KM 4(1981)
pag 73-85

OPTIMALE GEWICHTEN VOOR "EENZIJDIGE" TOETSINGSPROBLEMEN BIJ KRUISTABELLEN

Tom Snijders*
Mathematisch Instituut, R.U. Groningen.

Dit artikel is een bewerking van de tekst van een voordracht, gehouden op een gezamenlijke bijeenkomst van de Mathematische en Medisch-Biologische Sectie van de V.V.S. op 18 maart 1981, gewijd aan verdelingsvrije statistiek.

1. *Samenvatting*

Bij toepassingen van de statistiek komt men regelmatig waarnemingsgegevens tegen die worden samengevat in een tweedimensionale kruistabel - deze zullen we verder aangeven als een k × m tabel - en waarop toetsen worden losgelaten voor homogeniteit of onafhankelijkheid. De meest gebruikte toets hierbij is wel de χ^2 toets. In dit artikel gaat het over situaties waar de kategorieën van de k × m tabel geordend zijn, en men de nulhypothese van homogeniteit, dan wel onafhankelijkheid, liever wil toetsen tegen een ingeperkt (één- of tweezijdig) alternatief dan tegen het "alzijdige" alternatief waartegen de χ^2 toets vooral geschikt is. Dan gebruikt men vaak (stijgende) gewichten voor de kategorieën, bijvoorbeeld ekwidistante gewichten of gemiddelde rangnummers. In dit artikel zullen gewichten worden gegeven die in zekere zin optimaal zijn, en enkele voors en tegens van deze "optimale" gewichten zullen worden besproken.

2. *Formulering van de beschouwde toetsingsproblemen*

In dit artikel zal het gaan over het toetsen van homogeniteit in het

*Nu werkzaam bij het Econometrisch Instituut, Postbus 800, 9700 AV Groningen.

k-steekproevenprobleem bij één waar te nemen variabele met m mogelijke, geordende, uitkomsten; en over het toetsen van onafhankelijkheid tussen twee variabelen met k respektievelijk m mogelijke, geordende, uitkomsten. Het toetsen van univariate en bivariate symmetrie in een $1 \times k$, respektievelijk $k \times k$ tabel wordt behandeld in Schaafsma (1966), hoofdstuk 10 en Snijders (1979), §9.5. Voor de behandelde toepassingsproblemen zal een uniforme notatie worden gebruikt waarin de $k \times m$ tabel wordt aangegeven als de matrix

$$(n_{ij})_{1 \leq i \leq k, 1 \leq j \leq m}$$

met randtotalen

$$n_{i+} = \Sigma_j n_{ij}, \qquad n_{+j} = \Sigma_i n_{ij}$$

en met totaal

$$n = \Sigma_{i,j} n_{ij}.$$

2.1 *k steekproeven-problemen*

Bij k steekproeven-problemen geeft n_{ij} het aantal malen aan dat in de i^e steekproef uitkomst j wordt waargenomen. De veronderstelling dat het om k onafhankelijke random steekproeven gaat leidt ertoe, dat $(n_{i1},\ldots,n_{im})$ wordt opgevat als de uitkomst van een toevalsvariabele $(N_{i1},\ldots,N_{im})$ met de multinominale verdeling met parameters n_{i+} en

$$p_i = (p_{i1},\ldots,p_{im}),$$

de vektor van kansen waarvoor geldt

$$p_{ij} > 0, \qquad \Sigma_j p_{ij} = 1.$$

De nulhypothese van homogeniteit luidt

$$H : p_1 = p_2 = \ldots = p_k,$$

"alle steekproeven komen uit dezelfde kansverdeling". Bij de precieze omschrijving van de alternatieve hypothese maken we gebruik van het begrip "stochastische ordening". Van twee toevalsvariabelen X en Y met mogelijke uitkomsten 1,2,...,m en vektoren van kansen $q = (q_1,\ldots,q_m)$

en $r = (r_1,\ldots,r_m)$ waarbij

$$q_j = P\{X = j\}, \qquad r_j = P\{Y = j\},$$

zeggen we dat Y stochastisch groter is dan X, wanneer

$$P\{X \geq j\} \leq P\{Y \geq j\} \qquad \text{voor } 1 \leq j \leq m;$$

de verdelingsfunktie van X ligt dus nergens onder die van Y. We geven dit aan door

$$q \leq r.$$

Er zullen twee alternatieve hypothesen worden beschouwd bij deze nulhypothese van homogeniteit. De eerste daarvan is de alternatieve hypothese van een stijgende trend

$$A_1 : p_1 \leq p_2 \leq \cdots \leq p_k \quad \text{en } p_1 \neq p_k.$$

Dit betekent dus, dat de $(i+1)^e$ steekproef komt uit een kansverdeling die stochastisch groter is dan de i^e, voor $1 \leq i \leq k-1$. De tweede alternatieve hypothese is die van trend in een ongespecificeerde volgorde:

