

BAYESIAANSE STATISTIEK EN HET METEN VAN VOORKENNIS

Ivo W. Molenaar*)

Samenvatting

De Bayesiaanse statistiek onderscheidt zich van de traditionele door een andere interpretatie van het kansbegrip en door het gebruik van voorkennis. Een discussie van de voor- en nadelen leidt tot de conclusie dat met name bij het beslissen in onzekere situaties de Bayesiaanse methode in aanmerking komt, mits het mogelijk is de aanwezige voorkennis op een betrouwbare manier in een a priori kansverdeling om te zetten.

In een aantal psychologische experimenten is gebleken dat de met dat doel aan een proefpersoon ontlokte informatie niet zonder voorzorgen bruikbaar is: de antwoorden zijn onbetrouwbaar, soms onderling tegenstrijdig, te optimistisch en vatbaar voor verandering in de ondervragingsmethode. Combinatie van diverse methoden, met terugkoppeling van de antwoorden en herhaalde mogelijkheid van correctie, kan daarin mogelijk verbetering brengen. Met het door Lourens en Terlouw geconstrueerde interactieve computerprogramma SPAT zijn hiertoe enkele experimenten gedaan: de verbetering blijkt vooralsnog tegen te vallen. Tenslotte wordt een door Euverman en Vermulst ontwikkeld wiskundig model voor Bayesiaanse factoranalyse in het kort samengevat; ook hier is een conversationeel computerprogramma een belangrijk hulpmiddel.

*) Vakgroep Statistiek en Meettheorie, FSW, Rijksuniversiteit Groningen
Oude Boteringestr. 23, 9712 GC Groningen. De auteur dankt Lourens, Terlouw en Vermulst voor enkele redactionele suggesties.

1. Kansinterpretatie en voorkennisgebruik

De inferentiële of inductieve statistiek wordt in empirisch onderzoek gebruikt om tot uitspraken met een generaliserend karakter te komen. Voorzover een aselechte steekproef is getrokken, dan wel de onderzoekseenheden aselekt aan diverse condities zijn toegewezen, probeert de gebruikelijke toepassing van mathematische statistiek daarbij antwoord te geven op de vraag "wat gebeurt er als ik het overnieuw doe?" en daarmee de onderzoeker in staat te stellen systematische verschijnselen te onderscheiden van verschijnselen die kunnen worden toegeschreven aan de toevallige steekproeftrekking of de toevallige experimentele toedeling. Kans wordt daarbij geïnterpreteerd als relatieve frequentie onder een groot aantal (denkbeeldige) herhalingsen van het onderzoek.

De voorkennis waarover de onderzoeker al beschikt wordt bij het formuleren van de conclusies uit de data zorgvuldig buiten beschouwing gelaten. Deze voorkennis mag wel een rol spelen bij de opzet van het onderzoek (keuze van de vraagstelling en de meetmethoden, formulering van de hypothesen, vaststelling van de steekproefomvang), maar nadat deze keuzen zijn gemaakt wordt het ideaal nagestreefd dat elke statisticus uit de data exact dezelfde conclusies zou trekken. Voorzover dit niet het geval is (keuze van de toets, behandeling van uitbijters of ontbrekende waarnemingen) wordt dit door de meeste vakgenoten gezien als een tekortkoming in de wetenschappelijkheid: statistiek behoort ook in haar toepassingen een kunde en geen kunst te zijn.

