

Een toets tegen verloop voor een aantal kansen *)

door Constance van Eeden

Summary

A test against trend for a number of probabilities.

This is an expository paper on a test for the equality of a number of probabilities against the alternative hypothesis of a trend, first described by Constance van Eeden and J. Hemelrijk [2]. In section 2 the test is illustrated by means of an example and methods are given to determine the exact distribution of the test statistic under the null-hypothesis and an approximation for this distribution is described. Section 3 gives the general formulation of the test and section 4 contains a generalization. In section 5 methods are given for combining independent tests of this kind.

1. Inleiding

In dit artikel wordt een toets beschreven waarmee men kan onderzoeken of er in een rij kansen een stijgend of dalend verloop aanwezig is. Dit kan een verloop in de tijd zijn of een verloop in samenhang met een andere grootte, zoals leeftijd, gewicht e.d.

In par. 2 zullen wij deze toets aan de hand van een praktijk-voorbeeld beschrijven. De algemene formulering geven wij in par. 3, terwijl in par. 4 een generalisatie van de toets besproken wordt; in par. 5 zullen wij beschrijven hoe men een aantal onafhankelijke toetsen voor verloop kan combineren. Wij geven in dit artikel geen bewijzen; deze zijn elders gegeven door Constance van Eeden en J. Hemelrijk [2].

2. Beschrijving van de toets aan de hand van een voorbeeld

2.1. Het voorbeeld.

Het voorbeeld, dat wij zullen gebruiken om de toets toe te lichten is ontleend aan een onderzoek, dat uitgevoerd werd door het Adviesbureau voor Toegepaste Statistiek in Den Haag, in verband met maat-normalisatie voor de confectie-industrie. Bij een onderdeel hiervan onderzocht men of het oordeel van leden van een jury een verloop vertoonde in de tijd. Deze jury beoordeelde het al of niet passen van kledingstukken en om het verloop te onderzoeken liet men hen, tussen andere beoordelingen door, enige gelijkwaardige combinaties van mannequin en kledingstuk beoordelen. Een verloop in de gestrengheid der beoordeling zou bij een dergelijke proef b.v. kunnen op-

*) Rapport SP 44 van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam. De afdeling staat onder leiding van Prof. Dr D. van Dantzig.

treden als gevolg van vermoeidheid enerzijds, of van toenemende routine anderzijds.

Nummeren wij de combinaties in de volgorde waarin zij beoordeeld zijn en geven wij de goedkeuringen aan met een 1 en de afkeuringen met een 0 dan krijgen wij voor ieder jurylid een rij nullen en enen, b.v.:

TABEL 1. Resultaten van de beoordeling van het passen van gelijkwaardige combinaties door zes juryleden.

Jurylid nr	Combinatie nr (in tijdsvolgorde)								
	1	2	3	4	5	6	7	8	9
1	1	0	1	1	1	0	0	1	1
2	1	0	1	0	1	0	0	1	0
3	1	0	1	0	1	0	0	1	0
4	1	1	1	1	0	1	1	1	1
5	1	1	1	1	1	0	0	1	1
6	1	0	1	1	1	1	0	1	1

Als er nu een verloop in het oordeel van één der juryleden is, dan zal men dus voor dat jurylid of wel rechts meer goedkeuringen (dus meer enen) vinden dan links of wel rechts meer afkeuringen (dus meer nullen) dan links. Als grootheid om het verloop te onderzoeken kan men dus b.v. gebruiken de som van de aantallen nullen, die achter ieder der enen staan. Deze grootheid neemt grote waarden aan als het aantal afkeuringen stijgt in de tijd en kleine waarden als het aantal goedkeuringen stijgt in de tijd.

Voor jurylid no. 1 krijgen wij:

achter de eerste 1 staan 3 nullen

„ „ tweede 1 „ 2 „

„ „ derde 1 „ 2 „

„ „ vierde 1 „ 2 „

————— +

De grootheid wordt dus 9

Voor de overige juryleden vinden wij voor deze grootheid resp.: 13, 13, 4, 10 en 6.

Om de overschrijdingskans van het gevonden resultaat te kunnen vinden moeten wij de verdeling van de grootheid kennen onder de hypothese, dat het oordeel van het betrokken jurylid constant is in de tijd (dus onder de hypothese, dat voor dat jurylid de kans op goedkeuring constant is in de tijd). Wij zullen, om deze verdeling af te leiden, aantonen, dat de voorgestelde

toets samenhangt met de toets van Wilcoxon voor twee steekproeven en dat men van de tabellen voor deze toets gebruik kan maken.

2.2. Verband met de toets van Wilcoxon.

De toets van Wilcoxon (zie b.v. [5], [7] en [8]) dient om te onderzoeken of twee steekproeven uit eenzelfde populatie afkomstig zijn. Daarbij wordt de volgende toetsingsgroottheid gebruikt. Eerst worden de waarnemingen van de beide steekproeven tezamen naar grootte gerangschikt, waarbij eventuele gelijke waarnemingen boven elkaar worden geplaatst om aan te geven, dat zij op dezelfde plaats in de rangschikking behoren. Vervolgens vervangen wij de waarnemingen van de eerste steekproef alle door een 1 en die van de tweede alle door een 0. Op die wijze verkrijgen wij dus een rij van door elkaar staande enen en nullen, juist als in iedere regel van tabel 1. De toetsingsgroottheid U van de toets van Wilcoxon is nu gelijk aan de som van de aantallen nullen, die achter ieder der enen staan, waarbij op dezelfde plaats staan als half erachter staan geteld wordt, hetgeen dan een bijdrage $\frac{1}{2}$ tot U geeft. Bij ons voorbeeld in tabel 1 komt dit niet voor. Dit zou zich daar kunnen voordoen als twee of meer combinaties op hetzelfde tijdstip door eenzelfde jurylid beoordeeld werden of als de beoordelingen van alle juryleden tezamen werden beschouwd. De in par. 2.1 gedefiniëerde groottheid komt dus juist overeen met de U van Wilcoxon als wij de enen en de nullen uit een rij beoordelingen beschouwen als twee in volgorde van grootte gerangschikte steekproeven A en B , waarbij A bestaat uit de combinatienummers van de goedgekeurde combinaties en B uit die van de afgekeurde combinaties; wij zullen daarom dit symbool U ook voor deze groottheid gebruiken.

Voor jurylid no. 1 b.v. bestaat steekproef A uit de getallen 1, 3, 4, 5, 8 en 9 en B uit de getallen 2, 6 en 7.

Niet alleen de waarde van de groottheid komt echter overeen met de U van Wilcoxon, doch ook zijn waarschijnlijkheidsverdeling onder de getoetste hypothese is dezelfde als bij de toets van Wilcoxon. Dit kan men als volgt inzien. Indien de te toetsen hypothese juist is, d.w.z. als voor het betrokken jurylid de kans op goedkeuring constant (en wel gelijk aan p) is, dan is de kans op het verkrijgen van t_1 goed- en t_2 afkeuringen (waarbij $t_1 + t_2 = n$ het totale aantal beoordelingen is) gelijk aan $p^{t_1} (1-p)^{t_2}$, voor iedere mogelijke volgorde der goed- en afkeuringen. De kans op een bepaalde volgorde, bij gegeven waarden van t_1 en t_2 , is dus ook voor iedere mogelijke volgorde van deze t_1 goed- en t_2 afkeuringen dezelfde. Voor een rij van t_1

enen en t_2 nullen betekent dit dus, dat onder de getoetste hypothese alle mogelijke volgorden gelijke kans bezitten. Hierop berust echter ook juist de afleiding van de verdeling van de toetsingsgrootheid $\underline{U}$ ¹⁾ van Wilcoxon; de verdeling van $\underline{U}$ onder de hypothese van gelijke goedkeuransen is dus dezelfde als bij het toetsen van twee steekproeven, zodat alle voor deze toets berekende tabellen, evenals de normale benadering, gebruikt kunnen worden. Wij geven de formules voor deze benadering zonder de bewijzen.

