STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 47 (M 10)

Symmetrietoets.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
Amsterdam-O.
Statistische Afdeling.
S 47 (M 10)

Symmetrietoets[1].

Hypothese H_0: de waarnemingen $z_1,\dots,z_{n'}$ zijn afkomstig van n' onafhankelijk verdeelde stochastische grootheden, die alle symmetrisch ten opzichte van 0 verdeeld zijn[2]. Van deze toets bestaan meerdere versies $T_1,\dots,T_2''$. We bespreken eerst T_1 en T_2.

Toetsingsgrootheden. Deze worden als volgt uit $z_1,\dots,z_{n'}$ afgeleid:

1e. de waarnemingen, die gelijk aan 0 zijn worden weggelaten. Stel er blijven over: $z_1,\dots,z_n$.

2e. Hieruit worden de positieve waarnemingen gezocht. Stel dit zijn $x_1,\dots,x_{n_1}$, dus n_1 in aantal.

3e. De overblijvende negatieve waarnemingen worden van teken veranderd, zodat zij ook positief worden. Stel dit zijn dan $y_1,\dots,y_{n_2}$.

4e. De grootheden $x_1,\dots,x_{n_1}$ en $y_1,\dots,y_{n_2}$ worden, door elkaar, in afdalende grootte-volgorde opgeschreven. Stel dit geeft: $w_1,\dots,w_n$. (Komen er gelijken voor, dan worden deze in willekeurige volgorde geplaatst.)

5e. De groep waarden $w_1,\dots,w_n$ wordt verdeeld in twee groepen $w_1,w_2,\dots,w_r$ en $w_{r+1},\dots,w_n$, waarbij $w_r \neq w_{r+1}$ is en r zo dicht mogelijk bij de waarde $\frac{1}{2}n$ genomen wordt. Is n even, dan wordt $r=\frac{1}{2}n$, indien althans $w_{\frac{1}{2}n} \neq w_{\frac{1}{2}n+1}$ is. Zijn er twee mogelijk keuzen voor r, beide op gelijke afstand van $\frac{1}{2}n$, dan nemen wij $r > \frac{1}{2}n$. Is b.v. n oneven en $w_{\frac{1}{2}n-\frac{1}{2}} \neq w_{\frac{1}{2}n+\frac{1}{2}}$, dan nemen wij $r=\frac{1}{2}n+\frac{1}{2}$. Wij geven de waarden $w_1,\dots,w_r$ aan als groep A (die dus r elementen bevat) en de overigen als groep B. Alle elementen van A zijn dus groter dan ieder element van B[3].

---

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
2) Zetten wij hier a in plaats van 0, dan geldt H_0 voor $x_1-a,\dots,x_{n'}-a$.
3) In de oorspronkelijke publicaties over deze toets (zie de litteratuurverwijzingen aan het einde van dit memorandum) is een enigzins minder algemene definitie van r gegeven.
Alle stellingen blijven echter gelden, indien de hier gegeven definitie gebruikt wordt.

6e. Het aantal waarden van $x_1, \ldots, x_n$ die in A voorkomen noemen wij u.

De toetsingsgrootheden zijn $\underline{n_1}$ en $\underline{u}$, r is een hulpgrootheid.

<u>V.B.</u> 22 waarden z_i: 7,4/6,3/3,6/3,5/3,4/2,9/2,5/1,1/0/0/-1,3/-2,5/-3,2/-4,6/-4,6/-4,6/-4,8/-6,3/-7,0/-7,9/-8,0/-8,7.