De omgang met onzekerheid in het dagelijks leven, en ook in de dagelijkse onderzoekspraktijk, voldoet evenwel niet aan de beide in de vorige alinea's gesteide eisen. Een kabinetsformateur vertelt de journalisten dat hij de kans veertig procent acht dat hij in zijn opdracht zal slagen, een uitgever praat over de kans dat van een bepaald boek nog dit jaar een herdruk nodig zal zijn, en iedereen heeft het over de kans dat het morgen zal regenen. Van aselechte trekking of toedeling is in deze gevallen geen sprake, en het is uiterst gewrongen om een kunstmatig universum te postuleren van net zulke kabinetsformaties, net zulke boeken of net zulke weers-

omstandigheden, waarin dan het bedoelde verschijnsel met een bepaalde relatieve frequentie zou optreden. Kans wordt hier veeleer opgevat als de uitdrukking in een getal van de mate van redelijk geloof. Dit impliceert onmiddellijk dat men het over de slaagkans van een kabinetsformatie of de voorspelde verkoop van een boek oneens kan zijn, zonder dat één partij ongelijk heeft. Wel postuleert Carnap dat twee personen met exact dezelfde voorkennis redelijkerwijs tot dezelfde kansuitspraak behoren te komen, maar omdat zulke personen niet bestaan is de Bayesiaanse kans personalistisch: Uw voorkennis is de mijne niet.

Voor allerlei beslissingsproblemen kan men nu proberen met deze personalistische kansen te werken bij de keuze van een strategie die b.v. het verwachte nut moet maximaliseren. Vooral wanneer de voorkennis gering is en de onzekerheid groot, is het gewenst daarbij een steekproef of experiment te gebruiken ter aanvulling van de kennis. Volgens de Bayesiaanse theorie behoort dan de kennis na datavergaring de vorm te hebben van de a posteriori verdeling die volgens de stelling van Bayes ontstaat door de aannemelijkheidsfunctie van de data te combineren met de a priori verdeling van de relevante parameter(s).

Zowel de interpretatie van kans als mate van redelijk geloof als het model van kennisgroei volgens de stelling van Bayes zijn door de meeste statistici uiterst sceptisch bejegend. In de ogen van Van Dantzig (1957) zijn personalistische kansen, evenals fiduciële kansen, een onwetenschappelijke zaak: parameters zijn immers vaste maar onbekende grootheden, die niet volgens een kansverdeling fluctueren. Dat ben ik volstrekt met hem eens: zolang het frequentistisch kansbegrip wordt gehanteerd is een parameter vast (behoudens enkele vanzelfsprekende uitzonderingen zoals model II variantie-analyse of discriminant-analyse waar de parameters op een voor frequentisten acceptabele manier via een toevalsproces tot stand komen). Zoals boven aangeduid is de variantie van een a priori verdeling evenwel geen aanduiding van een mate van toevalsfluctuatie, maar van een mate van persoonlijke onzekerheid over een onbekende grootheid.

Van Zwet formuleerde een kleine tien jaar geleden soortgelijke bezwaren tijdens een lezing op de Statistische Dag, waarin Bayesiaanse statistiek

met zwarte kunst werd vergeleken. Hemelrijk drukte zich in een brief aan mij iets milder uit: als er al persoonlijke kansen bestaan zou hij die ter voorkoming van verwarring met "pansen" willen aanduiden. Hij betwijfelde of die pansen zich volgens de axioma's van de kansrekening zouden gedragen en zich volgens de regel van Bayes met kansen lieten combineren. Andere auteurs voeren inderdaad als bezwaar aan dat het effect van een steekproef op de subjectieve kansschatting van een proefpersoon lang niet altijd conform de regel van Bayes verloopt: kleine steekproeven hebben meer en grote steekproeven minder invloed dan men volgens die regel zou verwachten.

De in mijn ogen belangrijkste voordelen en bezwaren van de Bayesiaanse methode zijn hiermee genoemd: ze is ook toepasbaar waar herhaling van een lotingsmechanisme niet aan de orde is en ze laat een combinatie toe van voorkennis en waarnemingen, maar die combinatie van kans als voorkennis met kans als lotingsfrequentie is aanvechtbaar en leidt tot verlies aan objectiviteit. Over dit conflict zijn in het verleden heftige discussies gevoerd, waarbij de mathematische statistici overwegend voor het frequentistische standpunt hebben gekozen, terwijl de subjectivistische opvatting werd aangehangen door talrijke wetenschapsfilosofen en door enkele dissidente statistici zoals Lindley en Box. Er zijn maar weinig publicaties waarin beide posities zo objectief mogelijk worden vergeleken zonder dat de auteur partij kiest: een leerzaam voorbeeld is Barnett (1973).