In de tabellen, formules en nomogrammen worden voor de steekproefgrootten gewoonlijk de symbolen n en m gebruikt. Wij zullen, om verwarring te voorkomen de bovengegeven notatie (dus t_1 en t_2 voor de steekproefgrootten) aanhouden.

De verdeling van de grootheid $\underline{U}$ onder de te toetsen hypothese, die wij met H_0 zullen aangeven, is, voor het geval er geen gelijktijdige beoordelingen²⁾ voorkomen, getabelleerd in [7] voor t_1 en $t_2 \leq 10$ en in [5] voor t_1 en $t_2 \leq 8$; een nomogram van deze verdeling vindt men in [6] voor t_1 en $t_2 \leq 11$. De verdeling is symmetrisch ten opzichte van $\frac{1}{2} t_1 t_2$; de rechtséénzijdige overschrijdingskans wordt gedefiniëerd als de som van de waarschijnlijkheden van die waarden van $\underline{U}$ die niet kleiner zijn dan de gevonden waarde, de linkséénzijdige als de som van de waarschijnlijkheden van die waarden van $\underline{U}$ die niet groter zijn dan de gevonden waarde en de tweezijdige overschrijdingskans is gelijk aan tweemaal de kleinste éénzijdige, d.w.z. men neemt als tweezijdige kritieke zone Z met onbetrouwbaarheidsdrempel α een links- en een rechts-éénzijdige ieder met onbetrouwbaarheidsdrempel $\frac{1}{2} \alpha$.

Voor grotere waarden van t_1 en t_2 kan men gebruik maken van het feit dat $\underline{U}$ onder de te toetsen hypothese en bij de gegeven waarde van t_1 bij benadering normaal verdeeld is met gemiddelde

$$(2.1) \quad \mu = \frac{1}{2} t_1 t_2$$

en, als er geen gelijktijdige beoordelingen zijn, met variantie

$$(2.2) \quad \sigma^2 = \frac{1}{12} t_1 t_2 (n + 1).$$

De rechtséénzijdige overschrijdingskans is bij benadering gelijk aan het oppervlak der normale verdeling met gemiddelde 0 en spreiding 1 rechts van

$$\frac{U - \mu - \frac{1}{2}}{\sigma} \quad ^3)$$

1) Stochastische grootheden zullen onderscheiden worden van getallen (b.v. van de waarden die zij bij een experiment aannemen) door ze te onderstrepen.

2) Of, bij de toets voor twee steekproeven: geen gelijke waarnemingen.

3) Omdat wij hier een discrete verdeling benaderen met een continue passen wij de z.g. continuïteitscorrectie toe. In het algemeen neemt men hiervoor de halve afstand tussen twee opeenvolgende waarden der toetsingsgrootheid; deze afstand is hier gelijk aan 1.

waarin μ en σ gevonden worden uit (2.1) en (2.2), de linkséénzijdige overschrijdingskans is bij benadering gelijk aan het oppervlak links van

$$\frac{U - \mu + \frac{1}{2}}{\sigma}$$

en de tweezijdige overschrijdingskans is bij benadering tweemaal het oppervlak rechts van

$$\frac{|U - \mu| - \frac{1}{2}}{\sigma}.$$

Voor jurylid no. 1 zijn de steekproefgrootten van de steekproeven A en B resp. 6 en 3. Daar bij ons onderzoek zowel een stijging als een daling van de kans op goedkeuring in de tijd mogelijk geacht wordt, toetsen wij tweezijdig. In een tabel of nomogram van de toets van Wilcoxon vinden wij bij de gevonden waarde $U = 9$ een tweezijdige overschrijdingskans van 1. Voor de overige juryleden vinden wij voor deze overschrijdingskansen: 0,55; 0,55; 1; 0,67; 0,89.

Als men alleen wil onderzoeken of de kans op goedkeuring stijgt (dit kan men b.v. doen als men weet dat deze kans zeker niet daalt) dan kiest men een linkséénzijdige kritieke zone Z_L , daar bij stijging van deze kans slechts weinig nullen achter enen zullen staan; als men alleen wil onderzoeken of de kans daalt dan kiest men een rechtséénzijdige kritieke zone Z_R .

2.3. *Het geval van gelijke groepen gelijktijdige beoordelingen.*

Wij hebben hierboven aangegeven hoe men, voor ieder der juryleden apart, kan onderzoeken of de kans op goedkeuring stijgt of daalt in de tijd. Wij zullen nu beschrijven hoe men deze zes resultaten tot één toets kan combineren, als de kans op goedkeuring, voor ieder der combinaties apart, voor alle juryleden dezelfde is, niet alleen onder de getoetste hypothese, maar ook als deze niet juist is. Deze onderstelling is niet erg realistisch, maar terwille van het verschaffen van een numeriek voorbeeld wordt hij hier voor het moment toch ingevoerd¹⁾. Men kan dan de beoordelingen van alle juryleden tezamen nemen en wij krijgen dan in ons voorbeeld het volgende schema:

1) Hoe men de resultaten voor de zes juryleden kan combineren als de kans op goedkeuring niet voor alle juryleden dezelfde is zullen wij in par. 5 aangeven.

TABEL 2. Resultaten van de beoordeling van het passen van $k = 9$ gelijkwaardige combinaties door $m = 6$ juryleden.

Combinaties nr.			1	2	3	4	5	6	7	8	9	Totaal
Aantal	(A) goed-	keuringen	6	2	6	4	5	2	1	6	4	$36 = t_1$
	(B) af-		0	4	0	2	1	4	5	0	2	$18 = t_2$
totaal			6	6	6	6	6	6	6	6	6	$54 = n$

Wij passen nu weer de toets van Wilcoxon toe op analoge wijze als boven; de daar gevoerde bewijsvoering blijft onveranderd geldig.

Stel het aantal beoordeelde combinaties k , het aantal juryleden m en de totale aantallen goed- en afkeuringen resp. t_1 en t_2 . Wij beschouwen nu weer twee steekproeven A en B die tezamen $m \times$ de waarde 1, $m \times$ de waarde 2, enz. tot en met $m \times$ de waarde k bevatten en waarbij A bestaat uit de combinatienummers van de goedgekeurde combinaties (en B dus uit de combinatie-nummers van de afgekeurde combinaties).

In ons voorbeeld is $k = 9$, $m = 6$, $t_1 = 36$ en $t_2 = 18$. Steekproef A bestaat uit (zie tabel 2) 6 maal de waarde 1, 2 maal de waarde 2, 6 maal de waarde 3, enz., en steekproef B uit 0 maal de waarde 1, 4 maal de waarde 2, 0 maal de waarde 3, enz.

De waarde van de toetsingsgrootheid U vinden wij als volgt (zie de bovengegeven definitie): Er zijn in steekproef B 18 waarnemingen die groter zijn dan 1 en in A komt de waarde 1 zes maal voor. De bijdrage van deze zes enen tot U is dus $6 \times 18 = 108$. De bijdrage van ieder der twee waarden 2 tot U is $\frac{1}{2} \times 4 + 2 + 1 + 4 + 5 + 2 = 16$, dus van deze twee waarden tezamen is de bijdrage $2 \times 16 = 32$. Zo voortgaande vinden wij: $U = 372$.

De verdeling van de grootheid U onder de te toetsen hypothese is voor dit geval, waarbij gelijktijdige beoordelingen voorkomen, niet getabelleerd. Men zal dus of wel gebruik moeten maken van een benadering of wel de exacte verdeling zelf moeten uitrekenen.

Bij de benadering maakt men gebruik van het feit dat de grootheid U onder de hypothese H_0 en bij de gegeven waarde van t_1 bij benadering normaal verdeeld is met gemiddelde

$$(2.3) \quad \mu = \frac{1}{2} t_1 t_2$$

en variantie

$$\sigma^2 = \frac{t_1 t_2 m^2 (k^2 - 1)}{12 (n - 1)},$$

bij $n = km$ het totaal aantal beoordelingen voorstelt.
 Als continuïteitscorrectie neemt men weer de halve afstand tussen twee
 opeenvolgende waarden van U ; deze afstand is hier gelijk aan m .