A_2 : er bestaat een permutatie $\pi(1),\pi(2),\ldots,\pi(k)$ van de getallen $1,2,\ldots,k$ waarvoor geldt

$$p_{\pi(1)} \leq p_{\pi(2)} \leq \cdots \leq p_{\pi(k)} \quad \text{en } p_{\pi(1)} \neq p_{\pi(k)}.$$

Merk op dat deze tweede alternatieve hypothese, mits $m \geq 3$, nog altijd veel beperkender is dan de niet-ingeperkte alternatieve hypothese die luidt "de kansvektoren $p_1,\ldots,p_k$ zijn niet alle gelijk". Bij A_1 kan het groepsnummer worden opgevat als een ordinale variabele, bij A_2 als een nominale. Om de gedachten te bepalen kan men bij alternatief A_1 denken aan het toetsen van het verschil in werking tussen k verschillende (toenemende) doses van hetzelfde geneesmiddel, terwijl men bij alternatief A_2 kan denken aan het vergelijken van k verschillende geneesmiddelen; waarbij men vooral die afwijkingen van de nulhypothese graag wil ontdekken waarvoor (A_1) een grotere dosis een sterkere werking heeft, of (A_2) het ene middel een sterkere werking heeft dan het andere, zonder dat iets bekend is over de volgorde van de geneesmiddelen.

In de onderstaande tabel staan als voorbeeld gegevens over de afloop van een bepaalde vorm van meningitis, behandeld met penicilline die intramusculair (i=1) of intralumbaal (i=2) werd toegediend. Verderop zullen meer bijzonderheden over de wijze van experimenteren worden gegeven. Uitkomst j=1 betekent "volledige genezing", j=2 betekent "genezing maar met restverschijnselen", j=3 betekent "dodelijke afloop". Op elk van beide manieren werden n_i=90 patiënten behandeld.

		resultaat		
		1	2	3
methode	1	38	36	16
	2	42	44	4

2.2 *Onafhankelijkheidsproblemen*

Bij onafhankelijkheidsproblemen gaat het hier over een random steekproef ter grootte n uit een bivariate kansverdeling van een toevalsvariabele (X,Y). Het aantal mogelijke uitkomsten van X en Y is gelijk aan k, respektievelijk m. Hier geeft n_{ij} het aantal malen aan dat uitkomst (i,j) van (X,Y) wordt waargenomen.
De nulhypothese van onafhankelijkheid luidt

$$H : P\{X = i, Y = j\} = P\{X = i\}P\{Y = j\} \quad \text{voor alle } (i,j),$$

"X en Y zijn onafhankelijk". Ook hier worden twee alternatieve hypothesen beschouwd. De eerste daarvan is de alternatieve hypothese van positieve regressie-afhankelijkheid van Y op X:

$$A_3 : P\{Y \geqq j \mid X = i\} \text{ is een stijgende funktie van i, voor alle j.}$$

Deze vorm van afhankelijkheid wordt besproken in Lehmann (1959, 1966). Er is geen symmetrie tussen de rol van X en Y, maar Y wordt beschouwd als de "afhankelijke" en X als de "onafhankelijke variabele". Dit toetsingsprobleem lijkt bijzonder veel op het eerste toetsingsprobleem uit §2.1.

De tweede alternatieve hypothese is die van positieve kwadrantafhankelijkheid, of positieve associatie:

A_4 : $P\{X \geqq i, Y \geqq j\} \geqq P\{X \geqq i\}P\{Y \geqq j\}$ voor alle (i,j) met strikte ongelijkheid voor sommige (i,j).