Ik ben van mening dat het onmogelijk is om met wetenschappelijke argumenten overtuigend aan te tonen dat de ene opvatting over de hele linie superieur zou zijn aan de andere. Wie meent dat dit wel kan, zou lering kunnen trekken uit bijvoorbeeld de debatten tussen statistici over verdelingsvrije methoden, tussen logici over intuitionisme of tussen wetenschapsfilosofen over inductie. Ook wanneer binnen elke school een sluitend bouwwerk van redeneringen zou bestaan, kan de keuze van de uitgangspunten fundamenteel verschillend zijn en uiteindelijke berusten op persoonlijke afwegingen van plausibiliteit, doelmatigheid en redelijkheid.

Het is volgens mij een illusie dat het gebruik van statistische methoden gelijk te stellen is met de strikt objectieve toepassing van één uitputtend

en overtuigend stelsel van voorschriften. Ik ben het eens met de frequentisten dat objectiviteit bij het wetenschappelijk handelen nastrevenswaardig is. Maar het is onvermijdelijk dat persoonlijke afwegingen een rol spelen bij de opzet van een experiment of survey, bij de formulering van een statistisch model, bij de keuze van een schatter of toets en bij de vertaling van de wiskundige conclusie in een inhoudelijke. Leamer (1978) zegt hierover veel verstandige dingen, waarop ik door plaatsgebrek niet kan ingaan.

De psycholoog A.D. de Groot (1961, 1981) bepleit het streven naar intersubjectiviteit wanneer objectiviteit niet mogelijk blijkt te zijn. De communicatie met de vakgenoten door wetenschappelijke publicaties vergelijkt hij met het functioneren van een wetenschappelijk forum, dat claims van de onderzoekers bespreekt en weegt. Bij een gunstige uitslag kan gezegd worden dat "de geleerden het er over eens zijn dat...", maar ook bij een goed functionerend forum is dit een voorlopige conclusie die in het licht van nieuwe feiten of inzichten kan worden bijgesteld. Het verzamelen van data om een ander te overtuigen wordt besproken door Novick & Jackson (1974) in het kader van de Bayesiaanse methode.

Voor de evaluatie van die methode houdt het bovenstaande in dat ik alleen een persoonlijk standpunt kan formuleren, en daarin wil ik onderscheid maken naar de aard van de toepassing. Bij de tegenwoordige stand van de inzichten acht ik het als regel niet aanbevelenswaardig de Bayesiaanse methode te gebruiken bij de statistische evaluatie van experimenten of surveys die primair gericht zijn op de groei van algemene wetenschappelijke kennis. Zoals boven opgemerkt kent de klassieke aanpak ook wel aspecten waarbij een persoonlijke keuze vereist is, maar zij biedt toch meer houvast voor wetenschappelijk uitspraken die de bedoeling hebben de risico's van generalisatie in kaart te brengen. Bij persoonlijke en zakelijke beslissingen in onzekere situaties daarentegen is het zo goed mogelijk verdisconteren van alle beschikbare voorkennis en alle beschikbare data in een eindoordeel over een modelparameter vaak de beste weg om tot een optimale beslissing te geraken, zie bijv. Raiffa (1968). Daarnaast biedt het Bayesiaanse model een uitstekend raamwerk voor bijv. regressie naar het gemiddelde

(Novick & Jackson, 1974, p. 94), schatting van variantiecomponenten (Box & Tiao, 1953, ch.5), regressie in verwante kleine steekproeven (Molenaar & Lewis, 1979 en Jansen 1977) en schatting van verwante proporties (Lewis e.a. 1975).

