschappen wilde doen én geïnteresseerd was in wiskunde, ben ik na een jaar toch econometrie gaan studeren. Het is niet zo dat ik al vroeg wist dat ik iets met wiskunde zou gaan doen. Sommige mensen weten al vanaf kinds af aan dat ze bijvoorbeeld arts of wiskundige willen worden, maar bij mij is het nooit zo duidelijk geweest; ik heb altijd in veel dingen belangstelling gehad. Als ik geen econometrist zou zijn, maar bioloog of historicus, dan was ik ook gelukkig geweest. Wat ik wel erg

prettig vind, is de academische omgeving en het academische bestaan.

Waarom bent u econometrie aan de UvA gaan studeren?

Destijds was econometrie een behoorlijk nieuw vak en werd eigenlijk alleen in Amsterdam en Rotterdam gegeven. De voornaamste reden om voor de UvA te kiezen was dat de wiskunde- en statistiekvakken aan de wiskundefaculteit gedoceerd werden, in plaats van aan de economische faculteit zoals in Rotterdam. Op de UvA heb ik ook filosofie erbij gestudeerd maar uiteindelijk beviel me dat toch minder, omdat de richting van de filosofie aan de UvA nogal Duits georiënteerd was. Mijn interesse lag meer in de richting van de Engelse filosofie.

Wat heeft u na uw studie gedaan?

Na mijn doctoraalexamen ben ik twee jaar in Vancouver geweest. Ik had toen al een aantal artikelen gepubliceerd en daar heb ik toen een baan aangeboden gekregen. Vanuit Vancouver kreeg ik een aanbod om naar de London School of Economics te gaan. Dat was geweldig en ik heb daar een prachtige tijd gehad. Op een gegeven moment wilden we graag terug naar Nederland, onder andere vanwege de school van mijn kinderen. Toen ben ik ingegaan op een aanbod van Tilburg, die hoogleraren vanuit het buitenland aan wilden trekken voor onderzoek. Dat was weinig onderwijs en veel tijd voor research. Dat heb ik gedaan tot aan mijn pensioen.

Ik ben daarna ingegaan op een aanbod van de VU om een college econometrie (Econometrics I in het tweede jaar van de bachelor) te geven. Dat leek me een leuke uitdaging, want ik had jarenlang geen les gegeven aan bachelor-studenten.

Kunt u iets vertellen over uw werkzaamheden en interessegebieden?

Qua onderzoek heb ik aan allemaal verschillende dingen gewerkt. Ik ben vrij theoretisch begonnen in mijn loopbaan. Ik heb een boek geschreven over matrix differentiaalrekening (met Heinz Neudecker), wat nog steeds veel gebruikt wordt. Dat was nuttig want je komt voortdurend functies van matrices tegen. Het schrijven van een boek kost echter veel tijd. Ik ben er dan ook ongeveer tien jaar mee bezig geweest.

Waarom heeft u besloten zelf een boek te schrijven?

De markt voor de boeken waar ik naar zocht voor het vak Econometrics I, waarbij voor mij matrix-algebra belangrijk is, bleek volledig uitgedroogd. Daarom heb ik zelf het boek *Introduction to the Theory of Econometrics* geschreven. Er zijn namelijk veel econometrie boeken die heel weinig met matrices doen, zeker voor bachelor-studenten. Dat zijn meestal boeken die je kan omschrijven als 'econometrie voor economen'. Qua matrix-algebra gaan deze boeken niet zo heel ver. Als je in het eerste jaar van de studie lineaire algebra leert, dan is dat een ontzettend groot hulpmiddel wat je tot je beschikking hebt.

Wij hebben gelezen dat u onderzoek gedaan heeft naar sportstatistieken. Kunt u daar iets over vertellen?