In het voorbeeld is $t_1 = 36$ en $t_2 = 18$ dus

$$\mu = \frac{1}{2} \cdot 36 \cdot 18 = 324$$

$$\sigma^2 = \frac{36 \cdot 18 \cdot 6^2 \cdot (9^2 - 1)}{12 \cdot 53} = 2934$$

Verder is $U = 372$ en $m = 6$. De tweezijdige overschrijdingskans is dus
 bij benadering tweemaal het oppervlak van de normale verdeling met
 gemiddelde 0 en spreiding 1 rechts van

$$\frac{372 - 324 - 3}{\sqrt{2934}} = 0,83.$$

In een tabel der normale verdeling vinden wij voor deze overschrij-
 dingskans 0,41.

Voor kleine waarden van m en k kan men de exacte verdeling van U onder
 de te toetsen hypothese als volgt berekenen:

Stel de aantallen goedkeuringen der combinaties $1, 2, \dots, k$ resp. $a_1,$
 $a_2, \dots, a_k$, dan schrijft men bij de gegeven waarden van m, k en t_1 alle moge-
 lijke stellen waarden $a_1, a_2, \dots, a_k$ op en berekent de bijbehorende waarden
 van U . De kans op zo'n stel waarden $a_1, a_2, \dots, a_k$ onder de hypothese H_0
 en bij de gegeven waarde van t_1 is

$$\binom{m}{a_1} \binom{m}{a_2} \dots \binom{m}{a_k} / \binom{n}{t_1}.$$

(2.5)

Voor $k = 4, m = 2$ en $t_1 = 2$ vinden wij de volgende mogelijke
 waarden a_1, a_2, a_3, a_4 :

$\binom{n}{a} = \frac{n!}{a!(n-a)!}$; tabellen van deze binomiaalcoëfficiënten
 formule geven wij niet.

1) $\begin{array}{cccc c} 0 & 0 & 0 & 2 & 2 \\ 2 & 2 & 2 & 0 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$	2) $\begin{array}{cccc c} 0 & 0 & 1 & 1 & 2 \\ 2 & 2 & 1 & 1 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$	3) $\begin{array}{cccc c} 0 & 0 & 2 & 0 & 2 \\ 2 & 2 & 0 & 2 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$
4) $\begin{array}{cccc c} 0 & 1 & 0 & 1 & 2 \\ 2 & 1 & 2 & 1 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$	5) $\begin{array}{cccc c} 0 & 1 & 1 & 0 & 2 \\ 2 & 1 & 1 & 2 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$	6) $\begin{array}{cccc c} 1 & 0 & 0 & 1 & 2 \\ 1 & 2 & 2 & 1 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$
7) $\begin{array}{cccc c} 1 & 0 & 1 & 0 & 2 \\ 1 & 2 & 1 & 2 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$	8) $\begin{array}{cccc c} 0 & 2 & 0 & 0 & 2 \\ 2 & 0 & 2 & 2 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$	9) $\begin{array}{cccc c} 1 & 1 & 0 & 0 & 2 \\ 1 & 1 & 2 & 2 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$
10) $\begin{array}{cccc c} 2 & 0 & 0 & 0 & 2 \\ 0 & 2 & 2 & 2 & 6 \\ \hline 2 & 2 & 2 & 2 & 8 \end{array}$		

In tabel 3 staan voor ieder van deze tien mogelijke uitkomsten de waarde van U en de kans op die uitkomst, berekend volgens (2.5), vermeld.

TABEL 3. *Exacte verdeling van $\underline{U}$ onder de hypothese H_0 voor $k = 4$, $m = 2$ en $t_1 = 2$.*

mogelijk- heid nr	U	kans
1	0	$\frac{1}{28}$
2	2	$\frac{4}{28}$
3	4	$\frac{1}{28}$
4	4	$\frac{4}{28}$
5	6	$\frac{4}{28}$
6	6	$\frac{4}{28}$
7	8	$\frac{4}{28}$
8	8	$\frac{1}{28}$
9	10	$\frac{4}{28}$
10	12	$\frac{1}{28}$

De exacte verdeling van $\underline{U}$ onder de nulhypothese heeft dus in dit geval de volgende vorm (zie fig. 1).

Men kan nu bewijzen dat de verdeling van $\underline{U}$ onder de hypothese H_0 in dit geval (dus het geval van gelijke groepen van gelijktijdige beoordelingen) altijd symmetrisch is ten opzichte van $\frac{1}{2} t_1 t_2$. De exacte linker-, rechter- en

tweezijdige overschrijdingskansen worden net zo gedefinieerd als in par. 2.2 is aangegeven voor het geval $m = 1$, d.w.z. voor het geval dat er geen gelijktijdige beoordelingen zijn.

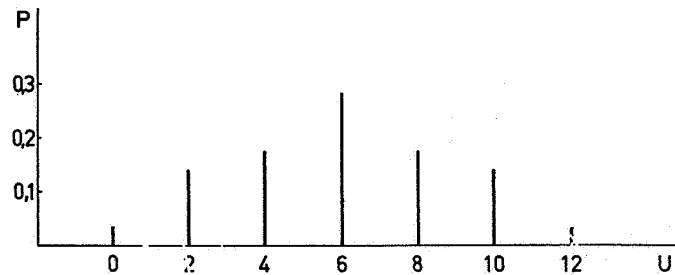


Fig. 1. Exacte verdeling van U onder de hypothese H_0 voor $k = 4$, $m = 2$ en $t_1 = 2$.

2.4. Het geval van ongelijke groepen van gelijktijdige beoordelingen.

Wij beschouwen nu het geval dat de aantallen beoordelingen voor de verschillende combinaties niet dezelfde zijn. Dit kan bij ons voorbeeld optreden als sommige combinaties door bepaalde juryleden niet beoordeeld worden. De beoordelingen kunnen dan als volgt worden samengevat:

TABEL 4. Beoordeling van het passen van k gelijkwaardige combinaties door een aantal juryleden.

Combinatie nr. (in tijdsvolgorde)		k	Totaal
Aantal	(A) goed	a_k	t_1
	(B) vindt men b.v. in [3].	b_k	t_2
		n_k	n

Wij zou
twee ste
2, en
co

de wa

Deze toets ne dat de combinatienummers met grote
waarden van n_i zwaarde dan degenen met kleine n_i . Neem b.v.
 $k = 3$ en stel de kans op goedkeu g resp. p_1 , p_2 en p_3 met $p_1 = p_3$ en $p_2 > p_1$
(zie fig. 2).

so x o n kunnen toepassen op
al de waarde 1, n_2 maal de waarde
atten en waarbij A bestaat uit de
combinaties (dus uit a_1 maal de
en met a_k maal de waarde k) en B
e waarde 2, enz. tot en met b_k maal

Dit is dus een geval waarin men noch van een stijgend, noch van een dalend verloop kan spreken, daar eerst een stijging en daarna een even grote daling optreedt. Nemen wij nu echter $n_1 < n_3$, dan telt de stijging $p_2 - p_1$ minder

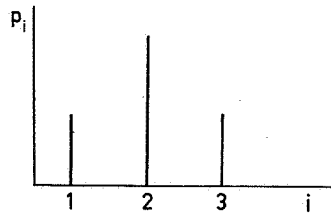


Fig. 2. Voorbeeld van een geval waarbij de kansen op goedkeuring wel ongelijk zijn, maar noch een overwegend stijgend, noch een overwegend dalend verloop vertonen.

mee dan de daling $p_2 - p_3$ en met behulp van de toets zal men dan, zoals in [2] bewezen is, voor voldoende grote n tot de conclusie komen dat de kans op goedkeuring daalt. Neemt men $n_1 > n_3$, dan zal men, voor voldoende grote n , tot de conclusie komen, dat de kans stijgt. Dit betekent dus in feite dat men in dergelijke gevallen ieder gewenst resultaat zou kunnen bereiken door met de verhoudingen der aantallen beoordelingen te manipuleren. Dit is bij gelijke aantallen uiteraard niet mogelijk.

