

ROBUUSTHEID EN ZICH AANPASSENDE
VERDELINGSVRIJE PROCEDURES ⁽¹⁾W. Albers ⁽²⁾

Samenvatting: Eerst wordt iets gezegd over robuuste procedures in het algemeen, met name over hoe deze zich tot de klassieke procedures verhouden. Vervolgens wordt meer in het bijzonder gekeken naar zich aanpassende procedures. Het blijkt dat dergelijke procedures bruikbaar kunnen zijn als de mate van aanpassing binnen redelijke grenzen gehouden wordt. Van dergelijke zich beperkt aanpassende procedures wordt tenslotte een aantal voorbeelden behandeld, voornamelijk op verdelingsvrij gebied.

1. Robuustheid

In de klassieke statistiek wordt voor een probleem een nauw omschreven model opgesteld. Binnen dat model wordt vervolgens de optimale procedure (bv. de schatter met de kleinste variantie of de toets met het hoogste onderscheidend vermogen) gezocht. Mogelijke andere procedures worden beoordeeld op hun relatieve doeltreffendheid ten opzichte van deze optimale procedure.

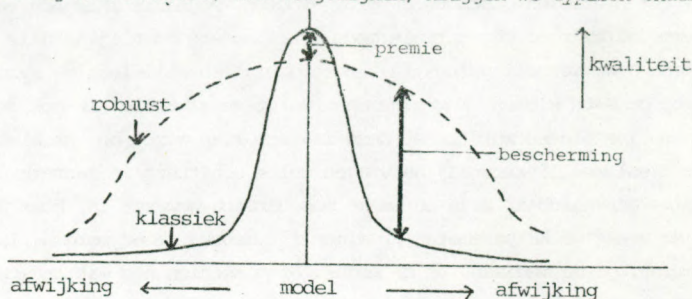
Een standaard voorbeeld betreft het schatten van de verwachting van een normale verdeling. Hier zijn X_1, \dots, X_N onderling onafhankelijk (o.o.) en ze hebben alle dezelfde verdelingsfunctie F , waarbij bovendien wordt verondersteld dat $F(x) = \Phi((x-\theta)/\tau)$ voor zekere $\tau > 0$ en θ , waarin Φ de standaard normale verdelingsfunctie voorstelt. Het steekproefgemiddelde \bar{X} is, zoals bekend, in deze situatie de ideale schatter voor θ .

Het probleem met de hierboven geschetste aanpak is echter dat de model aannamen meestal op z'n best bij benadering vervuld zijn. Op zich is dat nog niet zo erg, maar het blijkt bovendien dat vele klassieke procedures nogal gevoelig zijn voor afwijkingen van de model aannamen, m.a.w. ze zijn niet erg robuust. In het licht hiervan heeft het zin om te zoeken naar procedures die wel robuust zijn terwijl hun relatieve doeltreffendheid t.o.v. de optimale procedure voldoende hoog is.

(1) Dit is gebaseerd op een lezing gehouden op de "Mathemedistica", een gezamenlijke bijeenkomst van de Mathematisch-Statistische en de Medisch-Biologische Secties van de VVS, op 18 maart 1981, te Amsterdam, met als onderwerp "Verdelingsvrije methoden nog steeds actueel?"

(2) Technische Hogeschool Twente
Onderafdeling der Toegepaste Wiskunde

In navolging van Bickel (1976) kunnen we de situatie als volgt karakteriseren: om het model heen veronderstellen we een "supermodel" waarin bepaalde mogelijk geachte afwijkingen opgenomen zijn. Als we de kwaliteit van een klassieke en een robuuste procedure (b.v. het onderscheidend vermogen bij een toets of 1/variantie bij een schatter) dan vergelijken, krijgen we typisch onderstaand plaatje:



Als het model juist is, kost de robuuste procedure ons een zekere premie, in ruil waarvoor we een zekere bescherming genieten tegen de afwijking beschreven door het supermodel. Van geval tot geval zal men moeten afwegen welke premie redelijk is voor welke bescherming tegen welke afwijking.