Deze vorm van afhankelijkheid wordt besproken in Lehmann (1966) en Schaafsma (1966). Hierbij is er wel symmetrie tussen de rol van X en Y. Het is niet moeilijk in te zien, dat de ongelijkheid in de definitie van A_4 kan worden vervangen door

$$P\{X \leqq i, Y \leqq j\} \geqq P\{X \leqq i\}P\{Y \leqq j\},$$

zonder dat dit enig verschil maakt. In de genoemde referenties kan men vinden dat regressie-afhankelijkheid (A_3) impliceert dat positieve associatie (A_4) geldt.

3. *Lineaire toetsen*

Voor de eenzijdige toetsingsproblemen met de alternatieven A_1, A_3 en A_4 uit §2 worden, meestal zonder de alternatieve hypothese precies te omschrijven, vaak toetsen gebruikt die de nulhypothese verwerpen voor grote waarden van een toetsingsgrootheid van de vorm

$$\Sigma_{i,j} a_{ij} N_{ij}$$

(stochastische variabelen worden aangegeven met hoofdletters, hun uitkomsten met kleine letters). Hierbij is het getal a_{ij} een "gewicht" dat aan cel (i,j) van de k × m tabel wordt gegeven. De gewichten en/of de kritieke waarde zullen van de randtotalen afhangen. Dergelijke toetsen zullen we lineaire toetsen noemen. Beschouw bijvoorbeeld het twee steekproeven-probleem dat ontstaat door in §2.1 te nemen k = 2 en alternatief A_1. Twee veel gebruikte toetsen hiervoor (naast de χ^2 toets, die bij dit alternatief eigenlijk niet op zijn plaats is) zijn de Student t-toets waarbij men de uitkomsten vervangt door hun nummers, dus 1,2,...,m, of door andere vooraf bepaalde gewichten; en de Wilcoxon toets met gemiddeld rangnummer en knoopkorrektie. Beide toetsen hebben een toetsingsgrootheid van de vorm

$$(3.1) \quad (S^2)^{-\frac{1}{2}} \Sigma_{j=1}^{m} a_j \left(\frac{N_{2j}}{n_2} - \frac{N_{1j}}{n_1}\right)$$

met

$$(3.2)\quad S^2 = \frac{n}{n_1 n_2 (n-1)} \Sigma_{j=1}^{m} N_{+j} (a_j - a.)^2$$

$$a. = n^{-1} \Sigma_{j=1}^{m} N_{+j} a_j \quad .$$

Deze toetsingsgrootheid heeft asymptotisch (voor grote n) onder de nulhypothese een normale verdeling. Bij de eerstgenoemde toets (Student) is $a_j = j$, bij de tweede (Wilcoxon) is

$$(3.3)\quad a_j = \Sigma_{h=1}^{j-1} N_{+h} + \tfrac{1}{2}(N_{+j}+1) \quad .$$

Veel gehoorde argumenten voor deze keuzes voor de gewichten zijn hun eenvoud (die bij de gemiddelde rangnummers ook tot uiting komt in de eenvoudige formule voor S^2, met de bekende knoopkorrektie), "wat zou je anders doen" en (minder vaak) "het maakt toch niets uit hoe je de gewichten kiest".

Een andere methode die vaak wordt gebruikt is het konstrueren van modellen die de klasse van kansverdelingen onder het alternatief verder inperken, zodat er onder het alternatief één (reëelwaardige) parameter θ meer is dan onder de nulhypothese. Men probeert dan een inhoudelijk interessante parameter θ te konstrueren. Het toetsen van de nulhypothese komt dan neer op het toetsen van "$\theta = 0$". Zie hiervoor bv. Mc Cullagh (1980). Deze benadering kan heel zinvol zijn, vooral wanneer men (bijna) zeker meent te weten dat de nulhypothese niet geldt; het toetsen daarvan neemt dan een ondergeschikte (soms onmisbare en soms overbodige) plaats in. Hierbij moet men overigens oppassen niet de kunstfout te begaan om op grond van dezelfde kruistabel een "goed passend" model te kiezen èn de nulhypothese van homogeniteit (of onafhankelijkheid) in bovengenoemde vernauwde vorm "$\theta = 0$" te toetsen; dan kiest men namelijk een toets die bij de gegeven uitkomst extra gauw tot verwerping leidt, en kan men niet volhouden dat er getoetst wordt op het opgegeven signifikantienivo α. Ook McCullagh (op cit.) kan aan deze verleiding niet weerstaan.