$\therefore$ $r = 11$, $n_1 = 8$
$u = 2$

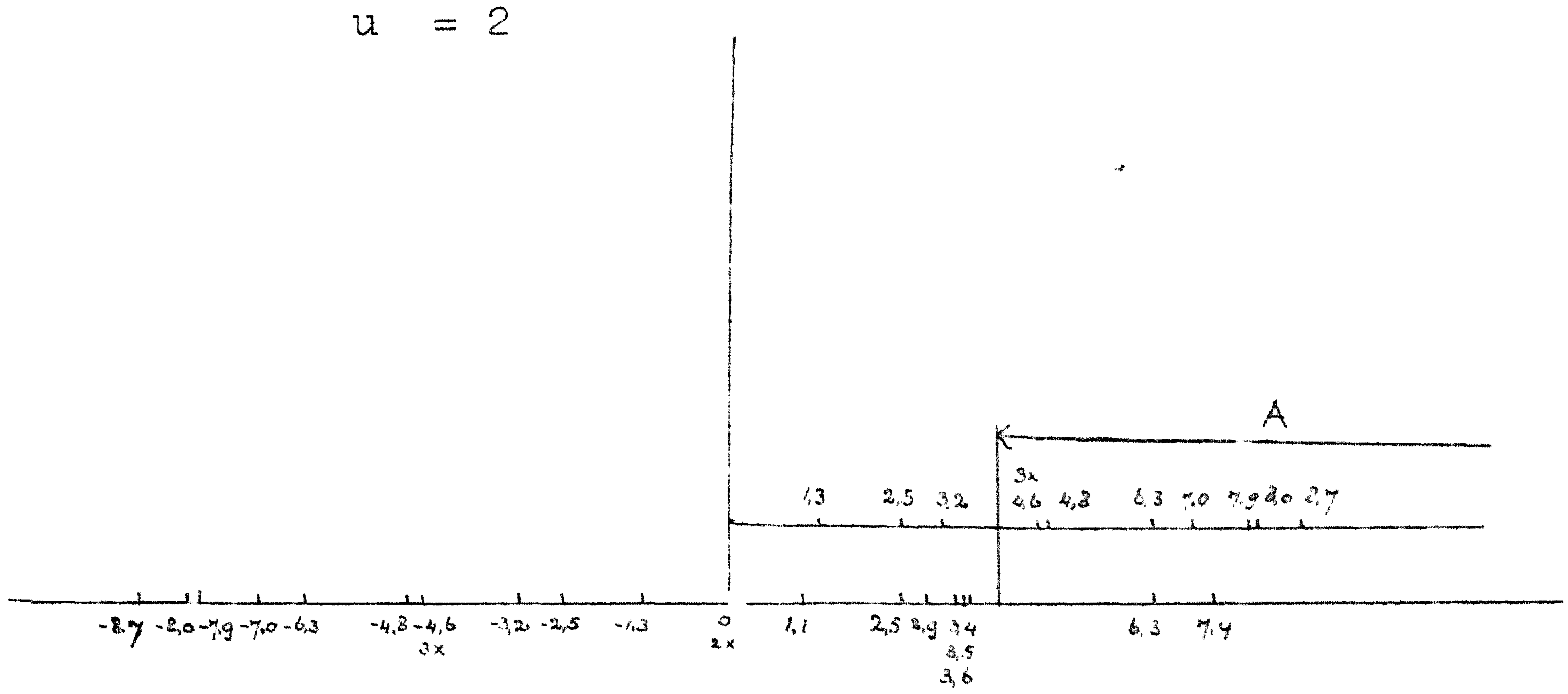

fig. 1

<u>Kritieke zônes</u>. Waarden van n_1, die dicht bij 0 of dicht bij n liggen, zullen, als H_0 juist is weinig, maar als H_0 onjuist is vaker, voorkomen. Grote resp. kleine waarden van u zullen eveneens, als H_0 juist is weinig voorkomen. Hierop berust de keuze van de bij T_1 en T_2 behorende kritieke zônes Z_1 resp. Z_2. Z_1 bevat grote en kleine waarden van n_1 en grote en kleine waarden van u, terwijl Z_2 bij grote waarden van n_1 in hoofdzaak grote waarden van u en bij kleine waarden van n_1 in hoofdzaak kleine waarden van u bevat. T_1 leidt bij voldoende grote n vrijwel steeds tot verwerping als de hypothese niet is vervuld. T_2 leidt echter alleen tot verwerping van H_0 als er veel <u>positieve</u> (resp, negatieve) waarden zijn, die <u>verder</u> van 0 verwijderd liggen dan de <u>negatieve</u> (resp. positieve). In sommige gevallen is het juist van belang om deze laatste afwijkingen van H_0 te ontdekken. In dat geval gebruikt men T_2 liever dan T_1. In fig. 2 is een schematisch voorbeeld gegeven van een serie waarnemingen, waarbij het aantal positieve groter is dan het aantal negatieve, terwijl deze positieve dichter bij 0 liggen dan de negatieve, zodat T_2 niet tot verwerping leidt.