Op een gegeven moment ben ik geïnteresseerd geraakt in sport, vooral tennis, doordat ik in Engeland naar televisie keek en ik de commentatoren nogal opmerkelijke uitspraken hoorde doen. Het is bij tennis moeilijk om interessant commentaar te geven, omdat het vergeleken met andere sporten minder commentaar behoeft. Hierdoor vallen tenniscommentatoren vaak terug op clichématige uitspraken als "Als je iemand gebroken hebt, is er een grotere kans om teruggebroken te worden" en "De zevende game is een uitzonderlijk belangrijke game". Ik heb op basis daarvan een lijstje gemaakt met 24 van dit soort uitspraken. Hiermee ben ik in Engeland naar IBM gegaan waar een medewerker mij de relevante data gaf en met behulp van deze data ben ik die uitspraken nagegaan. Daar heb ik later veel papers over geschreven en uiteindelijk zelfs een boek: *Analyzing Wimbledon*. Dat was eigenlijk het resultaat van een uit de hand gelopen hobby. Een van de mooie dingen aan sportdata is dat de data heel schoon en transparant zijn. Bij economische data is er altijd veel ruis, maar in de puntentelling van een tenniswedstrijd zitten nauwelijks fouten.

“Dit was het eerste onderzoek waarbij ik voor het slapen dacht: dit resultaat is erg ongewenst, klopt het wel?”

De meeste van de 24 uitspraken die ik had opgeschreven, hebben wij getoetst en ontkracht. Een bekende hypothese in tennis is "Als je begint met serveren in een set, levert dat een voordeel op". De redenering is dat de achterstand demotiverend werkt voor de tegenstander. Om dit te toetsen hebben we toentertijd bij zo'n 4000 sets bekeken wie er begint met serveren en wie de set wint. Wat bleek was dat degene die ontvangt vaker wint dan degene die serveert, behalve in de eerste set. Dit

vonden we gek, maar de verklaring hiervan is dat de betere speler, bij verwachting, de eerste set wint. De waarschijnlijke situatie is dat de betere speler als laatste serveert in de set, en dan de set wint. Dus de minder goede speler begint met serveren in de volgende set en omdat hij de minder goede speler is, verliest hij de volgende set. De sets zijn dus niet onafhankelijk van elkaar. Dat verklaart ook meteen waarom de eerste set steeds een uitzondering vormt.

Een ander onderwerp dat we onderzocht hebben is de strategie van de service. In tennis heb je twee service-mogelijkheden. Hoe moeilijk moet een tennisser de eerste service maken en hoe moeilijk de tweede zodat de kans op punten geoptimaliseerd wordt? Je kunt dan kijken in hoeverre spelers daarvan afwijken en of een speler, door een andere balans te kiezen tussen moeilijkheid van eerste en tweede service, de kans op winst van de wedstrijd kan verhogen.

Welk onderzoek dat u heeft gedaan sprak u het meeste aan?

Een interessant onderzoek dat ik gedaan heb gaat over CO₂-uitstoot en aerosols (fijnstof). Het bekende verhaal is dat de CO₂-uitstoot in de atmosfeer komt, waardoor er een vervuilde laag in de stratosfeer ontstaat. Hierdoor kan de warmte op de aarde moeilijker weg en ontstaat er een effect van opwarming. Dit is een wereldwijd effect; uitstoot in China beïnvloedt ons klimaat ook.

De aerosol-uitstoot, daarentegen, bestaat uit kleine korreltjes die mettertijd in de lucht samenklonteren tot een soort schild en zo het zonlicht reflecteren. Dit zorgt lokaal voor een koelend effect. In gebieden waar veel vervuiling is, daalt daardoor de temperatuur. In tegenstelling tot CO₂ is het effect van de uitstoot van aerosols lokaal, dus meer fijnstof in China heeft geen effect in Nederland.

Wat was de inhoud en uitkomst van uw onderzoek daarnaar?