Wij willen dit vermijden en zullen daartoe een toets beschrijven die in gevallen zoals dat van fig. 2, niet de mogelijkheid biedt tot het verkrijgen van tegengestelde uitkomsten door manipulatie met de aantallen en die, voor het geval dat $n_i = m$ voor iedere i , identiek is met de in par. 2.3 beschreven toets. De toets zal, voor voldoende grote n , wel tot verwerping leiden als de kans op goedkeuring overwegend stijgt of daalt, dus niet alleen als iedere kans groter (resp. kleiner) is dan de voorafgaande.

Algemeen geformuleerd zeggen wij dat de kansen $p_1, p_2, \dots, p_k$ een stijgend resp. een dalend verloop vertonen als voldaan is aan

$$(2.6) \quad \theta \stackrel{\text{def}}{=} \sum_{i < j} (p_i - p_j) < 0 \text{ resp. } > 0.$$

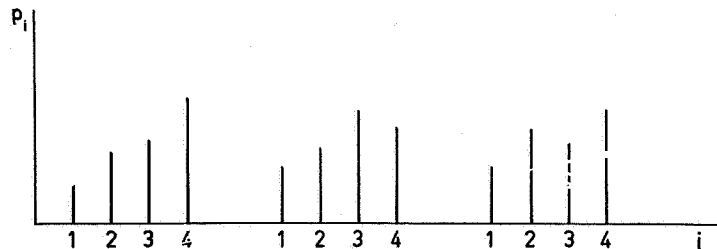


Fig. 3. Enige gevallen waarbij de kansen p_1, p_2, p_3, p_4 een stijgend verloop vertonen, d.w.z. waarbij $\theta < 0$ is.

In het voorbeeld van fig. 2 is deze grootheid $\theta = 0$ en in fig. 3 zijn een aantal voorbeelden gegeven waarbij $\theta < 0$ is.

Wij willen nu dus een toets hebben die in het tweezijdige geval tot verworping der nulhypothese leidt als $\theta \neq 0$ is. Als toetsingsgrootheid voor deze toets gebruiken wij een zuivere schatting van θ en verwerpen de nulhypothese ten gunste van een dalend verloop als deze grootheid grote positieve waarden aanneemt en ten gunste van een stijgend verloop als de grootheid grote negatieve waarden aanneemt.

Daar $\frac{a_i}{n_i}$ een zuivere schatting van p_i is ($i = 1, 2, \dots, k$), is

$$(2.7) \quad \underline{W} = \sum_{i < j} \sum \left(\frac{a_i}{n_i} - \frac{a_j}{n_j} \right) = \sum_{i < j} \sum \frac{a_i n_j - a_j n_i}{n_i n_j} = \sum_i (k + 1 - 2i) \frac{a_i}{n_i}$$

een zuivere schatting van θ .

Het gemiddelde van $\underline{W}$ onder de hypothese H_0 en bij de gegeven waarde van t_1 is

$$(2.8) \quad \mu = 0$$

en de variantie is

$$(2.9) \quad \sigma^2 = \frac{t_1 t_2}{n(n-1)} \sum_i \frac{(k+1-2i)^2}{n_i}.$$

Als $n_i = m$ voor iedere i dan is de toets gebaseerd op $\underline{W}$ identiek met de toets gebaseerd op de grootheid $\underline{U}$ van Wilcoxon berekend uit de twee steekproeven A en B . Dan geldt nl., zoals gemakkelijk na te rekenen is

$$(2.10) \quad \underline{W} = \frac{2 \underline{U} - t_1 t_2}{m^2}.$$

Hierin zijn alleen $\underline{W}$ en $\underline{U}$ stochastisch, zodat $\underline{W}$ een lineaire functie van $\underline{U}$ is. De toets verandert dus niet als wij van $\underline{W}$ op $\underline{U}$ overgaan of omgekeerd.

Voor kleine waarden van k en $n_1, n_2, \dots, n_k$ kan men de exacte verdeling van $\underline{W}$ berekenen op analoge wijze als in par. 2.3 aangegeven is voor $\underline{U}$. Wij schrijven dus weer alle mogelijke stellen waarden $a_1, a_2, \dots, a_k$ op en berekenen voor ieder geval de bijbehorende waarde van $\underline{W}$. De kans op zo'n stel waarden $a_1, a_2, \dots, a_k$ wordt hier

$$(2.11) \quad \binom{n_1}{a_1} \binom{n_2}{a_2} \dots \binom{n_k}{a_k} / \binom{n}{t_1}.$$

In fig. 4, 5 en 6 zijn voor een aantal gevallen de exacte verdelingen van W onder de hypothese H_0 getekend.

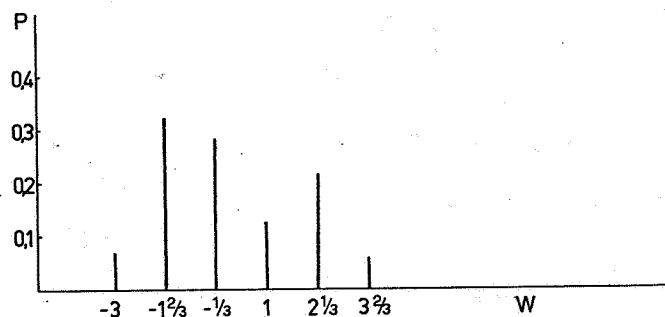


Fig. 4. De exacte verdeling van W onder de nulhypothese voor $k = 4$, $n_1 = n_3 = 1$, $n_2 = n_4 = 3$, $t_1 = 3$ en $t_2 = 5$.

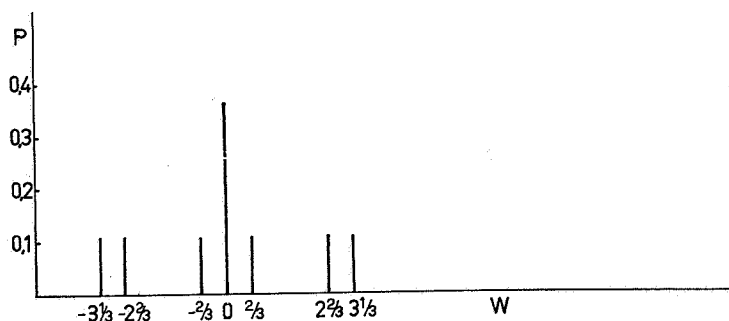


Fig. 5. Exacte verdeling van W onder de nulhypothese voor $k = 4$, $n_1 = n_4 = 1$, $n_2 = n_3 = 3$, $t_1 = 2$ en $t_2 = 6$.

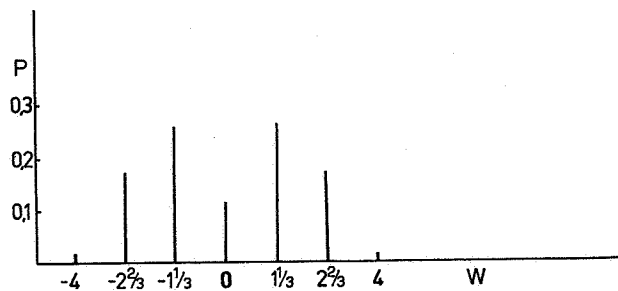


Fig. 6. Exacte verdeling van W onder de nulhypothese voor $k = 4$, $n_1 = n_3 = 1$, $n_2 = n_4 = 3$ en $t_1 = t_2 = 4$.

Uit deze voorbeelden zien wij dat de verdeling van $\underline{W}$ in het algemeen niet symmetrisch is (zie fig. 4) en dat de afstand tussen de opeenvolgende waarden van $\underline{W}$ niet altijd constant is (zie fig. 5). Verder blijkt uit de fig. 4 en 6 dat de verdeling meer dan één top kan bezitten.