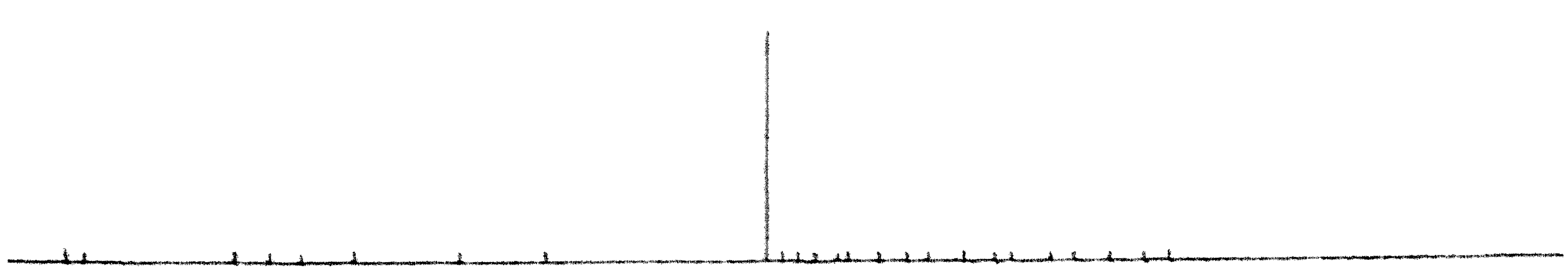

fig. 2

Van T_1 en T_2 bestaan ook éénzijdige versies, waarvan de beschrijving te ver zou voeren.

T_1' en T_2'.

Toetsingsgrootheden.

1e, 2e en 3e als boven. Eerste toetsingsgrootheid: $\underline{n}_1$.
4e: op $x_1,\dots,x_{n_1}$ en $y_1,\dots,y_{n_2}$ wordt de toets van Wilcoxon toegepast (vgl. S 47 (M 8)). De toetsingsgrootheden zijn $\underline{n}_1$ en de $\underline{U}$ van deze toets van Wilcoxon.

Kritieke zônes. Overwegingen analoog aan die voor T_1 en T_2 (met U in plaats van u) leiden tot analoge kritieke zônes Z_1' en Z_2', behorend bij T_1' en T_2'.

Opmerkingen: T_1 en T_2 zijn bijzonder geschikt voor een niet te groot aantal waarnemingen. Zij gelden ook voor niet continue verdelingen. T_1' en T_2' zijn alleen geschikt, als er geen of weinig paren (x_i,y_j) met $x_i=y_j$ zijn. Voor grote aantallen zijn T_1' en T_2' geschikter dan T_1 en T_2. Er is ook een versie voor grote aantallen (T_1'' en T_2''), die geheel analoog is met T_1' en T_2' met dien verstande, dat $\underline{u}$ in plaats van $\underline{U}$ wordt gebruikt (vgl. b.v. [2], blz. 77, § 6.4.5).

Litteratuur:

[1] J.Hemelrijk, A family of parameterfree tests for symmetry with respect to a given point, I, II. Proceedings van de Kon.Ned.Ak.v.Wet. 53 (1950), p.945-955. Indagationes Mathematicae 12 (1950), p. 340-350.

[2] - " - , Symmetrietoetsen, Diss., Den Haag 1950, Excelsior.