Om mogelijke toekomstscenario's door te rekenen, hebben we econometrische analyses uitgevoerd op een hele grote dataset, die collega's in Zürich verzameld hebben. We hebben onderzocht wat er zou gebeuren als de maatregelen tegen aerosols, die dus lokaal aan te pakken zijn, doorgevoerd zouden

worden. Als je de aerosols wilt aanpakken, moet je dus extra maatregelen tegen CO₂ nemen om voor het wegvalen van die koelende aerosol-laag te compenseren. Dit effect moet je meenemen als je prognoses doet, en dat hebben we dus uitgerekend. Dat gaf vrij alarmerende uitkomsten. Bij sommige onderzoeken maakt een klein verschil in resultaat niet heel veel uit, maar omdat dit wereldwijd zo'n belangrijk onderwerp is, hebben we de resultaten een aantal keer opnieuw berekend. Dit was ook het eerste onderzoek waarbij ik voor het slapen dacht: "Dit resultaat is erg ongewenst, klopt het wel?", juist omdat het over zo'n belangrijk probleem gaat. Uiteindelijk bleken onze resultaten inderdaad te kloppen.

De afspraken in Parijs waren voor veel mensen in de klimaatwetenschap een grote positieve verrassing. Plotseling waren er daadwerkelijke internationale afspraken, wat erg belangrijk is om klimaatverandering tegen te gaan. Hopelijk blijven deze afspraken staan nu er een nieuwe regering in Washington is die de invloed van de mens op de opwarming van de aarde ontkent.

Sometimes
you have
to look
at things
from a
different
perspective



Scriptieproblematiek

Een academische studie wordt afgerond met een scriptie waarin de student laat zien hoe hij of zij (voortaan steeds hij) alle geleerde kennis zelfstandig kan toepassen of verwerken in een theoretisch of praktisch probleem de EOR betreffend. In deze ene zin blijkt al de vele mogelijke varianten voor een scriptie. Het begint er al mee dat de EOR-studie opgesplitst is in twee (sub-)studies, namelijk een bachelor- en een masterfase, elk met een afrondende scriptie. We spreken dus over een bachelorscriptie en over een masterscriptie, waarbij opgetekend dient te worden dat wegens taalkundige consistentie “scriptie” is vervangen door “thesis”.

En zo staan deze ook in VuNet, Blackboard, studiegids, en andere media, en bovendien staan ze daarin vermeld als een van de verplichte vakken in het EOR-curriculum. In dit artikel zal ik een aantal zaken toelichten over de EOR-scripties. Daarnaast zal ik ook proberen te antwoorden op de twaalf meest prangende vragen die studenten stellen over scripties:

1. Wanneer mag ik beginnen met de scriptie?
2. Hoe vind ik een geschikt onderwerp voor mijn scriptie?
3. Hoe vind ik een scriptiebegeleider?
4. Kan ik een scriptie aan een stage koppelen?
5. Hoe kan ik het onderwerp inperken tot een onderzoeksvraag?
6. Hoe kom ik tot een plan van aanpak?
7. Waar vind ik goede bronnen, data, en informatie bij mijn onderwerp?
8. Moet ik programmeren?
9. Schrijf ik de scriptie in Nederlands of Engels?
10. Is het verplicht om de tekst in LaTeX te verwerken?
11. Hoeveel tijd krijg ik om de scriptie af te ronden?
12. Hoe wordt de scriptie beoordeeld?

Waarom is een scriptie verplicht?

Het formele antwoord is dat tegenwoordig vele leerdoelen en eindkwalificaties ten grondslag liggen aan een studie (of opleiding), bijvoorbeeld de afgestudeerde van onze EOR-studie heeft “een grondige kennis van de drie kerngebieden van de EOR”, en “is in staat om (bedrijfs)economische problemen te modelleren en met methoden uit de wiskunde en statistiek te analyseren”. Zo zijn er nog een heel stel. Ze vormen een dichtgetimmerd verhaal dat wordt weergegeven in een diagram van matrijzen en matrices met kolommen en rijen die zo met elkaar zijn verweven dat een normaal mens al snel de weg erin kwijt is. Maar als je goed kijkt zie je dat al die einddoelen en eindtermen samenkomen in het scriptie-onderdeel! Met andere woorden, wij (EOR-docenten) hebben voldoende aan het oordeel van de scriptie om te toetsen of wij de studenten in drie jaar (bachelorfase) of in één jaar (masterfase) hebben weten te onderrichten volgens de eisen en doelen die wij onszelf hebben opgelegd.