De exacte eenzijdige overschrijdingskansen worden weer gedefiniëerd als aangegeven is in par. 2.2. Voor tweezijdige toetsing beschouwen wij eerst het geval, dat de verdeling van $\underline{W}$ onder H_0 symmetrisch is. Dit doet zich voor

- a) als $t_1 = t_2$ (zie fig. 6),
- b) als $n_1 = n_k$, $n_2 = n_{k-1}$ enz. (zie fig. 5); hieraan is o.a. voldaan als $n_i = m$ voor iedere i .

In deze symmetrische gevallen neemt men als tweezijdige kritieke zone Z met onbetrouwbaarheidsdrempel α weer een links- en een rechtséénzijdige, ieder met onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha$. De tweezijdige overschrijdingskans is dan dus tweemaal de kleinste éénzijdige overschrijdingskansen.

Deze keuze van de kritieke zone heeft bij asymmetrische verdelingen echter het bezwaar, dat er een groot verschil ontstaan kan tussen α en de werkelijke onbetrouwbaarheid. Als de grootste of de kleinste waarde, die $\underline{W}$ aan kan nemen, een waarschijnlijkheid $> \frac{1}{2}\alpha$ heeft, wordt de zo gedefiniëerde tweezijdige kritieke zone bovendien éénzijdig. Het verschil tussen α en de werkelijke onbetrouwbaarheid is dan $\geq \frac{1}{2}\alpha$.

Verschillende andere methoden zouden nu in aanmerking kunnen komen.

Als men b.v. bij asymmetrische verdelingen de tweezijdige kritieke zone vormt door het bijeenzoeken van de waarden van $\underline{W}$ met de kleinste waarschijnlijkheden (zie [4]), totdat de gekozen onbetrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert, dan zal de werkelijke onbetrouwbaarheid in het algemeen dichter bij α liggen dan bij de bovengedefiniëerde kritieke zone Z , terwijl de kritieke zone dan een maximaal aantal waarden van $\underline{W}$ bevat. Wij zullen de zo gevormde tweezijdige kritieke zone Z_1 noemen. De tweezijdige overschrijdingskans wordt hier de som van de waarschijnlijkheden die niet groter zijn dan de waarschijnlijkheid van de gevonden waarde van $\underline{W}$.

In fig. 7 is de verdeling van $\underline{W}$ getekend voor $k = 2$, $n_1 = 6$, $n_2 = 4$, $t_1 = 3$ en $t_2 = 7$.

De tweezijdige kritieke zone Z met onbetrouwbaarheidsdrempel 0,25¹⁾

1) Wij gebruiken in dit en volgende voorbeelden deze grote onbetrouwbaarheidsdrempel voor illustratieve doeleinden.

bestaat hier uit de waarde $W = -\frac{3}{4}$. De werkelijke onbetrouwbaarheid is $\frac{1}{30} = 0,033$. De tweezijdige kritieke zone Z_1 met dezelfde onbetrouwbaarheidsdrempel bestaat uit de waarden $W = -\frac{3}{4}$ en $W = \frac{1}{2}$. De onbetrouwbaarheid is hier $\frac{1}{30} + \frac{1}{6} = 0,20$.

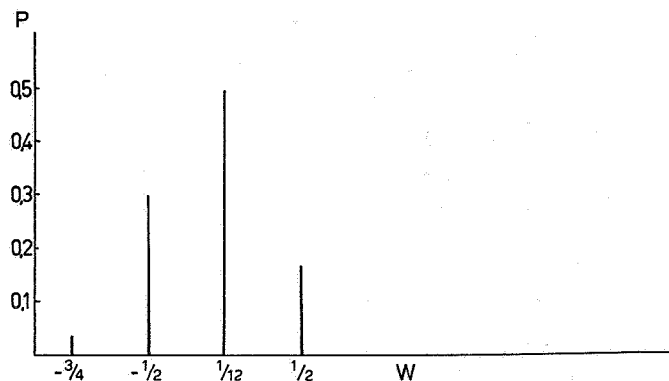


Fig. 7. Exacte verdeling van $\underline{W}$ onder de nulhypothese voor $k = 2$, $n_1 = 6$, $n_2 = 4$, $t_1 = 3$ en $t_2 = 7$.

De tweezijdige kritieke zone Z_1 heeft echter bij verdelingen met meer dan één top het bezwaar dat het linker- en rechterdeel der kritieke zone niet altijd ieder een aaneengesloten geheel vormen.

Bij het voorbeeld van fig. 4 bestaat de kritieke zone Z_1 met onbetrouwbaarheidsdrempel 0,25 uit de waarden $W = -3$, $W = 1$ en $W = 3\frac{2}{3}$.

Om nu te komen tot een algeme methode voor het stapsgewijze opbouwen van een tweezijdige kritieke zone, die de genoemde nadelen niet bezit en die voor symmetrische gevallen met de gebruikelijke overeenstemt, stellen wij eerst een aantal principes vast, die wij bij deze opbouw in de aangegeven volgorde willen toepassen. Dit zijn:

- het linker- en rechterdeel der kritieke zone vormen bij iedere stap ieder een aaneengesloten geheel,
- alle waarden, die tot een kritieke zone behoren, behoren ook tot alle kritieke zones met grotere onbetrouwbaarheidsdrempels,
- bij iedere stap wordt het verschil tussen de onbetrouwbaarheden van linker- en rechterdeel van de kritieke zone zo klein mogelijk gehouden,

- d) de onbetrouwbaarheid van de kritieke zone wordt, voor iedere onbetrouwbaarheidsdrempel α zo dicht mogelijk bij α gebracht,
- e) indien de principes a, ..., d bij een bepaalde stap niet één, doch twee mogelijke nieuwe waarden aanwijzen als volgende bouwsteen voor de kritieke zone, worden deze tegelijk aan de kritieke zone toegevoegd.

Deze principes sluiten inderdaad de twee bovenbeschreven methoden (Z en Z_1) als algemene methode uit; zij leiden bovendien in het symmetrische geval tot $Z = Z_1$, terwijl verder een ondubbelzinnige opbouw verkregen wordt, die, op aanschouwelijke wijze, als volgt beschreven kan worden. Het bewijs van de aequivalentie van deze opbouw en de bovengenoemde principes geven wij summier aan door de gebruikte principes tussen haakjes te vermelden.

Wij beginnen met van de grootste en de kleinste waarde, die $\underline{W}$ aan kan nemen, diegene te kiezen met de kleinste waarschijnlijkheid (c); zijn de waarschijnlijkheden van deze twee waarden aan elkaar gelijk, dan nemen wij beide waarden (e). De volgende waarde van $\underline{W}$, die wij bij de kritieke zone nemen, is een waarde links of een waarde rechts zo, dat het linker- en rechterdeel ieder een aaneengesloten geheel vormen (a) en dat het verschil tussen de onbetrouwbaarheden links en rechts in absolute waarde zo klein mogelijk is (c). Zo gaan wij door iedere keer links of rechts een waarde van $\underline{W}$ bij de kritieke zone te nemen, totdat de gekozen onbetrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert.

Als het toevoegen van een waarde links in absolute waarde hetzelfde verschil tussen links en rechts geeft als het toevoegen van een waarde rechts, dan wordt van deze twee waarden degene met de kleinste waarschijnlijkheid toegevoegd (b en d). Tenslotte: is het verschil tussen links en rechts in een bepaald stadium gelijk aan 0 en hebben de volgende linker- en rechterwaarde van $\underline{W}$ dezelfde waarschijnlijkheid dan worden beide waarden toegevoegd (of, als men daarmee α overschrijdt, geen van beide; e).

Wij zullen de zo gedefiniëerde kritieke zone aanduiden als Z_2 . De tweezijdige overschrijdingskans wordt gedefiniëerd als de onbetrouwbaarheid van de kleinste kritieke zone Z_2 , die het gevonden resultaat bevat.