4. *"Optimale" gewichten*

Voor deze toetsingsproblemen bestaan er (tenzij $k = m = 2$) geen exakt of asymptotisch (voor $n \to \infty$) meest onderscheidende toetsen van nivo α. Dit hangt er mee samen, dat het alternatief meer dan 1 extra parameter heeft ten opzichte van de nulhypothese. Wanneer men toch binnen een objectivistisch kader een "optimale" toets probeert te vinden, kan men een compromiskriterium hanteren dat tot een toets leidt die van zeker standpunt gezien een goed onderscheidingsvermogen heeft, zonder dat het onderscheidingsvermogen in sommige parameterwaarden al te slecht wordt. Het kriterium van de strengste (most stringent) toets is zo'n compromiskriterium. Het is afkomstig van Abraham Wald (1942); wie iets wil weten over de definitie en de meest bekende eigenschappen raadplege Lehmann (1959), hoofdstuk 8. De te noemen gewichten leveren toetsen die binnen de klasse van lineaire toetsen (voor het toetsen van homogeniteit tegen alternatief A_2 de klasse van "kwadratische" toetsen, zie verderop) een optimaliteitseigenschap hebben, EAMS genaamd, die een asymptotische versie is van het kriterium van de strengste toets, speciaal ontworpen voor toetsingsproblemen met een ingeperkt alternatief. Voor de definitie hiervan, en de noodzakelijke bewijzen, verwijs ik u naar Snijders (1979). (Wie de euvele moed heeft dat boekje te raadplegen, raad ik aan te beginnen met hoofdstuk 9; daar staan een paar dingen die lijken op dit artikel).

Deze gewichten zijn bepaald door Schaafsma (1966). Hierbij wordt gebruik gemaakt van Abelson & Tukey (1963), waarin het begrip "maximin kontrast" wordt ingevoerd. Dit begrip is in de toetsingstheorie vooral bruikbaar bij meer-dimensionale ingeperkte alternatieven, zoals die voorkomen in de hier behandelde toetsingsproblemen. Andere toepassingen van maximin kontrasten worden gegeven door Hirotsu (1978) voor het toetsen tegen interaktie bij variantie-analyse, en Stuart (1980) voor het toetsen van goodness of fit.

De lineaire toets met EAMS gewichten voor het toetsen van homogeniteit tegen een stijgende trend (alternatief A_1) gebruikt de toetsingsgrootheid

$$(4.1) \qquad (S^2)^{-\frac{1}{2}} \Sigma_{i=1}^{k} b_i \Sigma_{j=1}^{m} a_j \frac{N_{ij}}{n_{i+}}$$

waarin

$$a_1 = 0,\ a_j - a_{j-1} = \{\frac{n}{N_{+,j-1}} + \frac{n}{N_{+j}}\}^{\frac{1}{2}}$$

$$b_i = \{c_{i-1}(n - c_{i-1})\}^{\frac{1}{2}} - \{c_i(n-c_i)\}^{\frac{1}{2}}$$

$$(4.2) \quad c_i = \Sigma_{n=1}^{i} n_{h+}, \qquad c_o = 0$$

$$S^2 = (n-1)^{-1}\{\Sigma_{i=1}^{k} n_{i+}^{-1} b_i^2\}\{\Sigma_{j=1}^{m} N_{+j}(a_j - a.)^2\}$$

$$a. = n^{-1}\Sigma_{j=1}^{m} N_{+j} a_j \ .$$

Er kan worden opgemerkt dat $a_1 < a_2 < \ldots < a_m$ en $b_1 n_{1+}^{-1} < b_2 n_{2+}^{-1} < \ldots < b_k n_{k+}^{-1}$, terwijl $\Sigma_i b_i = 0$. De laatste eigenschap leidt ertoe dat de keuze voor de oorsprong (hier $a_1 = 0$) voor de a_j er niet toe doet. Deze toetsingsgrootheid kan worden opgevat als een kontrast tussen de gemiddelden der k steekproeven, wanneer aan uitkomst j het getal a_j wordt toegekend. Onder de nulhypothese heeft de toetsingsgrootheid (4.1) asymptotisch voor $n \to \infty$ een standaardnormale verdeling. De normalisatiefaktor S^2 is zo gekozen dat de voorwaardelijke variantie van (4.1), gegeven de randtotalen, onder de nulhypothese gelijk is aan 1.