Om een concreet voorbeeld te geven keren we terug naar het eerder genoemde voorbeeld van het schatten van de verwachting van een normale verdeling. Een supermodel dat hier vaak (zie o.a. Huber (1972 en Hampel (1978)) van toepassing blijkt te zijn, is het "grove-fouten" model. Hier wordt voor de verdelingsfunctie F gekozen

$$F(x) = (1-\epsilon) \phi\left(\frac{x-\theta}{\tau}\right) + \epsilon H(x-\theta),$$

waarbij ϵ een klein positief getal is en H hoort bij een symmetrische verdeling met zwaardere staarten dan die van de door $\phi((x-\theta)/\tau)$ gegeven normale verdeling. De waarnemingen komen dus over het algemeen nog uit een en dezelfde normale verdeling, maar een kleine fraktie is afkomstig uit een duidelijk afwijkende verdeling. Deze situatie kan ontstaan door storingen in gevoelige meetapparatuur, door afleesfouten (koma op de verkeerde plaats), etc..

Het zal duidelijk zijn dat \bar{X} , de optimale schatter onder het normale model, niet robuust is onder dit supermodel: door middel van één enkele voldoende extreme uitschieter kan \bar{X} iedere gewenste- of liever ongewenste - waarde gegeven worden. Een robuust alternatief is hier het zogeheten " α -getrimde gemiddelde" \bar{X}_α , dat verkregen wordt door de 100 α % kleinste en de 100 α % grootste waarnemingen te verwijderen en het gemiddelde van de resterende waarnemingen te nemen. Dus voor $\alpha = 0$ krijgen we \bar{X} terug en voor $\alpha \rightarrow \frac{1}{2}$ krijgen we de mediaan. Naarmate α groter is, is \bar{X}_α minder gevoelig voor uitschieters. Om een indruk te geven van de premie die hiervoor betaald moet worden, merken we op dat onder een exakt normaal model de relatieve doeltreffendheid van $\bar{X}_{0.05}$ en $\bar{X}_{0.10}$ t.o.v. \bar{X}

respektievelijk 97% en 94% is. Dit zijn weliswaar de asymptotische resultaten, maar de Monte Carlo schattingen stemmen hier voor $N = 20$ al keurig mee overeen (zie Huber (1972), blz. 1064). Merk op dat bv. $\bar{X}_{0,05}$ voor $N = 20$ neerkomt op het weglaten van de grootste en van de kleinste waarde, waarna het gemiddelde van de resterende 18 waarden genomen wordt.

Het hierboven beschouwde "grove-fouten" model is maar één van de mogelijke supermodellen rond het normale model. Een andere mogelijkheid is bijvoorbeeld om voor F de verdelingsfunctie behorend bij een willekeurige symmetrische verdeling om θ te kiezen. Nog algemener wordt de situatie als ook de symmetrie-aanname losgelaten wordt. Met deze laatste stap wordt het probleem echter wel essentieel moeilijker. Bij de vragen welke schatters in aanmerking komen, welke premies aanvaardbaar zijn en welke bescherming gewenst is, komt dan namelijk nog de vraag welke parameter er eigenlijk geschat moet worden. Is dat de verwachting, of de mediaan, of de modus, of misschien nog wat anders? Bij een symmetrische verdeling speelt dit probleem niet, omdat daar het centrum van de verdeling ondubbelzinnig bepaald is.

Tot slot van deze paragraaf wijzen we er op dat in alle bovengenoemde gevallen het normale model op de normaliteit aangevallen werd. Het is ook mogelijk (en wenselijk) om zich af te vragen wat er gebeurt als X_1, \dots, X_N niet o.o. zijn, of niet identiek verdeeld. Dit geeft weer aanleiding tot andere klassen van supermodellen.

