KM 43 (1993)
pg. 5-7

STATISTIEK ANDERS

Arie ten Cate *

Men overweegt iets te laten onderzoeken door een laboratorium in Delft, of een in Utrecht. Beide laboratoria hebben hun goede en slechte eigenschappen. Men komt er niet uit en werpt een munt op. Het wordt Delft. Vraag: wanneer straks de meetresultaten uit Delft worden geanalyseerd, moet dan rekening gehouden worden met de kans dat het Utrecht had kunnen worden?

We antwoorden hierop natuurlijk ontkennend. Utrecht heeft met de meetresultaten niets te maken. Een goede analyse is gebaseerd op feiten (en een model); gebeurtenissen die niet zijn opgetreden doen niet ter zake. Als we een statistisch model opvatten als de kansverdeling van data met een vector van onbekende parameters, brengt dit ons tot de aannemelijkheidsfunctie: de kansverdeling van de data wanneer de waargenomen data-waarden daarin worden ingevuld, zodat een functie van alleen de parameters ontstaat. De aannemelijkheidsfunctie geeft voor iedere waarde van de parametervector de kans op (of kansdichtheid van) hetgeen gebeurd is. Ze is bepaald op een willekeurige positieve multiplicatieve constante na.

Principe: we waarderen parameterwaarden *uitsluitend* naar hun aannemelijkheid (= de waarde van de aannemelijkheidsfunctie); hoe meer aannemelijk, des te plausibeler.

Uitgaande van dit Principe heeft de meest aannemelijke parametervector uiteraard onze belangstelling. Niet vanwege fraaie asymptotische eigenschappen, maar omdat het volgens het Principe de meest plausibele is.

Een aanduiding van de *nauwkeurigheid* van de informatie die de data geven omtrent de parameters, halen we uit het verloop van de aannemelijkheidsfunctie rond het maximum. Geen standaardfouten, betrouwbaarheidsintervallen of significantietoetsen. Deze drie worden door ons Principe uitgesloten. Ze hebben betrekking op sommeringen (of integralen) over de steekproefruimte en kunnen niet uit de aannemelijkheidsfunctie worden afgeleid.

---

* Centraal Planbureau, Van Stolkweg 14, 2585 JR 's-Gravenhage; telefoon 070-3383380.
Deze korte notitie is bedoeld om – op een enigszins pamflet-achtige manier – aandacht te vragen voor de literatuur over de '*likelihood*-school', die meer aandacht verdient dan zij krijgt. Deze korte notitie is niet bedoeld als beschrijving van de werkwijze op het CPB.

Dit hoeft geen dramatische consequenties te hebben. Beschouw bijvoorbeeld het simpele geval van een trekking uit een normale verdeling met onbekende te schatten verwachting en bekende variantie; de getrokken waarde is de meest aannemelijke schatting van de gezochte verwachting. Het valt eenvoudig in te zien dat $\pm 2\sigma$ rond de trekking (ruwweg 95% betrouwbaarheid) identiek is aan het 'aannemelijkheidsinterval' dat begrensd wordt door de parameterwaarden die e^2=7½ maal zo weinig aannemelijk zijn als de meest aannemelijke waarde. Deze laatste definitie is compatibel met het Principe. Zie Statistische Onderzoekingen M28 van het CBS, uit 1986, bijlage C, voor een vergelijking tussen betrouwbaarheidsintervallen en aannemelijkheidsintervallen rond geschatte kwantielen.

De vraag of we een hypothese omtrent de parameters van een statistisch model moeten verwerpen, hangt ervan af hoeveel minder aannemelijk die hypothese is in vergelijking met de maximale aannemelijkheid. Bijvoorbeeld met bovengenoemde 7½ als kritieke waarde van de aannemelijkheidsverhouding – in plaats van de gebruikelijke kans van 5% op ten onrechte verwerpen. (Beide getallen zijn natuurlijk even willekeurig.)

De statistiek wordt op deze manier eenvoudiger. Het is niet meer nodig om de kansverdeling van een schatter te bepalen. Geen geschatte covariantiematrix $E(-H^{-1})$ meer, maar eenvoudigweg $-H^{-1}$ zelf. H is de Hessiaan van de log-aannemelijkheid in het optimum: de matrix van tweede afgeleiden naar de parameters. (De tweede afgeleide is een maatstaf van kromming; hoe groter/kleiner de diagonale elementen van $-H^{-1}$, des te vlakker/spitser verloopt de aannemelijkheidsfunctie rond het optimum.)