In het geval k = 2, dus het 2 steekproevenprobleem, is de toetsingsgrootheid (4.1) gelijk aan (3.1) met S^2 als in (3.2) en de gewichten a_j en a. als in (4.2). Een ekwivalente formule voor het geval dat m = 2 en $k \geqq 2$ (toetsen van gelijkheid van k kansen tegen trend) is te vinden in Snijders (1979), §9.2.

Voor het toetsen van homogeniteit tegen trend in een ongespecificeerde richting (alternatief A_2) kan men de aandacht richten op "kwadratische" toetsen met een toetsingsgrootheid van de vorm

$$(4.3) \quad S^{-2}\Sigma_{i=1}^{k} n_{i+}\{\Sigma_{j=1}^{m} a_j(\frac{N_{ij}}{n_{i+}} - \frac{N_{+j}}{n})\}^2$$

waarin

$$S^2 = (n-1)^{-1}\Sigma_{j=1}^{m} N_{+j}(a_j - a.)^2$$

en a. als in (3.2) of (4.2). Neemt men voor a_j het gemiddeld rangnummer (3.3) dan krijgt men de Kruskal-Wallis toets met knoopkorrektie. Ook hierin worden EAMS gewichten geleverd door a_j als in (4.2). Onder de nulhypothese heeft de toetsingsgrootheid (4.3) asymptotisch een χ^2 verdeling met k-1 vrijheidsgraden.

De lineaire toets met EAMS gewichten voor het toetsen van onafhankelijkheid tegen positieve regressie-afhankelijkheid van Y op X (alternatief A_3) wordt weer gegeven door (4.1) en (4.2); pro forma mag men n_{i+} in N_{i+} veranderen. Het was niet best als deze twee, vrijwel gelijke, toetsingsproblemen tot verschillende toetsen zouden leiden.

De lineaire toets met EAMS gewichten voor het toetsen van onafhankelijkheid tegen positieve kwadrant-afhankelijkheid (alternatief A_4) gebruikt de toetsingsgrootheid

$$(4.4)\quad (S^2)^{-\frac{1}{2}}\Sigma_{i=1}^{k}\Sigma_{j=1}^{m}N_{ij}(a_i - a.)(b_j - b.)$$

waarin

$$a_i - a_{i-1} = \{\frac{n}{N_{i-1,+}} + \frac{n}{N_{i+}}\}^{\frac{1}{2}}$$

$$b_j - b_{j-1} = \{\frac{n}{N_{+,j-1}} + \frac{n}{N_{+j}}\}^{\frac{1}{2}}$$

$$S^2 = (n-1)^{-1}\Sigma_{i=1}^{k}N_{1+}(a_i - a.)^2\Sigma_{j=1}^{m}N_{+j}(b_j - b.)^2$$

$$a. = n^{-1}\Sigma_{i=1}^{k}N_{i+}a_i$$

$$b. = n^{-1}\Sigma_{j=1}^{m}N_{+j}b_j \,.$$

Ook hier doet de keuze van de oorsprong voor de a_i en de b_j er niet toe, en kan men bijvoorbeeld nemen $a_1 = b_1 = 0$ (of $a. = b. = 0$, enz.). De toetsingsgrootheid (4.4) heeft onder de nulhypothese asymptotisch weer een standaardnormale verdeling. Na deling door $(n-1)^{\frac{1}{2}}$ kan men (4.4) opvatten als een variant van de rang-korrelatie-coëfficiënt van Spearman, met EAMS gewichten in plaats van de gebruikelijke gemiddelde rangnummers.

5. *Tweezijdige alternatieven*

Van de in §2 omschreven alternatieve hypothesen A_1, A_3 en A_4

kunnen ook tweezijdige versies worden gegeven. Voor A_1 is dat bijvoorbeeld

$$A_1^2 : p_1 \leqq p_2 \leqq \dots \leqq p_k \text{ of } p_1 \geqq p_2 \geqq \dots \geqq p_k;$$
$$\text{en } p_1 \neq p_k,$$

"er is een stijgende of een dalende trend". De omschrijving van de tweezijdige versie van A_4 is "er is positieve of negatieve associatie". Bij deze toetsingsproblemen kan men dezelfde toetsingsgrootheden gebruiken, en hiermee een tweezijdige toets uitvoeren. Deze toets is dan weer "EAMS" in de klasse van alle tweezijdige lineaire toetsen.