Dat is natuurlijk van de gekke, het gaat niet om de docenten. Het gaat erom dat de student uiteindelijk iets geleerd heeft waarmee hij de markt op kan

gaan en zich kan ontwikkelen tot EOR-professional. Of misschien wil hij zich verder ontwikkelen tot een EOR-wetenschapper. De EOR-studie is zo ingericht dat de student de daartoe benodigde informatie, theoretische kennis en praktische handvaardigheid krijgt aangereikt middels vele vakken, practica en noem maar op, en dat de student laat zien dat hij die kennis en kunde kan reproduceren en etaleren tijdens huiswerkopdrachten en tentamens. Hoewel je met zo'n achtergrond ongetwijfeld een goede professional kunt worden, vindt "men" dat een academicus meer in zijn mars heeft en dat moet laten zien door eens zelfstandig aan een probleem of aan een onderwerp te werken. Je kunt je dan niet verschuilen of meeliften met groepswork, je hebt niets aan een perfect geheugen; nee, je moet je ingraven in het onderwerp, inventief zijn, associëren, oplossingen vinden, en uiteindelijk een leesbaar rapport schrijven over dat proces. Haal maar nachten door overlevend op cafeïne, kauw dat potlood eens kapot, blaas je laptop op met te complexe berekeningen, vul de papierbak met mislukte spinsels, ga dagenlang niet bellen, mailen, appen, chatten, tweeten, liken,... En dat noemen we het scriptieproces en het eindproduct noemen we de scriptie. Als je dat lukt mag je met trots de EOR-bul ophalen.

De bachelor thesis

Het voorjaarsemester (periodes 4, 5, 6) van het derde EOR-jaar is het tijdvak dat gereserveerd is voor het scriptieproces, dus zeg maar als je "uit de minor komt". Voor de langzame student zal dat de facto in zijn vierde of vijfde studiejaar zijn want formeel zijn er wel enige voorwaarden, met name (i) 120 studiepunten gehaald, en (ii) keuze tot specialisatie X als de vakken X.I en X.II van jaar 2 zijn gedaan. Want de scriptie zal geschreven worden over een onderwerp in een van de drie EOR-specialisaties, Econometrie, Operations Research, of Wiskundige Economie. Het onderwerp is meestal gebaseerd op een onderzoeksartikel dat gepubliceerd is in een wetenschappelijk tijdschrift. De bedoeling is dat de scriptant het artikel grondig bestudeert, begrijpt, en in eigen woorden kan beschrijven, en dat hij vervolgens aan de slag gaat met uitbreidingen of modelaanpassingen of andere toepassingen of andere data sets of andere oplostechneken, enz. Dit alles in overleg met de begeleider. Een lijst van

geschikte artikelen en bijbehorende begeleiders is vanaf het begin van periode 4 te vinden op de Blackboard pagina van het "scriptievak". Dat beantwoordt de eerste drie prangende vragen.

Hieruit blijkt dat het belangrijk is dat je je inschrijft (aanmeldt) voor het scriptievak in Blackboard, en dat je niet afwacht op het VuNet. Want behalve de artikelenlijst staat in Blackboard een document met heel veel overige informatie over het scriptieproces waarmee de meeste andere prangende vragen worden beantwoord. Met name moet je de deadlines goed in de gaten houden. En ook is in Blackboard een opmaakdocument te vinden dat verwerkt is in LaTeX en dat je kunt gebruiken om de goede layout van de scriptie te produceren.