2. Zich aanpassende procedures

Als we nog even terugkijken naar het plaatje in de vorige paragraaf, dan zien we dat het het meest aantrekkelijk zou zijn om de stippellijn aan te houden met evenwel een uitstapje omhoog via de doorgetrokken lijn in de buurt van het model. Met andere woorden, als het model (vrijwel) juist is zouden we de optimale procedure willen gebruiken, terwijl we de robuuste procedure alleen willen toepassen als het model niet klopt. Op zich weten we natuurlijk niet in welk van deze situaties we verkeren, maar misschien kunnen we daarvan een idee krijgen door vooraf een blik op de waarnemingen te werpen. Bijvoorbeeld, als deze er normaal uitzien, gebruiken we gewoon \bar{X} om θ te schatten, en alleen als de waarnemingen een verdachte indruk maken, nemen we onze toevlucht tot een getrimd gemiddelde. Kortom, het lijkt een aantrekkelijk idee om een zich aanpassende (engels: adaptive) procedure te gebruiken. Merk op dat dit volkomen geoorloofd is, mits we bij het bepalen van de verdeling van de korresponderende schatter of toetsingsgrootheid ook rekening houden met het feit dat en de manier waarop we van te voren naar de waarnemingen gekeken hebben.

Een voorbeeld uit de literatuur dat nauw aansluit bij het bovenstaande is dat van Jaeckel (1971). Hierin wordt als schatter een zich aanpassend getrimd gemiddelde \bar{X}_α gebruikt, waarin $\hat{\alpha} = \hat{\alpha}(X_1, \dots, X_N)$ zo gekozen wordt dat de geschatte variantie van \bar{X}_α minimaal is.

Voor een tweede voorbeeld beschouwen we de situatie van het één-steekproef probleem: X_1, \dots, X_N zijn o.o., alle met verdelingsfunctie $F(x-\theta)$, waarin de dichtheid f van F symmetrisch is om 0. Het gaat er dan om $H_0: \theta = 0$ tegen bv. $H_1: \theta > 0$ te toetsen. Omdat F onbekend is wordt een rangtoets gebruikt. De toetsingsgrootte S_N hiervan wordt als volgt ingevoerd: laat $0 < Z_1 < \dots < Z_N$ de geordende rij der $|X_1|, \dots, |X_N|$ zijn en laat R_1, \dots, R_N de bijbehorende rij van rangnummers zijn, dus $|X_{R_j}| = Z_j$. Dan definiëren we voor iedere zogeheten score functie J op $(0,1)$:

$$S_N = \sum_{j=1}^N J\left(\frac{R_j}{N+1}\right) u(X_j),$$

waarin $u(x) = 1$ voor $x > 0$ en $u(x) = 0$ elders. Als J konstant is, telt S_N simpelweg het aantal positieve X_j 's en hebben we de tekentoets. Als $J(t) = t$ geeft S_N de som van de rangnummers van de positieve waarnemingen, wat de rangteken-toets van Wilcoxon oplevert. Een derde voorbeeld betreft de van der Waerden toets, waarbij $J(t) = \Phi^{-1}([1+t]/2)$, met Φ^{-1} de inverse van Φ .

Rangtoetsen zijn verdelingsvrij, dus vinden we ongeacht welke F de juiste is, altijd dezelfde onbetrouwbaarheid α . Het onderscheidend vermogen hangt echter wel af van F : er blijkt bij elke F een optimale score functie J , en dus ook een optimale rangtoets, te zijn. Zo is de tekentoets het best voor een dubbel-exponentiële verdeling, de toets van Wilcoxon voor een logistische verdeling en de van der Waerden toets voor een normale verdeling. Gezien deze situatie lijkt een zich aanpassende procedure ook hier een aantrekkelijke mogelijkheid: schat eerst met behulp van de waarnemingen de onderliggende F en gebruik dan de S_N met de voor die F optimale J .