Deze discussie van het grondslagenprobleem is noodgedwongen onvolledig: men raadplege bijvoorbeeld Barnett (1973), Wilkinson (1977) of Lindley & Smith (1972). Hier wordt alleen nog opgemerkt dat de eisen aan een statistisch model volgens mij voor een Bayesiaan even stringent behoren te zijn als voor een frequentist: aselechte en onafhankelijke trekking, en assumpties zoals een bepaalde verdeling of gelijke varianties moeten immers garanderen dat de aannemelijkheidsfunctie de informatie correct weergeeft. Bovendien moet de voorkennis onafhankelijk van de data worden vastgesteld om vermenigvuldiging in de regel van Bayes toelaatbaar te maken.

2. Het specificeren van voorkennis

Vooral onder econometristen (b.v. Zellner, 1971) bestaat de neiging om wel kansen aan parameters toe te kennen, maar de invloed van persoonlijke voorkennis te weren door de a priori verdeling neutraal of niet-informatief te kiezen; zie ook Box & Tiao (1973). Naar mijn mening wordt daarbij toch tekort gedaan aan het begrip "voorkennis", want een vakkundig onderzoeker zal maar zelden tevoren geen enkel idee hebben over de plausibiliteit van diverse parameterwaarden. De kunst is alleen om de ietwat vage en ongearticuleerde voorkennis die er ongetwijfeld is te preciseren tot een wiskundig uitgedrukte kansverdeling, die zich dan leent voor een voorspelling, een beslissingskeuze of een combinatie met de aannemelijkheidsfunctie tot een a posteriori verdeling gegeven de data. Soms zal men behoefte hebben aan een gevoeligheidsanalyse om te zien bij welke kleine veranderingen in de a priori verdeling de conclusies nog stabiel blijven; zie daarvoor bijv. Vijn & Molenaar (1981) en Leamer (1978).

De ondervraging van een onderzoeker omtrent de aanwezige voorkennis is een gecompliceerde zaak, omdat die voorkennis niet in de vorm van een kansverdeling in het menselijk geheugen ligt opgeslagen. De ondervraagde

moet niet alleen zijn/haar geheugen raadplegen om de relevante informatie te selecteren, maar die informatie moet ook nog tot een kansverdeling worden vertaald. Daarbij doet zich het probleem voor dat zelfs onderzoekers die geregeld met statistische methoden omgaan nogal veel denkfouten maken bij het ongewapend oordelen over kansen. Het is niet mogelijk hier de talrijke studies over dit verschijnsel te behandelen: men raadplege bijv. Einhorn & Hogarth (1981), Peterson & Beach (1967), Tversky (1974), Hogarth (1975, 1980), Tversky & Kahnemann (1983).

De vraag blijft dan wel of er een ondervragingsmethode te vinden is die daarmee rekening houdt en tot voldoende betrouwbare resultaten leidt. Perfecte betrouwbaarheid is een illusie, maar ook niet nodig. Zo zegt Leamer (1978, p. 41): "anyone who insists that personal probabilities be assessed precisely cannot perform statistical inference, for the same reason that no spaceship would ever have reached the moon if measurement of lengths had to be perfectly precise before construction could commence. To say that something is 6.5 centimeters long is only to say that for the purposes at hand we may proceed as if it were".

De mogelijkheden om subjectieve voorkennis omtrent een proportie betrouwbaar te formaliseren zijn psychometrisch en experimenteel-psychologisch onderzocht door Lourens & Terlouw in ZWO-project 56.145. Aan Lourens (1984) wordt hier een korte schets ontleend van enkele problemen die zich bij zo'n onderzoek voordoen.