Maar laat ik toch nog even kort ingaan op de prangende vragen. 1 t/m 3 zijn behandeld. 4. In principe verdiep je je in een wetenschappelijk artikel. In uitzonderlijke gevallen is een stage mogelijk. 5 t/m 7 zijn de punten waarin je je kunt onderscheiden. Een goede afstemming met je begeleider is nuttig. 8. Bijna alle EOR-scripties bevatten programmeeronderdelen (modellen schatten, simulaties, numerieke en statistische methoden, enz). 9. Is je begeleider niet het Nederlands machtig, dan is Engels verplicht, anders mag je zelf kiezen tussen Nederlands en Engels (geen andere taal). 10. Bij voorkeur verwerk je de scriptie in LaTeX en converteer je die naar pdf formaat. In Blackboard staat een LaTeX-opmaakdocument. 11. Klemmend advies is de scriptie op 1 juli afgerond te hebben, want tot die dag mag je op begeleiding rekenen. In de maand juli is je begeleider waarschijnlijk op vakantie of naar congressen. Zo niet, dan zou je zijn laatste commentaren kunnen verwerken. Zorg dat op 31 juli al je cijfers (inclusief die van de scriptie) bij het studiesecretariaat binnen zijn, en dat je je scriptie hebt geüpload naar de bibliotheek. 12. De scriptie wordt beoordeeld door je begeleider en een tweede lezer. De laatste is meestal een docent van de EOR-afdeling. Zij lezen je scriptie grondig door waarna je, in een eindgesprek met hen, reageert op hun commentaar en hun vragen beantwoordt. Daarna beoordelen zij de scriptie aan de hand van een beoordelingstabel en leggen zij het cijfer vast. Deze beoordelingstabel kun je ook weer vinden in de Blackboard pagina.

De master thesis

De scriptie waarmee je de masterfase afrondt is van een andere kaliber dan de bachelor thesis. Nu wordt een duidelijke eigen bijdrage van je verwacht. De bedoeling is dat je aan een probleem werkt en liefst weet op te lossen. Juist daarom is het nu heel goed mogelijk en door sommigen toegejuicht, dat het een onderzoeksproject betreft bij een bedrijf of instelling. Als je dan betrokken bent bij zo'n project via een stage of via je werk, dan krijg je bij wijze van spreken het probleem gratis aangereikt. Vervolgens is dan de eerste stap dat je het probleem helder formuleert in een onderzoeksvraag, of in een probleemstelling, want daardoor baken je het probleem af en krijg je een goed inzicht in het geschikte EOR-vakgebied. Dat geeft weer een leidraad aan het vinden van relevante literatuur, en aan het stipuleren van mogelijke theorieën of modellen waarmee je aan de slag kunt gaan. Deze fase van een scriptieproces naar aanleiding van een stage is dus uitermate belangrijk en doe je in samenspraak met het stagebedrijf, en met de beoogde begeleider van de EOR-afdeling. Maar hoe vind je een stagebedrijf als je geïnteresseerd bent om je master thesis te koppelen aan een stage? Er zijn vele mogelijkheden, via je eigen netwerk, via Kraket, via het FEWEB Office of Career Services, via vacature aankondigingen op de Blackboard pagina van de master thesis, via een zoektocht op Internet, enzovoorts.

Wat betreft Blackboard, net zoals voor de bachelor thesis, is er ook een pagina voor de master thesis. Het is zaak je hiervoor aan te melden (of in te schrijven) bij aanvang van je masterstudie, want ook hier staat (i) een document met informatie over het scriptieproces, (ii) een opmaakdocument in LaTeX, (iii) de beoordelingstabel. Maar daarnaast zijn er lijstjes met scriptie-onderwerpen, verschijnen er aankondigingen van vacatures (zowel van stages als van banen), en aankondigingen van het thesis seminar.