Als we ons vervolgens afvragen of de procedures uit de bovenstaande voorbeelden bevredigend werken, dan blijkt het antwoord in het eerste geval bevestigend en in het tweede geval ontkennend te zijn. Jaeckel's procedure heeft bv. voor $N = 20$ een relatieve doeltreffendheid van 90% t.o.v. \bar{X} onder het normale model. Bij de zich aanpassende rangtoetsen lijken echter zulke enorme steekproeven nodig te zijn voordat de asymptotische optimaliteit zichtbaar wordt, dat ze voor praktisch gebruik volmaakt ongeschikt zijn (zie o.a. Huber (1972), blz. 1058 en Hájek en Šidák (1967), blz. 259). Als verklaring ligt voor de hand dat de aanpak in het tweede geval door de immense keuzevrijheid te ambitieus is. Het eerste voorbeeld is veel bescheidener, daar wordt slechts een enkele parameter, te weten het trimmingspercentage, aangepast. De moraal lijkt dus dat zich aanpassende procedures behalve intuïtief aantrekkelijk ook best praktisch bruikbaar kunnen zijn, maar dat dan wel de mate van aanpassing binnen redelijke grenzen moet blijven.

3. Beperkte aanpassing; voorbeelden

In het licht van het voorafgaande zullen we verder alleen nog zich aanpassende procedures beschouwen waarbij de aanpassing tot een eindig aantal parameters beperkt blijft. Het onderliggende supermodel kan dan gekarakteriseerd worden door (θ, τ) , waarin θ de te schatten of te toetsen parameter en $\tau = (\tau_1, \dots, \tau_m)$ de aan te passen parameter is. Binnen dit supermodel begrepen is het model, gekarakteriseerd door (θ, τ_0) , waarbij τ_0 een bekende waarde, vaak 0, van τ is. Met deze formulering zijn we overigens weer bij een klassiek probleem terug: onder het model hebben we alleen te maken met de parameter θ , maar onder het supermodel worden we geconfronteerd met de storingsparameter (engels: nuisance parameter) τ . Voorbeelden van problemen waarin dergelijke storingsparameters een rol spelen zijn er te over. Het meest bekend is waarschijnlijk het probleem van het toetsen van de verwachting θ van een normale verdeling terwijl de variante τ^2 gelijk is aan een bekende waarde τ_0^2 , dan wel onbekend is. In het eerste geval kan de toets gebaseerd worden op \bar{X}/τ_0 , in het tweede geval dient de t-toets toegepast te worden.

Om te beginnen keren we nu terug naar het tweede voorbeeld uit de vorige paragraaf, dat van de rangtoetsen met zich aanpassende scorefunctie J . Beperkte aanpassing leidt hier tot de volgende opzet. Kies een verdelingsfunctie F_0 voor onder het model, b.v. $F_0 = \Phi$, dan fungeert als basis-scorefunctie de hierbij behorende J_0 , bv. $J_0(t) = \Phi^{-1}([1+t]/2)$. Kies ook een supermodel, bv. inhoudend afwijking van normaliteit in de richting van zwaardere staarten, en neem een daarbij behorende scorefunctie als korrektiemogelijkheid, bv. $h(t) = t$ of $h(t) = 1$. Gebruik dan een toets met scorefunctie $J_{\hat{\tau}} = J_0 + \hat{\tau}h$, waarin $\hat{\tau}$ een schatter is die grotere waarden aanneemt naarmate de afwijking van het model groter is en die onder het model consistent is voor τ_0 (hier $\tau_0 = 0$). Deze aanpak is gebruikt in Albers (1979) en Albers (1980).