Allereerst moeten criteria worden opgesteld voor een "goede" meting van een subjectieve kansverdeling; de reacties van de ondervraagde moeten bijvoorbeeld aan de axioma's van de waarschijnlijkheidsrekening voldoen, bij verschillende meetmethoden elkaar niet tegenspreken, stabiel in de tijd zijn (tenzij intussen een nieuwe kennisbron is aangeboord), tot een geldige representatie van de bedoelde kennis leiden, en dergelijke. Omdat uit eerder empirisch onderzoek blijkt dat de meest ondervragingsmethoden afzonderlijk niet aan zulke eisen voldoen (Schütt, 1981) hebben Lourens en Terlouw een combinatie van methoden opgenomen in hun meetinstrument. Voor een goede meting mag evenwel niet worden geëist dat de ware parameterwaarde de modus van de verdeling is, en evenmin dat alle ondervraagden dezelfde verdeling

produceren: zij zullen immers doorgaans over verschillende en feilbare voorkennis beschikken.

Het belang van een conversationeel computerprogramma voor de specificatie van a priori verdelingen is eerder benadrukt door Isaacs en Novick (1980) in hun pakket CADA (Computer Assisted Data Analysis). Lourens en Terlouw bouwen in hun eigen computerprogramma SPAT (Subjective Probability Assessment Technique) hierop voort door drie ondervragingsmethoden te combineren, met terugkoppeling en gelegenheid tot herziening zowel per methode als na vergelijking van de drie resultaten. In een experiment met 96 proefpersonen werd onder meer het effect van de volgorde van de drie methoden onderzocht, alsmede de hertest betrouwbaarheid na een week en de invloed van een groepsdiscussie tussen telkens vier proefpersonen. De voorlopige resultaten bevestigen eerdere bevindingen dat ongetrainde proefpersonen doorgaans veel moeite hebben met taken van dit type: men overschat de eigen voorkennis, geeft soms onlogische antwoorden en is te gauw tevreden met een door SPAT getoonde voorlopige verdeling.

Zowel eerdere publicaties (Hogarth, 1975, 1980) als eigen onderzoek (Lourens, 1981) doen vermoeden dat training en aanbieding van de juiste hulpmiddelen in deze onbevredigende situatie verandering kunnen brengen. Dit is in elk geval met succes gebeurd bij meteorologen (Murphy, 1983). Uit de experimenteel-psychologische literatuur blijkt in elk geval dat een ongewapend oordeel over kansen niet zonder meer betrouwbaar is.

3. Bayesiaanse factoranalyse

Een toepassing van het in de voorgaande paragrafen behandelde is de door Euverman en Vermulst (1983) ontwikkelde Bayesiaanse schattingsmethode voor factoranalyse, waarvan enkele hoofdpunten hier worden aangestipt.

Stel dat een stochastische vector $\underline{x}$ van p variabelen worden uitgedrukt in een vector $\underline{f}$ van $k < p$ gemeenschappelijke factoren aangevuld met unieke

componenten door het model

$$\begin{matrix} \underline{x} & = & \Lambda & \underline{f} & + & \underline{u} & + & \underline{\mu} \\ \text{px1} & & \text{pxk} & \text{kx1} & & \text{px1} & & \text{px1} \end{matrix}$$

waarbij $\underline{x}$ verwachtingsvector $\underline{\mu}$ heeft en covariantiematrix Σ gegeven door

$$\Sigma = \Lambda \Phi \Lambda' + \Psi.$$

Zoals gebruikelijk wordt verondersteld $E \underline{f} = 0$, $E \underline{u} = 0$, $E \underline{f} \underline{u}' = 0$, terwijl Ψ de diagonale matrix van unieke varianties $E \underline{u} \underline{u}'$ en Φ de correlatiematrix van de k factoren is; Λ is de patroonmatrix.