Natuurlijk blijft het ook heel goed mogelijk een zogenaamde theoretische scriptie te schrijven. Dat wil zeggen dat je een probleem onderzoekt dat vanuit een wetenschappelijke belangstelling voortgekomen is. Meestal wordt het onderwerp of probleem aangereikt door een docent. Op

Blackboard (thesis vak) staan enige lijstjes van docenten met onderwerpen, maar niet alle docenten doen dat! Als kandidaat-scriptant mag je vrijelijk "shoppen" bij docenten wier onderzoek jou aanspreekt. Heb je helemaal geen idee, dan kun je terecht bij de contactpersoon van de door jou gekozen specialisatie, en hij zal je dan verder op weg helpen.

Ook hier geldt, begin op tijd, periode 2 of 3 bijvoorbeeld, want het scriptieproces neemt vaak een half jaar in beslag, zeker als je daarnaast nog "gewone" vakken volgt, en zeker als je de scriptie baseert op een stage. Maandelijks wordt een thesis seminar gehouden waarin startende scriptanten hun onderwerp in 5 minuten presenteren aan studenten en docenten. Als je de scriptie succesvol hebt afgerond, dan geef je in het thesis seminar een eindpresentatie erover van maximaal een half uur. Daarna wordt de scriptie en de presentatie beoordeeld door je begeleider en een tweede lezer, ook weer aan de hand van e beoordelingstabel. Dat moet allemaal uiterlijk 31 augustus gebeurd zijn om nog in het academische jaar je studie af te ronden. Mocht je dat niet lukken, bijvoorbeeld omdat de stage langer duurt, of omdat je pas in april (of nog later) bent begonnen, of omdat het programmeerwerk uitloopt, of omdat... er zijn allerlei redenen, dan zul je je moeten herinschrijven in de MSc EOR. Maar dan gaat er een regeling in werking die de faculteit FEWEB heeft ingesteld. Die regeling heet "grade cap", en houdt het volgende in: als een master student is gestart met het scriptieproces vóór 1 augustus, maar pas afrondt na 1 september (m.a.w., in het volgende academische jaar), dan is het cijfer van de scriptie hoogstens een 6.

Met deze sombere mededeling besluit ik mijn verhaal.

*Ad Ridder
Thesis coördinator EOR.*

Puzzeltijd

Schaapjes tellen

Ad heeft weleens last van slapeloze nachten. Omdat het gewone schaapjes tellen niet echt een uitdaging meer is, is Ad overgestapt op wiskundige rijen en reeksen. Na de rij van Fibonacci is hij nu aanbeland bij de partiële sommen van de harmonische reeks:

$$S_k = \sum_{n=1}^k \frac{1}{n} = \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{k}$$

Na een aantal nachten vraagt Ad zich het volgende af: Bestaat er een $k > 1$ waarvoor S_k een geheel getal is?

Drie koningen

Drie koningen heten (in willekeurige volgorde) Waar, Onwaar en Willekeur. Waar spreekt altijd de waarheid, Onwaar liegt altijd en Willekeur antwoordt willekeurig met de waarheid of een leugen. Bernd wil de identiteiten van de drie koningen achterhalen en mag hiertoe drie vragen stellen. Hierbij gelden de volgende randvoorwaarden:

- Het mogen enkel ja/nee vragen zijn.
 - Bernd mag zelf kiezen aan welke koning hij de vraag stelt en hij mag meerdere vragen aan dezelfde koning stellen.
 - De koningen verstaan wel Nederlands, maar spreken het niet. Ze antwoorden met 'pi' en 'mu', maar vooraf weet Bernd niet wat 'ja' en wat 'nee' betekent.
- Gelukkig is Boris zeer intelligent, net als zijn vader. Welke vragen kan Bernd stellen om met zekerheid de identiteiten van de drie koningen te achterhalen?