De kwaliteit van de resulterende aanpassingstoets $\psi_{\hat{\tau}}$ en die van de optimale toets ψ_0 worden onder het model als volgt vergeleken. Als er N waarnemingen voor ψ_0 worden gebruikt, laat dan $N + d_N$ het aantal zijn dat voor $\psi_{\hat{\tau}}$ nodig is om hetzelfde onderscheidend vermogen te bereiken. De premie is dus het extra aantal waarnemingen d_N . Het blijkt dat er redelijk simpele schatters $\hat{\tau}$ gekonstrueerd kunnen worden waarvoor $d = \lim_{N \rightarrow \infty} d_N$ eindig is. Merk op dat dit tweede orde resultaat sterker is dan de uitspraak dat $\psi_{\hat{\tau}}$ en ψ_0 asymptotisch even doeltreffend zijn. Met dat laatste wordt namelijk bedoeld dat $(N+d_N) / N \rightarrow 1$ voor $N \rightarrow \infty$, wat nog bepaald niet inhoudt dat d_N een eindige limiet heeft. Voor concrete gevallen kan d berekend worden. In het bovengenoemde geval met J_0 de van der Waerden score functie en h die van Wilcoxon, is d gelijk aan 4.0, terwijl als h die voor de tekentoets is, we $d = 6.7$ vinden.

Hierna gaan we in op situaties waarin het supermodel niet (alleen) een afwijking van normaliteit inhoudt, maar (ook) een afwijking van de aanname van onafhankelijkheid. Ook hier geldt dat beperking nodig is: willekeurige afhanke-

lijkheid laat een veel te grote keuzevrijheid. We beperken ons daarom in eerste instantie tot m -afhankelijkheid: in een rij X_1, X_2, \dots zijn X_i en X_j onafhankelijk als $|i-j| > m$. Verder nemen we om te beginnen ook normaliteit aan: we beschouwen een steekproef X_1, \dots, X_N uit een stationair normaal proces met als marginale verdelingsfunctie $\phi((x-\theta)/\sigma)$. In deze situatie is m -afhankelijkheid equivalent met een voortschrijdend gemiddelde (engels: moving average) model, d.w.z. $X_i = \sum_{k=0}^m b_k Z_{i-k}$, waarin de b_k onbekende konstanten zijn en de Z_i o.o. normaal verdeeld. In deze situatie kan $H_0 : \theta = 0$ tegen $H_1 : \theta > 0$ getoetst worden met behulp van $\bar{X} / \hat{\sigma}(\bar{X})$, waarin $\hat{\sigma}(\bar{X})$ een geschikte schatter voor de spreiding $\sigma(\bar{X})$ van \bar{X} is. Bij onafhankelijkheid ($m=0$) neemt men uiteraard $\hat{\sigma}(\bar{X}) = N^{-1/2}S$, met $S^2 = \sum_{i=1}^N (X_i - \bar{X})^2 / (N-1)$, wat leidt tot de t -toets. Voor m positief kan gebruikt worden $\hat{\sigma}(\bar{X}) = N^{-1/2}S(1 + 2\sum_{k=1}^m \hat{\tau}_k)^{1/2}$, waarin de $\hat{\tau}_k$ de zogeheten seriële korrelatiecoëfficiënten zijn (zie Anderson (1971)), die op hun beurt de korrelaties τ_k tussen X_i en X_{i+k} schatten. Zo geformuleerd past ook dit geval in de eerder gegeven algemene beschrijving met (θ, τ) onder het supermodel en (θ, τ_0) onder het model. Het model houdt hier in dat $\tau_0 = 0$, we hebben dan onafhankelijkheid en de optimale toets ψ_0 is de t -toets. Deze is niet robuust tegen m -afhankelijkheid, de afwijking in het supermodel: de feitelijke onbetrouwbaarheid zal verschillen van de gewenste α , zelfs asymptotisch. De aanpassingstoets $\psi_{\hat{\tau}}$ heeft echter wel asymptotisch de juiste α .