Merk op dat bij deze nauwkeurigheidsaanduidingen (daling van aannemelijkheid tot een bepaalde waarde, of de tweede afgeleide van de log-aannemelijkheid) niets geschat wordt. Vergelijk het gebruik van een aannemelijkheidsfunctie met het kijken naar een wazige verte: de wazigheid maakt deel uit van hetgeen je ziet.

Tenslotte passen we het Principe toe op nog twee onderwerpen: stopregels en 'censurering'.

Beschouw een aantal waarnemingen die kunnen worden opgevat als trekkingen uit een tweewaardige verdeling. De kansverdeling van de fractie 'successen' hangt af van de stopregel. Bijvoorbeeld, wanneer gestopt wordt zodra een bepaald aantal *successen* is bereikt, dan is deze fractie geen zuivere schatter van de kans op succes. (Gemiddeld een overschatting.) Ook kritieke grenzen voor het testen van een hypothese omtrent de gezochte kans zijn in dat geval anders dan wanneer men stopt na een bepaald aantal *trekkingen*.

Verplaatst men zich nu eens in de situatie van degene die de experimenten die aldus gemodelleerd worden, heeft verricht. Zij weet precies wat er is gedaan en ze krijgt de leiding over de statistische analyse, terwijl de baas van het laboratorium op safari gaat. Volgens de traditionele statistiek is het nu niet te bepalen wat de resultaten vertellen over de te schatten kans, als het onbekend is waarom de baas indertijd heeft bepaald dat er gestopt moest worden – omdat er voldoende 'successen' waren opgetreden, of juist omdat het duidelijk was geworden dat er onvoldoende successen waren, of omdat het geld op was, of omdat iemand voortzetting ontraadde...

Met ons Principe doet zich dit probleem niet voor. De aannemelijkheidsfunctie van de gezochte kans is onafhankelijk van de stopregel. De fractie successen is dan ook altijd de meest aannemelijke waarde.

Wat betreft censurering kan ik niet beter doen dan het klassieke voorbeeld van Pratt in de discussie bij Birnbaum (1962) aan te halen – in navolging van Edwards (1972) en Berger en Wolpert (1988). Hierbij is de censurering niet effectief en heeft zij geen effect op de aannemelijkheidsfunctie.

Een ingenieur doet een aantal nauwkeurige spanningsmetingen. De metingen lopen uiteen van 75 tot 99 volt. Een orthodoxe statisticus onderzoekt de metingen en geeft een betrouwbaarheidsinterval voor de verwachtingswaarde, daarbij de metingen opvattend als trekkingen uit een normale verdeling. Later bezoekt hij het laboratorium en merkt dat de gebruikte voltmeter maar tot 100 volt gaat, zodat de populatie 'gecensureerd' is. De ingenieur vertelt echter dat hij ook een voltmeter heeft die tot 1000 volt gaat, bedoeld voor spanningen boven de 100 volt. Gelukkig waren die er niet: de ingenieur vertelt dat hij de dag na de metingen ontdekte dat de 1000-volt-meter kapot was. De statisticus verzekert zich ervan dat de ingenieur de metingen niet zou hebben uitgesteld als er een spanning boven 100 volt geweest was en vertelt de ingenieur dat voor een exact betrouwbaarheidsinterval een nieuwe statistische analyse nodig is, vanwege censurering. De ingenieur is perplex. "Er is precies hetzelfde gebeurd als wanneer die meter niet kapot was. Ik heb precies hetzelfde van de metingen geleerd als wanneer die meter wel beschikbaar was geweest. Straks vraagt u me nog naar mijn oscilloscoop."

Literatuur

Berger, J.O. en R.L. Wolpert, 1988, The likelihood principle (second edition). Institute of Mathematical Statistics, Hayward

Birnbaum, A., 1962, On the foundations of statistical inference. JASA 57, pp. 269-326

Edwards, A.W.F., 1972, Likelihood. Cambridge University Press, Cambridge