In tabel 5 en fig. 8 is de verdeling van $\underline{W}$ onder de hypothese H_0 gegeven voor $k = 4$, $n_1 = n_2 = 1$, $n_3 = 2$, $n_4 = 11$ en $t_1 = 4$.

De bovengedefiniëerde kritieke zone Z_2 met onbetrouwbaarheidsdrempel 0,25 vinden wij nu als volgt:

TABEL 5. *Exacte verdeling van $\underline{W}$ onder de hypothese H_0 voor $k = 4$, $n_1 = n_2 = 1$, $n_3 = 2$, $n_4 = 11$ en $t_1 = 4$.*

W	$P[W]$
$-1^{6/11}$	0,040
$-1^{7/22}$	0,242
$-1^{1/11}$	0,242
$-3^{3/11}$	0,008
$-1^{1/22}$	0,080
$2^{2/11}$	0,121
$1^{6/11}$	0,008
$1^{21/22}$	0,080
$2^{2/11}$	0,121
3	0,001
$3^{5/22}$	0,016
$3^{5/11}$	0,040

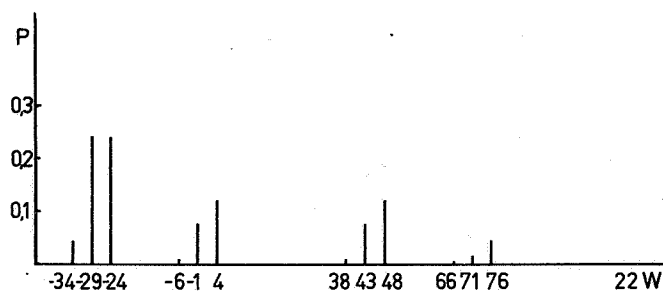


Fig. 8. *Exacte verdeling van $\underline{W}$ onder de hypothese H_0 voor $k = 4$, $n_1 = n_2 = 1$, $n_3 = 2$, $n_4 = 11$, $t_1 = 4$ en $t_2 = 11$. 1)*

De grootste en de kleinste waarde van $\underline{W}$ hebben hier dezelfde waarschijnlijkheid dus wij nemen ze beide bij de kritieke zone²⁾. De onbetrouwbaarheden zijn dan links en rechts 0,040. Vervolgens kiezen wij één der waarden $3^{5/22}$ en $-1^{7/22}$; nemen wij $3^{5/22}$ dan is het verschil tussen de onbetrouwbaarheden links en rechts in absolute waarde $0,040 + 0,016 = 0,040 + 0,016 = 0,056$; nemen wij $-1^{7/22}$ dan wordt dit verschil

1) Het merkwaardige karakter van deze verdeling ontstaat doordat n_4 zoveel van de overige aantallen verschilt. Terwille van de duidelijkheid van de figuur is op de horizontale as niet W zelf, maar $22 W$ uitgezet.

2) Hadden wij als onbetrouwbaarheidsdrempel 0,05 gekozen, dan was het dus niet mogelijk geweest beide waarden bij de kritieke zone te nemen. In dit geval bevat de kritieke zone Z_2 dus geen enkele waarde van W .

$0,040 + 0,242 - 0,040 = 0,242$. Wij kiezen dus $3^5/22$. De onbetrouwbaarheden links en rechts zijn dan resp. $0,040$ en $0,056$. Zo doorgaande vinden wij voor de tweezijdige kritieke zone Z_2 met onbetrouwbaarheidsdrempel $0,25$ de waarden $W = -1^6/11, 3^5/11, 3^5/22, 3$ en $2^2/11$.

In tabel 6 staan voor ieder der mogelijke waarden van $\underline{W}$ de tweezijdige overschrijdingskans (dus de onbetrouwbaarheid van de kleinste kritieke zone die die waarde van W bevat) vermeld.

TABEL 6. Tweezijdige overschrijdingskansen voor de mogelijke waarden van $\underline{W}$ bij $k = 4$, $n_1 = n_2 = 1, n_3 = 2, n_4 = 11$ en $t_1 = 4$.

$W^1)$	Tweezijdige over- schrijdingskansen
$-1^6/11$ (1)	0,080
$-1^7/22$ (5)	0,460
$-1^1/11$ (9)	0,911
$-3/11$ (11)	1
$-1/22$ (10)	0,991
$2/11$ (8)	0,669
$1^8/11$ (7)	0,548
$1^{21}/22$ (6)	0,540
$2^2/11$ (4)	0,218
3 (3)	0,097
$3^5/22$ (2)	0,096
$3^5/11$ (1)	0,080

Wij hebben hierboven dus aangegeven hoe men de exacte overschrijdingskansen kan vinden.

Voor grote waarden van k en (of) $n_1, n_2, \dots, n_k$ kan men gebruik maken van het feit dat $\underline{W}$ onder de hypothese H_0 bij benadering normaal verdeeld is met gemiddelde en variantie volgens (2.8) en (2.9). Wij hebben reeds eerder opgemerkt dat men bij zo'n benadering als continuïteitscorrectie de halve afstand tussen de opeenvolgende waarden van de toetsingsgrootheid neemt. Nu is, als $n_1, n_2, \dots, n_k$ niet alle aan elkaar gelijk zijn, deze afstand niet altijd constant. Het is echter mogelijk om, zonder alle mogelijke waarden van $\underline{W}$ uit te rekenen, aan te geven wat de kleinste afstand tussen twee opeenvolgende waarden van W is en wij nemen nu als continuïteitscorrectie de helft van deze kleinste afstand.

Deze kleinste afstand kunnen wij als volgt berekenen:

1) De cijfers tussen haakjes geven de volgorde aan waarin de waarden van W bij de kritieke zone genomen zijn.

TABEL 7. Exacte en benaderde links- en rechtséénzijdige overschrijdingskansen voor $k = 4$, $n = 8$ en enige waarden van n_1 , n_2 , n_3 en n_4 .

t_1	1			2			3			4		
	W	exact	benaderd	W	exact	benaderd	W	exact	benaderd	W	exact	benaderd
$n_i = 2$ ($i = 1, 2, 3, 4$)	$-1\frac{1}{2}$ $-1\frac{1}{2}$ $1\frac{1}{2}$ $1\frac{1}{2}$	0,25 0,25	0,1867 0,1867	-3 -2 -1 0 1 2 3	0,0357 0,1786 0,1786 0,0357	0,0436 0,1539 0,1539 0,0436	$-3\frac{1}{2}$ $-2\frac{1}{2}$ $-1\frac{1}{2}$ $-1\frac{1}{2}$ $1\frac{1}{2}$ $1\frac{1}{2}$ $3\frac{1}{2}$	0,0357 0,1071 0,1071 0,0357	0,0336 0,1112 0,1112 0,0336	-4 -3 -2 -1 0 1 2 3 4	0,0143 0,0714 0,2 0,2 0,0714 0,0143	0,0192 0,0694 0,1867 0,1867 0,0694 0,0192
$n_1 = n_3 = 1$ $n_2 = n_4 = 3$	-1 $1\frac{1}{3}$ 3	0,5 0,125	0,6026 0,0351	-2 $-2\frac{2}{3}$ $2\frac{2}{3}$ 2 $3\frac{1}{3}$	0,2143 0,1071	0,2148 0,0571	-3 $-1\frac{2}{3}$ $-1\frac{1}{3}$ 1 $2\frac{1}{3}$ $3\frac{2}{3}$	0,0714 0,3929 0,2679 0,0536	0,1075 0,2981 0,1849 0,0559	-4 $-2\frac{2}{3}$ $-1\frac{1}{3}$ 0 $1\frac{1}{3}$ $2\frac{2}{3}$ 4	0,0143 0,1857 0,1857 0,0143	0,0436 0,1539 0,1539 0,0436
$n_1 = n_4 = 1$ $n_2 = n_3 = 3$	-3 $-1\frac{1}{3}$ $1\frac{1}{3}$ 3	0,125 0,125	0,0401 0,0401	$-3\frac{1}{3}$ $-2\frac{2}{3}$ $-2\frac{2}{3}$ 0 $2\frac{2}{3}$ $2\frac{2}{3}$ $3\frac{1}{3}$	0,1071 0,1071	0,0668 0,0668	$-3\frac{2}{3}$ -3 $-2\frac{1}{3}$ -1 $-1\frac{1}{3}$ $1\frac{1}{3}$ 1 $2\frac{1}{3}$	0,0536 0,2143 0,2143	0,0681 0,1170 0,1170	-4 $-3\frac{1}{3}$ $-2\frac{2}{3}$ -2 $-2\frac{2}{3}$ 0 $2\frac{2}{3}$ 2	0,0143 0,1429 0,1429	0,0559 0,0968 0,0968