6. *Toepassing op het voorbeeld; verschil met de gemiddelde rangnummers als gewichten*

De gewichten a_1, a_2 en a_3 uit (4.2) voor de 2 × 3 tabel aan het eind van §2.1 zijn achtereenvolgens

$$-6.07 \qquad 2.23 \qquad 15.35 \quad .$$

Deze gewichten zijn zo gestandaardiseerd dat $a. = 0$ en $S^2 = 1$. De gemiddelde rangnummers (3.3), met dezelfde standaardisatie, zijn

$$-7.10 \qquad 4.26 \qquad 11.35 \quad .$$

De toetsingsgrootheid (3.1) heeft voor de EAMS gewichten de waarde 2.12 en voor de gemiddelde rangnummers de waarde 1.45. In dit geval levert de EAMS toets dus een behoorlijk signifikant, en de Wilcoxon toets met gemiddelde rangnummers een niet-signifikant resultaat. Dat is beslist niet in alle gevallen zo; van deze twee toetsen heeft geen van beide een uniform hoger onderscheidingsvermogen dan de ander (zie het begin van §4). Dat het in dit geval wel zo is, ligt aan een karakteristiek verschil tussen de EAMS gewichten en de gemiddelde rangnummers. Voor de EAMS gewichten (4.2) geldt

$$a_j - a_{j-1} = \{\frac{n}{N_{+,j-1}} + \frac{n}{N_{+j}}\}^{\frac{1}{2}}$$

en voor de gemiddelde rangnummers (3.3)

$$a_j - a_{j-1} = \tfrac{1}{2}(N_{+,j-1} + N_{+j}).$$

Dit leidt ertoe dat wanneer de uitkomst j weinig voorkomt, dus N_{+j}/n klein is, het EAMS gewicht a_j veel zal verschillen van a_{j-1} en a_{j+1}, terwijl het gemiddeld rangnummer a_j weinig zal verschillen van a_{j-1} en a_{j+1}. Het omgekeerde geldt als uitkomst j veel voorkomt. Als gevolg hiervan is de toets met EAMS gewichten meer gevoelig voor verschillen in kansen op weinig voorkomende uitkomsten, terwijl de toets met gemiddelde rangnummers meer gevoelig is voor verschillen in kansen op veel voorkomende uitkomsten. Welnu, bij de tabel uit §2.1 zit het verschil tussen de twee behandelingsmethoden vooral in de frekwenties van de weinig voorkomende uitkomst j = 3 (overlijden).

Dat het "compromiskriterium" EAMS deze eigenschap moet hebben, kan als volgt worden ingezien. Bij weinig voorkomende uitkomsten j is de kans kleiner dat eventuele verschillen tussen de populaties duidelijk tot uiting komen dan bij veel voorkomende uitkomsten. Om de weinig voorkomende uitkomsten toch nog een "rol van betekenis" te laten spelen, moeten ze in de toetsingsgrootheid een extra gewicht krijgen. Dit gebeurt ook bij de χ^2 toetsingsgrootheid (de χ^2 toets is asymptotisch strengst tegen het niet-ingeperkte alternatief).

7. *Verontschuldiging en diskussie*

Geachte lezer, mijn verontschuldiging! Ik heb u in het ootje willen nemen. Toen ik de voordracht hield waarvan dit artikel een bewerking is, hield onmiddellijk voor mij prof. H. de Jonge een voordracht. Hij had het o.a. over het vergelijken van ekwidistante gewichten $a_j = j$ en gemiddelde rangnummers (3.3) bij het toetsen van homogeniteit met een toetsingsgrootheid van de vorm (3.1). De twee hiermee verkregen toetsen vergeleek hij door ze toe te passen op een groot aantal getallenvoorbeelden, waarvan sommige uit de praktijk kwamen en andere waren gekonstrueerd om "allerlei verschillende gevallen" te krijgen. Het resultaat was dramatisch: er was bijna nooit een noemenswaard verschil tussen de uitkomsten van de twee toetsingsgrootheden. Een ondersteuning dus van het argument, dat de keuze van de gewichten er weinig toe doet. Weliswaar volgt uit het in §6 genoemde verschil tussen de EAMS gewichten en de gemiddelde rangnummers, dat deze twee tot sterker verschillende toetsingsgrootheden zullen kunnen leiden dan ekwidistante gewichten en gemiddelde rangnummers; maar de verschillen zullen meestal niet groot zijn.