Mediarecensies

Antwoorden editie 7-1

Brug

Antwoord:

Ja, in 19 minuten kunnen de vier avonturiers allemaal naar de overkant komen:

1. Witek en Nina steken samen de brug over, hier doen ze 3 minuten over.
2. Witek brengt de zaklamp terug, hij doet hier 2 minuten over.
3. Vervolgens gaan Nikki en Joël naar de overkant, hier doen ze 8 minuten over.
4. Nina brengt de zaklamp terug, zij doet hier 3 minuten over.
5. Tot slot gaan Witek en Nina samen naar de overkant in 3 minuten.

Stap 2 en 4 kunnen ook verwisseld worden.

Elfstedentocht

Antwoord:

10 minuten.

De eerste keer dat hij een sloot kiest, is de verwachte tijd tot een stad (Workum óf terug in Hindeloopen) gelijk aan de optelling van de producten van de kans op een sloot en de tijd die hij moet schaatsen. Dit komt neer op $\frac{1}{4} \times 7 + \frac{1}{4} \times 4 + \frac{1}{4} \times 3 + \frac{1}{4} \times 6 = 5$.

Echter is hij dan met kans een half terug in Hindeloopen. In dat geval maakt hij weer een keuze en is hij een verwachte 5 minuten later weer in een stad. Weer is hij met kans $\frac{1}{2}$ terug in Hindeloopen en dit proces herhaalt zich zo. De kans dat hij voor een derde keer moet kiezen is dan $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Dit gaat zo door: $5 + \frac{1}{2} \times 5 + \frac{1}{4} \times 5 + \frac{1}{8} \times 5 + \dots = 5 \times (1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots) = 5 \times 2 = 10$ minuten.

De laatste afleiding is gemaakt door te zien dat het een meetkundige reeks betreft met factor $r = \frac{1}{2}$.

Deze puzzels zijn geplaatst in samenwerking met MIcompany.



The Second Machine Age

Erik Brynjolfsson & Andrew McAfee

The Second Machine Age is een boek geschreven door Brynjolfsson (prof. aan MIT) en McAfee over de effecten van technologische vooruitgang op onze economie. De eerste 'Machine Age' was de industriële Revolutie. Dit boek gaat over de Second Machine Age: het effect van computertechnologie op de welvaart.

In het begin van het boek wordt een merkwaardige eigenschap beschreven. Activiteiten die mensen vaak moeilijk vinden (zoals logica en wiskunde) zijn computationeel makkelijk te automatiseren, maar dingen die wij makkelijk vinden zoals waarneming en beweging vereisen veel rekenkracht. Dit heet Moravec's paradox. Maar tegenwoordig worden zelfs hier grote vooruitgangen geboekt: de zelfrijdende auto leek tien jaar geleden voor veel mensen volkomen onhaalbaar, maar ondertussen zijn er al miljoenen kilometers afgelegd door dergelijke voertuigen.

De auteurs geven veel redenen waarom deze grote vooruitgangen juist nú geboekt worden; ze leggen uit waarom de IT-industrie de enige industrie is waar exponentiële groei (denk aan Moore's Law, maar ook bij kostenefficiëntie van harddrives) over lange tijd gehandhaafd kan blijven.

In het tweede deel worden de economische voor- en nadelen van deze voortschrijdende technologieën besproken, en wordt er beschreven hoe dit voor een toename van de totale welvaart zorgt, maar ook voor een stijging in de inkomensongelijkheid. In de laatste hoofdstukken worden beleidsadviezen voorgesteld. Het boek bevat aansprekende voorbeelden met een heldere structuur en is goed te lezen. De laatste hoofdstukken bevatten veel clichés, maar de rest van het boek is een interessant overzicht van de economische implicaties van technologische innovaties.

The Circle

Dave Eggers



In The Circle schetst Dave Eggers een futuristische samenleving waarin privacy volledig tot het verleden behoort. Het technologiebedrijf The Circle heeft door de jaren heen een stevige positie opgebouwd op het gebied van technologie en innovatie, en doet meteen denken aan een bedrijf als Google of Facebook. Met een eigen campus die voorzien is van alle gemakken, is het een aantrekkelijke plek voor 'newbies' in de technologiewereld.