Net als in het vorige voorbeeld kan de asymptotische premie d berekend worden. Die blijkt hier gelijk te zijn aan $m u_{\alpha}^2$, waarin $u_{\alpha} = \Phi^{-1}(1-\alpha)$ (bv. $u_{\alpha} = 1.65$ voor $\alpha = 0.05$ en $u_{\alpha} = 2.33$ voor $\alpha = 0.01$). Dus bescherming tegen m -afhankelijkheid kost asymptotisch per te schatten korrelatiecoëfficiënt u_{α}^2 extra waarnemingen. Details over een en ander zijn te vinden in Albers (1978a). Hierin wordt ook het pendant van bovenstaand geval besproken, namelijk dat van het autoregressieve model. Hier is de rol van de X_i en de onbekende Z_i verwisseld: nu is $Z_i = X_i + \sum_{k=1}^m \tau_k X_{i-k}$, waarin de τ_k onbekende konstanten zijn. X_i is dus op te vatten als een lineaire combinatie van zijn m directe voorgangers plus een onafhankelijke storingsterm Z_i . Ook voor dit geval blijkt d gelijk te zijn aan $m u_{\alpha}^2$.

Tot slot beschouwen we kort bovenstaand autoregressief geval wanneer bovendien de normaliteitsaannname vervalt. Omdat de Z_i o.o. zijn en symmetrisch verdeeld om $\theta(1 + \sum_{k=1}^m \tau_k)$, kan nu $H_0 : \theta = 0$ tegen $H_1 : \theta > 0$ in principe getoetst worden met behulp van een rangtoets gebaseerd op Z_1, \dots, Z_N in plaats van op X_1, \dots, X_N . In principe, want de Z_i zijn immers onbekend. Er zijn echter schatters $\hat{\tau}_k$ voor de autoregressieve konstanten τ_k bekend (zie weer Anderson (1971)), zodat we wel kunnen beschikken over de $\hat{Z}_i = X_i + \sum_{k=1}^m \hat{\tau}_k X_{i-k}$. Het blijkt (zie Albers (1978 b)) dat de rangtoets gebaseerd op $\hat{Z}_1, \dots, \hat{Z}_N$ in eerste orde equivalent is met de rangtoets gebaseerd op Z_1, \dots, Z_N . In het bijzonder is de toets dus asymptotisch verdelingsvrij. Misschien bestaat intuïtief de indruk dat dit resultaat vanzelfsprekend is omdat het al dan niet kennen van een storingsparameter asymp-

totisch immers niet zou uitmaken. Deze redenering gaat echter alleen op in speciale gevallen, waarbij θ en τ in een bepaalde zin loodrecht op elkaar staan, zoals bijvoorbeeld bij plaats- en schaalparameters in een symmetrische verdeling het geval is. In het algemeen geldt dit niet. Tegenvoorbeelden zijn onder andere het toetsen of een verdeling symmetrisch is wanneer het mogelijke symmetriepunt al dan niet gegeven is en het toetsen of er een verschuiving in de verwachtingswaarde van een verdeling is opgetreden binnen een steekproef wanneer het mogelijke omslagpunt al dan niet gegeven is.

Literatuurlijst

- Albers, W (1978 a). "Testing the mean of a normal population under dependence". Ann. Statist. 6, 1337-1344.
- Albers, W. (1978 b). "One-sample rank tests under autoregressive dependence". Ann. Statist. 6, 836-845.
- Albers, W. (1979). "Asymptotic deficiencies of one-sample rank tests under restricted adaptation". Ann. Statist. 7, 944-954.
- Albers, W. (1980). "Asymptotic expansions for the power of adaptive rank tests in the one-sample problem". In: "Statistique non Paramétrique Asymptotique (ed. J.P. Raoult), Springer Lecture Notes in Math. 821, 108-158.
- Anderson, T.W. (1971). "The statistical analysis of time series". Wiley, New York.
- Bickel, P.J. (1976). "Another look at robustness: a review of reviews and some new developments". Scand. J. Statist. 3, 145-168.
- Hájek, J. and Šidák, Z. (1967). "Theory of rank tests". Academic Press, New York.
- Hampel, F. (1978). "Modern trends in the theory of robustness". Math. Operationsforsch. Statist., Ser. Statistics, 9, 425-442.
- Huber, P.J. (1972). "Robust statistics: a review". Ann. Math. Statist. 43, 1041-1067.
- Jaeckel, L.A. (1971). "Some flexible estimates of location". Ann. Math. Statist. 42, 1540-1552.