Om toch nog enige indruk op u te maken heb ik een getallenvoorbeeld bedacht waarin het verschil tussen de twee vergeleken toetsingsgrootheden wel flink uitpakt, en daarbij een verhaal verzonnen waarin uitkomst $j = 3$, waar het effekt vooral lijkt op te treden, van levensbelang is. Het voorbeeld is dus fiktief en niet representatief.

Wat voor argumenten kunnen o.a. een rol spelen bij de keuze van de gewichten in een lineaire toetsingsgrootheid?
Soms (maar zelden) heeft de onderzoeker inhoudelijke redenen om bepaalde gewichten te kiezen, oftewel een deel-alternatief te kiezen waartegen de toets vooral een goed onderscheidingsvermogen moet hebben. Als die inhoudelijke argumenten niet worden aangevoerd, wordt de keuze gedeeltelijk een kwestie van smaak. Er bestaat immers voor deze toetsingsproblemen geen uniform meest onderscheidende toets, ook niet binnen de klasse van lineaire toetsen (of binnen een andere aardige klasse). Een praktisch argument voor ekwidistante gewichten en ook voor gemiddelde rangnummers is hun ingeburgerdheid. Dat de EAMS gewichten iets ingewikkelder zijn uit te rekenen is sinds de opkomst van de zakrekenmachine geen bezwaar meer (bovendien heb ik een in PASCAL geschreven programma van drs. S. Knijpstra dat de toetsen uitrekent, en dat u op aanvraag kunt krijgen). Wat het onderscheidingsvermogen betreft, volgt uit de opmerkingen in §6 dat de EAMS gewichten extra letten op weinig voorkomende uitkomsten, terwijl bv. de gemiddelde rangnummers extra letten op veel voorkomende uitkomsten. Dit kan ertoe leiden, dat op inhoudelijke gronden voor het ene of voor het andere stel gewichten wordt gekozen - of voor de "tussenliggende" ekwidistante gewichten.

De optimaliteitseigenschap van de EAMS gewichten, het belangrijkste argument om ze te gebruiken, kan in losse bewoordingen als volgt worden benaderd. Bij het voorbeeld uit §2.1 blijkt in §6, dat de statisticus zich een buil zou hebben gevallen wanneer hij gemiddelde rangnummers in plaats van EAMS gewichten had geadviseerd. (Daarbij ga ik er maar van uit, dat dat het effekt echt is (d.w.z. de alternatieve hypothese geldt), of dat de statisticus de onderzoeker aan zich wil verplichten door diens resultaten signifikant te kunnen praten.) Maar er zijn andere voorbeelden te geven waarin de statisticus zich juist een buil zou vallen door EAMS gewichten te adviseren in plaats van gemiddelde rangnummers. Welnu: wanneer we de alternatieve hypothese kiezen zoals in §2 is gedaan, dan is grootste buil die de statisticus zich zou kunnen vallen bij gebruik van de toets met EAMS gewichten, kleiner dan de grootste buil die hij zich zou kunnen vallen bij gebruik van enige andere lineaire toets (grof gezegd). Dit onder de voorwaarde dat de steekproeven niet te klein zijn (minstens lopend in de tientallen). Meer preciese formuleringen zijn te vinden in Snijders (1979).

Referenties

Abelson, R.P. & J.W. Tukey (1963), Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. Ann. Math. Stat. 34, 1347-1369.

Hirotsu, C. (1978), Ordered alternatives for interaction effects. Biometrika 65, 561-570.

Lehmann, E.L. (1959), Testing statistical hypotheses. Wiley and Sons, New York.

Lehmann, E.L. (1966), Some concepts of dependence. Ann. Math. Stat. 37, 1137-1153.

McCullagh, P. (1980), Regression models for ordinal data (with discussion). J. Roy. Stat. Soc. B 42, 109-142.

Schaafsma, W. (1966). Hypothesis testing problems with the alternative restricted by a number of inequalities. Noordhoff, Groningen.

Snijders, T.A.B. (1979), Asymptotic optimality theory for testing problems with restricted alternatives. MC Tracts 113, Mathematical Centre, Amsterdam.

Stuart, M. (1980), Components of χ^2 for testing distributional assumptions against certain restricted alternatives. J. Am. Stat. Ass. 75, 625-633.