Wanneer de vector $\underline{x}$ bij N aselekt gekozen elementen is waargenomen, en normaal verdeeld mag worden verondersteld zijn steekproefgemiddelden $\bar{x}$ en steekproefcovariantiematrix S afdoende voor $\underline{\mu}$ en Σ , en daarmee voor meest aannemelijke schatting van de elementen van $\underline{\mu}$, Λ , Φ en Ψ door maximaliseren van

$$(2\pi)^{-\frac{1}{2}N} |\Sigma|^{-\frac{1}{2}N} \exp \left[-\frac{1}{2} N \text{tr} (S \Sigma^{-1} + \Sigma^{-1} (\bar{x} - \underline{\mu}) (\bar{x} - \underline{\mu})') \right].$$

Nu zal een deskundig onderzoeker doorgaans over voorkennis beschikken over de relatie tussen de p variabelen en de k factoren. Gebruik van die voorkennis heeft enkele bijkomende voordelen: het vermindert de kans op negatieve variantieschattingen (Heywood case), het legt een unieke oplossing vast terwijl anders elke orthogonale rotatie tot een even goede oplossing leidt, en het vermindert de gevoeligheid van de schattingsmethode voor eventuele uitbijters.

Omdat in veel toegepast onderzoek correlaties gemakkelijker numeriek te preciseren zijn dan covarianties, worden de parameters (de elementen van Λ , Φ , Ψ en $\underline{\mu}$) eerst ten dele getransformeerd door over te gaan op $\delta_i = \sigma(x_i)$ en $\omega_{ij} = \rho(x_i, x_j)$. Euverman & Vermulst veronderstellen dat de voorkennis omtrent de standaardafwijkingen δ_i kan worden benaderd door inverse chi verdelingen (zie b.v. Novick & Jackson, 1974 of Box & Tiao, 1973). Voor de correlaties ω_{ij} en ϕ_{hj} worden voor eenvoudige modellering

normale verdelingen gebruikt, hoewel die zich niet tot het interval $[-1,1]$ beperken. Omdat de gemiddelden niet interessant zijn wordt de voorkennis daarover verwaarloosd en een constante dichtheid voor μ gekozen.

In de bijbehorende computerprogrammatuur dient de gebruiker eerst aantal en namen van de variabelen en de factoren op te geven. Dan volgt een ondervraging naar de voorkennis omtrent δ_i , waarvan de resultaten worden omgerekend in de geschatte parameters van de inverse chi verdelingen. Een soortgelijke stap volgt voor de correlaties ω_{ij} en ϕ_{hj} , waarbij de mogelijkheid bestaat $\phi = 1$ te kiezen als een orthogonale oplossing wordt verlangd. Nu is het mogelijk dat de resulterende a priori verdeling een niet verwaarloosbare simultane kans zou toekennen aan parametercombinaties die niet toelaatbaar zijn in verband met de eisen van positief definitie Σ , positieve diagonalelementen van Ψ en dergelijke. In dat geval volgt een vraag en antwoord spel dat leidt tot een toelaatbare herspecificatie die zo min mogelijk van de vorige specificatie afwijkt. De resultaten worden daarna gecombineerd met de data van de steekproef tot een a posteriori verdeling, waarbinnen met een aangepast quasi-Newton algoritme de simultane modale schatters worden bepaald.

Voor enkele toepassingen en voor een nadere beschrijving wordt verwezen naar Euverman & Vermulst (1983). Voor het hier gehouden betoog is het vooral van belang dat ook in een ingewikkeld model met veel parameters een formalisering van de voorkennis mogelijk is gebleken, en dat de numerieke problemen door een verstandig gebruik van de computer oplosbaar waren. Het hangt uiteraard van de aard van het onderzoek af of de aldus verkregen Bayesiaanse schatters te verkiezen zijn boven de meest aannemelijke en boven de via eventuele andere methoden verkregen waarden.

Wanneer men na lezing van dit artikel eerder geneigd zou zijn de Bayesiaanse methoden de hun passende plaats onder de zon toe te kennen (niet groter en niet kleiner dan de afweging van de voor- en nadelen wettigt), en over een wegwijzer naar verdere literatuur beschikt, dan heeft dit artikel voldaan aan de bedoelingen die ik er mee had.