Als nieuwe werknemer begin je op de afdeling Customer Experience, waar je wordt bijgebracht hoe belangrijk het sociale netwerk voor het bedrijf is. Niet alleen met de buitenwereld, maar vooral ook voor de mensen binnen de 'Inner Circle', allemaal op vrijwillige basis natuurlijk: "Sharing is Caring".

Met uitvindingen als PastPerfect (een perfecte stamboom-historie) en SoulSearch (het fysiek opsporen van iedere willekeurige persoon) laat The Circle zien waar het technologisch gezien toe in staat is. Maar centraal in het boek staat SeeChange, het plaatsen van praktisch onzichtbare cameraatjes om in eerste instantie als toezicht te dienen. De obsessie voor data leidt er echter toe dat ook belangrijke figuren 'openbaar gaan'. De motto's "Privacy is theft" en "Secrets are lies" liegen er dan ook niet om.

Het uiteindelijke doel van de 'Three Wise Men', de oprichters van het bedrijf, is de cirkel compleet te maken, dat wil zeggen een systeem optuigen waarin alle informatie staat opgeslagen en voor ieder doeleinde kan dienen. Hoewel een aantal passages door een overvloed aan detail wat lastiger wegleest en hier en daar wat extra diepgang aangebracht had mogen worden, zet dit boek je aan het denken door zijn relevantie. Technologie in combinatie met big data zal een grote impact hebben op onze toekomstige maatschappij, het is aan ons de taak deze op een goede manier vorm te geven.

Agenda

3 juli - Inhousedag Aegon

Op 3 juli organiseert Aegon een inhousedag voor leden van Kraket. Een mooie gelegenheid voor studenten om een beeld te krijgen van de sfeer en het werken bij Aegon.

4 oktober - Casedag

Op 4 oktober zal de jaarlijkse Casedag van Kraket plaatsvinden. De Casedag biedt derdejaarsstudenten en hoger de kans om kennis te maken met verschillende bedrijven, zowel door middel van het bijwonen van bedrijfscases, als tijdens de lunch en netwerkborrel. De locatie en deelnemende bedrijven zullen in de aanloop naar de Casedag toe bekend worden gemaakt.

Publicaties

Blasques Albergaria Amaral, F., Koopman, S.J., Mallee, M.I.P. & Zhang, Z. (2016). Weighted Maximum Likelihood for Dynamic Factor Analysis and Forecasting with Mixed Frequency Data. *Journal of Econometrics*, 193 (2), 405-417.

Estevez Fernandez, M.A., Borm, P. & Lazarova, E. (2016). Transfers and exchange-stability in two-sided matching problems. *Theory and Decision*, 81 (1), 53-71.

Groen, E.A., Zanten, H.H.E. van, Heijungs, R., Bokkers, E.A.M. & Boer, I.J.M. de (2016). Sensitivity analysis of greenhouse gas emissions from a pork production chain. *Journal of Cleaner Production*, 129 (August), 202-211.

Julien-Laferriere, A., Bulteau, L., Parrot, D., Marchetti-Spaccamela, A., Stougie, L., Vinga, S., Mary, A. & Sagot, M.-F. (2016). A Combinatorial Algorithm for Microbial Consortia Synthetic Design. *Scientific Reports*, 6:29182.

Sitters, R.A. & Wang, C. (2016). On some special cases of the restricted assignment problem. *Information Processing Letters*, 116 (11), 723-728.

Stelling-Konczak, A., Hagenzieker, M.P., Commandeur, J.J.F., Agterberg, M.J.H. & Wee, B. van (2016). Auditory localisation of conventional and electric cars: laboratory results and implications for cycling safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, 41B (August), 227-242.

SECTOR is een uitgave van



Kraket is de studievereniging voor Econometrie en Operationele Research aan de Vrije Universiteit te Amsterdam. De naam Kraket staat voor 'Kritische Aktuarissen en Econometristen'.