

Ondeugdelijke statistiek

Tweeten 4

Vind ik leuk

90 personen vinden dit leuk.

Prof.dr. Peter Grünwald

Hoogleraar statistiek en zelflerende computer systemen
Centrum Wiskunde & Informatica en Universiteit Leiden

1. Wat is de belangrijkste wetenschappelijke ontwikkeling in uw vakgebied?

Statistiek is de laatste 10 jaar steeds meer geïntegreerd geraakt met *machine learning*, het vakgebied binnen de informatica dat zich bezighoudt met zelflerende computers en software. Veel spraakherkenningssoftware is bijvoorbeeld zelflerend: u leest de computer een tekst voor die hij al kent; uw stemgeluid wordt opgeslagen in de vorm van een enorme hoeveelheid nullen en enen, die de computer probeert te matchen aan de tekst. Naarmate u meer voorleest leert de computer uw stem steeds beter kennen, zodat hij u gaandeweg ook gaat begrijpen wanneer u een tekst inspreekt die hij nog niet kent. Dit 'leren' komt neer op het vinden van patronen in een grote hoeveelheid data, en is dus in feite een geavanceerde vorm van statistiek. Ontwikkelingen binnen de spraakherkenning, robotica en data-analyse op het internet hebben de laatste tien jaar tot een kruisbestuiving tussen statistiek en informatica geleid die zijn weerga niet kent. Zo hebben ideeën uit de informatica er bijvoorbeeld voor gezorgd dat het aloude dogma dat statistiek alleen "correlatie" kan aantonen, en nooit "causatie" enigszins op de helling is komen te staan – in sommige, speciale gevallen kan het wellicht toch wel. Omgekeerd hebben de oorspronkelijk puur statistische *Lasso*-methodes het mogelijk gemaakt om uiterst efficiënt grote hoeveelheden complexe data te analyseren – met veel minder rekentijd dan informatici voor mogelijk hadden gehouden.

2. Op welke wetenschappelijke doorbraak hoopt u?

Allereerst hoop ik op een *sociologische* doorbraak. Vrijwel al het experimentele onderzoek in de medische wetenschap, de psychologie en de biologie is gebaseerd op zogenoemde p-waardes. Er kan gepubliceerd worden als een resultaat 'significant' ("p-waarde kleiner dan 5%") is. Dit is echter een gedateerde statistische methode die vaak ook nog eens verkeerd toegepast wordt. Tegelijkertijd is wetenschap steeds meer een industrie geworden, waarbij men zoveel mogelijk significante resultaten probeert te 'scoren'. Hoe opzienbender het resultaat, hoe groter de kans dat het de media haalt. De combinatie van scoringsdrift en slechte statistiek zorgt voor een schrikbarend hoog percentage van niet-reproduceerbare (zeg maar gewoon: foute) onderzoeksresultaten. Dit geldt zelfs voor onderzoek in toptijdschriften. De epidemioloog John Ioannidis heeft in het geruchtmakende artikel *Why Most Published Research Findings Are False (PLoS Medicine 2, 2005)* 45 van de meest geciteerde recente onderzoeksartikelen in de geneeskunde geanalyseerd. 34 van de claims in deze artikelen zijn later opnieuw getoetst, en 14 daarvan bleken bij nader inzien onjuist. Maar liefst 41%! In de psychologie lijkt de situatie nog erger. Bij vrijwel al het 'leuke' onderzoek dat de krant haalt kun je grote statistische vraagtekens zetten. Vorig jaar was er nog een artikel dat de New York Times haalde, waarin gerenommeerde psychologen "precognitie" aantoonde: mensen zouden in de toekomst kunnen kijken. Ook hier werd weer ondeugdelijke statistiek gebruikt. Maar in feite valt de onderzoekers persoonlijk niet veel te verwijten, want hun statistische methoden zijn volstrekt gangbaar in de psychologie.