Literatuur

- Barnett, V. (1973) *Comparative Statistical Inference*, Wiley, New York.
- Box, G.E.P. & Tiao, G.C. (1973) *Bayesian inference in statistical analysis*, Addison Wesley, Reading (mass.).
- Chiu, W.K. & Leung, M.P.Y. (1980) A new Bayesian approach to quality control charts, *Metrika*, 27, 243-253.
- Dantzig, D. van (1957) Statistical priesthood, *Statistica Neerlandica*, 11, 1-16 & 185-200.
- Einhorn, T.J. & Hogarth, R.M. (1981) Behavioral decision theory. *Annual review of Psychology* 32, 53-88.
- Euverman, T.J. & Vermulst, A.A. (1983) *Bayesian factor analysis*, dissertatie, Rijksuniversiteit Groningen, november 1983.
- Groot, A.D. de (1961) *Methodologie*, Mouton, 's-Gravenhage.
- Groot, A.D. de (1981) Over regels en taken van het forum in de gammawetenschappen, oratie Groningen, Boom, Meppel.
- Hogarth, R.M. (1975) Cognitive processes and the assessment of subjective probability distributions, *Journal of the American Statistical Association*, 70, 271-294.
- Hogarth, R.M. (1980) *Judgement and choice, the psychology of decision*, Wiley, New York.
- Isaacs, G.L. & Novick, M.R. (1980) *Computer-assisted data analysis*, Lindquist Center for Measurement, University of Iowa.
- Jansen, G.G.H. (1977) *An application of Bayesian statistical methods to a problem in educational measurements*, dissertatie Rijksuniversiteit Groningen.
- Leamer, E.E. (1978) *Specification Searches*, Wiley, New York.
- Lewis, C., Wang, M.M. & Novick, M.R. (1975) Marginal distributions for the estimation of proportions in m groups, *Psychometrika*, 40, 63-76.
- Lindley, D.V. & Smith, A.F.M. (1972) Bayes estimates for the linear model (with discussion), *Journal of the Royal Statistical Society B* 34, 1-41.

- Lourens, P.F. (1984) The formalization of knowledge by specification of subjective probability distributions: an experimental approach, dissertatie Rijksuniversiteit Groningen.
- Molenaar, I.W. (1977) On Bayesian formula scores for random guessing in multiple choice tests, *British Journal of Mathematical and Statistical Psychology*, 30, 79-89.
- Molenaar, I.W. & Lewis, Ch. (1979) Bayesian m-group regression: a survey and an improved model, *Methoden en Data Nieuwsbrief (MDN) SWS VvS* 4, 62-72.
- Molenaar, I.W. (1980) An insurance policy against unexpected data, *Kwantitatieve Methoden (KM) VvS*, 1, 49-74.
- Murphy, A. (1983) Probabilistic weather forecasting in the Netherlands and the United States: some recent results. Paper presented at the Ninth Conference on Subjective Probability, Utility and Decision Making, Groningen 1983.
- Novick, M.R. & Jackson, P.H. (1974) *Statistical methods for educational and psychological research*, McGraw Hill, New York.
- Peterson, C.R. & Beach, L.R. (1967) Man as an intuitive statistician, *Psychological Bulletin* 68, 29-46.
- Raiffa, H. (1968) *Decision analysis, Introductory lectures on choices under uncertainty*, Addison Wesley.
- Schütt, K.P. (1981) *Wahrscheinlichkeitsschätzungen im Computer-Dialog: Theorie, Methoden und eine experimentelle Studie zur Schätzung von subjectiven Wahrscheinlichkeiten*, Poeschel, Stuttgart.
- Tversky, A. (1974) *J. Royal Stat. Society Series B* 36, 148-159.
- Tversky, A. & Kahneman, D. (1983) Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement. *Psychological Review* 90, 293-315.
- Vijn, P. & Molenaar, I.W. (1981) Robustness regions for dichotomous decisions, *Journal of Educational Statistics*, 6, 205-235.
- Wilkinson, G.N. (1977) On resolving the controversy in statistical inference (with discussion), *Journal of the Royal Statistical Society B38*, 119-171.
- Zellner, A. (1971) *An introduction to Bayesian inference in econometrics*, Wiley, New York.