	"	"				$\frac{3}{2}$ $3\frac{2}{3}$	$\frac{1}{2}$ $0,536$	$\frac{1}{2}$ $0,681$	$\frac{1}{2}$ $3\frac{1}{3}$ 4	$\frac{1}{2}$ $0,1429$ $0,0143$	$\frac{1}{2}$ $0,0968$ $0,0559$
$n_1=n_2=1$ $n_3=2$ $n_4=4$	$-\frac{3}{4}$ $-\frac{1}{2}$ 1 3	0,5 0,125	0,3085 0,0113	$-\frac{1}{2}$ $-\frac{1}{4}$ -1 $\frac{1}{4}$ $\frac{1}{2}$ $\frac{2}{4}$ $\frac{2}{2}$ 4	0,2143 0,1071 0,0357	0,2033 0,0749 0,0096	0,0714 0,2857 0,1071 0,0357	0,1251 0,1562 0,0455 0,0336	-3 $-\frac{3}{4}$ $-\frac{2}{2}$ $-\frac{1}{4}$ -1 - $\frac{3}{4}$ $\frac{3}{4}$ 1 $\frac{1}{4}$ $\frac{2}{2}$ $\frac{2}{4}$ 3	0,0143 0,1286 0,1286 0,0143	0,0655 0,0838
$n_1=1$ $n_2=n_3=2$ $n_4=3$	-1 $-\frac{1}{2}$ $\frac{1}{2}$ 3	0,375 0,125	0,2843 0,0188	-2 $-\frac{1}{2}$ -1 $-\frac{1}{2}$ 0 1 2 $\frac{2}{2}$ $\frac{3}{2}$	0,1071 0,1429 0,0714	0,1562 0,0985 0,0301	0,0179 0,125 0,1964 0,0893 0,0179	0,0778 0,1230 0,1230 0,0778 0,0262	$-\frac{3}{2}$ -3 $-\frac{2}{2}$ -2 $-\frac{1}{2}$ -1 - $\frac{1}{2}$ 0 $\frac{1}{2}$ 1 $\frac{1}{2}$ 2 $\frac{2}{2}$ 3 $\frac{3}{2}$	0,0286 0,0714 0,1 0,1 0,0714 0,0286	0,0516 0,0838 0,1292 0,1292 0,0838 0,0516

De bijdrage h_i die één goedkeuring van de i^e combinatie tot W geeft is $(k + 1 - 2i)/n_i$. Verplaatsen wij dus een goedkeuring van de i^e naar de j^e combinatie dan verandert W in absolute waarde met $|h_i - h_j|$ en de kleinste afstand tussen twee opeenvolgende waarden van W is dus, daar wij het totale aantal goedkeuringen constant, nl. gelijk aan t_1 , houden, de kleinste positieve waarde van $|h_i - h_j|$.

In het voorbeeld van fig. 5 was $k = 4$, $n_1 = n_4 = 1$, $n_2 = n_3 = 3$.

Voor h_i vinden wij dus

i	h_i
1	3
2	$1/3$
3	$-1/3$
4	-3

De kleinste positieve waarde van $|h_i - h_j|$ is dus $2/3$ en de continuïteitscorrectie is dus $1/3$.

In tabel 7 zijn voor $k = 4$, $n = 8$ en een aantal waarden van n_1 , n_2 , n_3 en n_4 de exacte verdeling van $\underline{W}$ vergeleken met deze normale benadering. In de tabel staan de rechts- en de linkséénzijdige overschrijdingskansen vermeld tot en met de eerste overschrijdingskans die (exact berekend) groter dan of gelijk aan 0,1 is. De benadering met de normale verdeling zal in het algemeen beter worden naarmate n groter wordt en $|t_1 - t_2|$ en $\max_{i,j} |n_i - n_j|$ kleiner worden. Het verdient dus aanbeveling er bij de proefopzet, indien mogelijk, voor te zorgen dat de aantallen n_i onderling niet te veel verschillen en dus gevallen zoals die van fig. 8 zoveel mogelijk te vermijden.

3. Algemene formulering

De in par. 2 beschreven toets kan als volgt algemeen geformuleerd worden:

Gegeven k ($k \geq 2$) onafhankelijke reeksen van onafhankelijke experimenten, waarbij ieder experiment één der mogelijke uitkomsten succes of mislukking heeft. De i^e reeks bestaat uit n_i experimenten met $\underline{a}_i$ successen en $\underline{b}_i$ mislukkingen; p_i is de kans op succes voor ieder experiment van de i^e reeks ($i = 1, 2, \dots, k$) en

$$(3.1) \quad \underline{t}_1 \stackrel{\text{def}}{=} \sum_i \underline{a}_i, \quad \underline{t}_2 \stackrel{\text{def}}{=} \sum_i \underline{b}_i, \quad n \stackrel{\text{def}}{=} \sum_i n_i.$$

Deze waarnemingen kunnen wij als volgt samenvatten:

TABEL 8. Waarnemings-schema.

Reeks nr		1	2	.	.	.	k	totaal
kans op succes		p_1	p_2	.	.	.	p_k	
aan- tal	successen	$\underline{a}_1$	$\underline{a}_2$	.	.	.	$\underline{a}_k$	t_1
	mislukkingen	$\underline{b}_1$	$\underline{b}_2$	.	.	.	$\underline{b}_k$	t_2
	experimenten	n_1	n_2	.	.	.	n_k	n

De hypothese, die wij willen toetsen luidt

$$(3.2) \quad H_0 : p_1 = p_2 = \dots = p_k.$$

Hiervoor gebruiken wij de toetsingsgrootheid

$$(3.3) \quad \underline{W} = \sum_{i < j} \sum \left(\frac{\underline{a}_i}{n_i} - \frac{\underline{a}_j}{n_j} \right) = \sum_i (k + 1 - 2i) \frac{\underline{a}_i}{n_i},$$

waarvoor geldt

$$(3.4) \quad E(\underline{W} \mid t_1, H_0) = 0,$$

$$(3.5) \quad \sigma^2(\underline{W} \mid t_1, H_0) = \frac{t_1 t_2}{n(n-1)} \sum_i \frac{(k+1-2i)^2}{n_i}.$$

De grootheid $\underline{W}$ is onder de hypothese H_0 en onder de voorwaarde $t_1 = t_1$ bij benadering normaal verdeeld met gemiddelde en variantie volgens (3.4) en (3.5).

De tweezijdige toets is, onder zekere niet zeer strenge voorwaarden, waarvoor wij verwijzen naar [2], asymptotisch onderscheidend ten opzichte van de klasse van alternatieve hypothesen die voldoet aan

$$(3.6) \quad \lim_{n \rightarrow \infty} \frac{1}{k^2} \sum_{i < j} (p_i - p_j) \neq 0.$$

Als voor $n \rightarrow \infty$ k constant blijft dan gaat (3.6) over in

$$(3.7) \quad \theta \stackrel{\text{def}}{=} \sum_{i < j} (p_j - p_i) \neq 0.$$

Wil men als alternatieve hypothesen alleen een dalend verloop beschouwen dan kiest men een rechtséénzijdige kritieke zone, bestaande uit grote positieve waarden van $\underline{W}$; beschouwt men als alternatieven alleen een dalend verloop dan neemt men een linkséénzijdige kritieke zone, bestaande uit grote negatieve waarden van $\underline{W}$.