Dit alles leidt tot slechte toepassingen (denk aan medische adviezen die later weer ingetrokken moeten worden), slechte wetenschap (onderzoek bouwt voort op een hypothese die achteraf onjuist blijkt), en natuurlijk geldverspilling – veel te veel onderzoek blijkt achteraf voor niets te zijn geweest. Wat dat betreft kan men best van een crisis spreken. Niet voor niets wijdde *Science* er onlangs zijn voorpagina aan. Het moge duidelijk zijn: wat mij betreft moet de hele methodologie op de schop. Er moeten andere, betere statistische methoden gebruikt worden. Die methoden bestaan wel degelijk, maar de meeste onderzoekers zijn zelf geen statistici, en kennen deze methoden niet. Statistiek wordt gezien als het saaie sluitstuk van spannend onderzoek; men doet gewoon wat men in de, vaak achterhaalde, boeken vindt, en is weinig kritisch. Bovendien: wanneer men betere statistische methoden zou gebruiken, zou men veel minder vaak overtuigende resultaten vinden. Er zou daarom veel minder gepubliceerd kunnen worden. En daar hebben wetenschappers en wetenschappelijke tijdschriften nou net geen belang bij...

Maar waarom vormen de statistici dan niet een gezamenlijk front? Waarom zetten zij niet gezamenlijk paginagrote advertenties in *Nature* en *Science* waarin ze schrijven "Julie moeten allemaal Methode X gaan gebruiken, die is veel beter!" De reden is dat er verschillende richtingen bestaan binnen de statistiek. Terwijl de meeste statistici het erover eens zijn dat de p-waarde methode niet goed toegepast wordt, is er geen eensgezindheid over hoe het dan wel moet. Iedere stroming heeft zijn eigen favoriete methoden. Begrijp me goed – al deze methoden zijn beter dan de status quo. Maar statistici kunnen het er maar niet met elkaar over eens worden welke methode dan ook echt de beste is. Door de combinatie van gevestigde belangen enerzijds en onenigheid binnen de statistiek anderzijds ben ik tamelijk sceptisch over de mogelijkheid om de historisch gegroeide statistische en publicatiecultuur te doorbreken. Maar o, wat zou het een belangrijke doorbraak zijn!

De onenigheid binnen de statistiek brengt mij op de tweede, puur *wetenschappelijke* doorbraak waar ik op hoop. Er zijn thans twee hoofdstromingen binnen de statistiek, de Bayesiaanse en de frequentistische. Binnen deze stromingen bestaan er weer allerlei deelstromingen. Deze groepen hebben allemaal een fundamenteel andere kijk op het vakgebied. Zou het niet mogelijk moeten zijn om een enkele, overkoepelende theorie te formuleren? Dat zou het gebied in wetenschappelijk opzicht interessanter en aantrekkelijker maken – het wordt nu door andere wiskundigen toch vaak gezien als de vreemde eend in de wiskundebij – en in praktisch opzicht zouden we, wie weet, die zo gewenste betere methoden ook echt kunnen introduceren. Ik roep al een tijdje: *we need more GUTS - Grand Unified Theory of Statistics*.

3. Wat is de waarde van uw vakgebied voor de samenleving?

Met statistiek gaat er veel mis, maar zonder statistiek waren we helemaal nergens in bijvoorbeeld de medische wetenschap. Ik durf te stellen dat we zonder moderne statistiek gemiddeld allemaal een paar jaar minder zouden leven. Ook in *rechtszaken* speelt statistiek een steeds grotere rol – denk aan DNA bewijs, dat vaak van de vorm is 'de kans dat een willekeurig iemand dit profiel heeft is 1 op 100000'. En *last but not least*: zonder statistiek geen *Google*: om een goede zoekmachine te maken moet je voortdurend op zoek naar patronen in de gigantische hoeveelheid data op het web – statistiek dus. Zoals Hal Varian, de *chief economist* van Google het zei in 2009: *I keep saying the sexy job in the next ten years will be statisticians* (sic).

Andere bijdragen in [Digitale informatie](#), [Wiskunde](#)

Tweeten 4

Vind ik leuk

Vind dit als eerste van je vrienden leuk.