4. Een generalisatie van de toets

De toetsingsgrootheid $\underline{W}$ die wij in het voorafgaande definiëerden is een bijzonder geval van de grootheid

$$(4.1) \quad \underline{W} = \sum_i g_i \frac{a_i}{n_i},$$

waarbij $g_1, g_2, \dots, g_k$ voor iedere k gegeven getallen zijn, die voldoen aan

$$(4.2) \quad \sum_i g_i = 0.$$

Voor deze grootheid geldt

$$(4.3) \quad E(\underline{W} | t_1, H_0) = 0,$$

$$(4.4) \quad \sigma^2(\underline{W} | t_1, H_0) = \frac{t_1 t_2}{n(n-1)} \sum_i \frac{g_i^2}{n_i}$$

en $\underline{W}$ is onder de hypothese H_0 en onder de voorwaarde $t_1 = t_1$ bij benadering normaal verdeeld met gemiddelde en variantie volgens (4.3) en (4.4).

De tweezijdige toets is (zie [2]) asymptotisch onderscheidend ten opzichte van de klasse van alternatieve hypothesen

$$(4.5) \quad \lim_{n \rightarrow \infty} \sum_i g_i p_i / \sum_i |g_i| \neq 0.$$

Voor vaste k gaat (4.5) over in

$$(4.6) \quad \theta \stackrel{\text{def}}{=} \sum_i g_i p_i \neq 0.$$

De keuze der grootheden $g_1, g_2, \dots, g_k$ hangt dus af van de alternatieve hypothesen die men wenst te beschouwen. Zo kunnen bij het voorbeeld der juryleden de tijdsintervallen tussen de opeenvolgende beoordelingen verschillend zijn; wenst men hiermee bij de toets rekening te houden, dan kan men dit doen door de verschillen $p_i - p_j$ zwaarder te tellen naarmate het tijdsinterval tussen de i^e en de j^e beoordeling groter is. Is T_i het tijdsinterval tussen de i^e en de 1^e beoordeling (dus $T_1 = 0$) dan kan men als alternatieve hypothesen b.v. nemen

$$(4.7) \quad \sum_{i < j} (p_i - p_j) (T_j - T_i) \neq 0.$$

Stellen wij (4.7) gelijk aan θ (zie (4.6)) dan volgt hieruit voor de grootheden $g_1, g_2, \dots, g_k$

$$(4.8) \quad g_i = \sum_j T_j - k T_i \quad i = 1, 2, \dots, k.$$

Zijn de tijdsintervallen tussen de opeenvolgende beoordelingen aan elkaar gelijk dan is $T_i = (i-1) T_2$ ($i = 1, 2, \dots, k$) en (4.8) gaat dan over in

$$(4.9) \quad g_i = \frac{kT_2}{2} (k+1-2i)^{1)} \quad (i = 1, 2, \dots, k).$$

De toets, gebaseerd op de grootheden (4.8), is dus, als de tijdsintervallen alle aan elkaar gelijk zijn, identiek met de in de vorige paragrafen besproken toets.

De exacte verdeling van de grootheid $\underline{W}$, gedefinieerd door (4.1), kan men weer berekenen zoals in par. 2.4 is aangegeven en voor grotere waarden van k en (of) $n_1, n_2, \dots, n_k$ kan men gebruik maken van de normale benadering. Als continuïteitscorrectie kan men hierbij de helft van de kleinste afstand tussen twee opeenvolgende waarden van W nemen; deze kleinste afstand is hier gelijk aan de kleinste positieve waarde van $|(g_i/n_i) - (g_j/n_j)|$.

Opmerking

Als $k = 2$ dan volgt uit (4.2)

$$(4.10) \quad g_2 = -g_1.$$

De door (4.1) gedefinieerde toetsingsgrootheid wordt dan

$$(4.11) \quad \underline{W} = g_1 \left(\frac{a_1}{n_1} - \frac{a_2}{n_2} \right) = \frac{g_1}{n_1 n_2} (a_1 n_2 - n_1 a_2).$$

Hieruit blijkt, dat de toets voor $k = 2$ voor iedere waarde van g_1 identiek is met de methode der 2×2 -tabel (zie b.v. [1]). De afstand tussen twee opeenvolgende waarden van W is hier constant en gelijk aan $g_1 n / n_1 n_2$.

In het voorbeeld van fig. 7 was $k = 2$, $n_1 = 6$ en $n_2 = 4$. Verder was $g_i = k+1-2i$, dus $g_1 = 1$. De afstand tussen twee opeenvolgende waarden van W is hier dus $g_1 n / n_1 n_2 = 5/12$.

5. Het combineren van een aantal onafhankelijke toetsen voor verloop van kansen

5.1. Beschrijving van de methode aan de hand van een voorbeeld.

Stel, bij het voorbeeld der juryleden, dat de kans op goedkeuring niet voor alle juryleden dezelfde is en dat men de hypothese wenst te toetsen dat voor ieder jurylid de kans op goedkeuring constant is in de tijd.

1) Hieruit volgt

$$\sum_{i < j} (p_i - p_j) (j - i) = \frac{k}{2} \sum_{i < j} (p_i - p_j).$$

Deze hypothese, waarbij dus de weinig realistische onderstelling van par. 2.3 vermeden wordt, kunnen wij toetsen tegen de alternatieve hypothesen H_1 , inhoudende dat de kans op goedkeuring een verloop vertoont, dat voor een overwegend aantal der juryleden in dezelfde richting ligt, maar ook, op andere wijze, tegen de alternatieve hypothesen H_2 dat de kans op goedkeuring voor een overwegend aantal der juryleden een verloop vertoont, onverschillig in welke richting.

Om de bovengenoemde hypothese te toetsen berekent men voor ieder jurylid de grootheid $\underline{U}$, zoals in par. 1 en 2 is aangegeven, met zijn gemiddelde en variantie volgens (2.1) en (2.2).¹⁾

Wij vinden dan (zie tabel 1)

jurylid nr.	U	μ	σ^2
1	9	9	15
2	13	10	$16\frac{2}{3}$
3	13	10	$16\frac{2}{3}$
4	4	4	$6\frac{2}{3}$
5	10	7	$11\frac{2}{3}$
6	6	7	$11\frac{2}{3}$

Wil men nu als alternatieven de hypothesen H_1 beschouwen, dan neemt men als toetsingsgrootheid de som van de waarden van $\underline{U}$, die men voor de zes juryleden gevonden heeft. Noemen wij deze grootheid $\underline{U}'$ dan is in ons voorbeeld

$$U' = 9 + 13 + 13 + 4 + 10 + 6 = 55.$$

Onder de hypothese H_0 is deze grootheid $\underline{U}'$ bij benadering normaal verdeeld met als gemiddelde de som van de gemiddelden en als variantie de som van de varianties. Het gemiddelde wordt in het voorbeeld dus $9 + 10 + 10 + 4 + 7 + 7 = 47$ en de variantie wordt: $15 + 16\frac{2}{3} + 16\frac{2}{3} + 6\frac{2}{3} + 11\frac{2}{3} + 11\frac{2}{3} = 78\frac{1}{3}$; de spreiding is dus $\sqrt{78\frac{1}{3}} = 8,85$.

Daar in dit voorbeeld de afstand tussen twee opeenvolgende waarden van de toetsingsgrootheid $\underline{U}'$ gelijk is aan 1 nemen wij als continuïteitscorrectie $\frac{1}{2}$ en de tweezijdige overschrijdingskans is dus bij benadering tweemaal het oppervlak der normale verdeling met gemiddelde 0 en spreiding 1 rechts van

$$\frac{55 - 47 - \frac{1}{2}}{8,85} = 0,85.$$

1) Men kan ook de grootheid $\underline{W}$, gedefinieerd door (2.7) gebruiken, met gemiddelde en variantie volgens (2.8) en (2.9); dit maakt in het voorbeeld geen verschil daar hier $n_i = 1$ voor iedere i en voor ieder jurylid.

In een tabel van de normale verdeling vinden wij voor deze overschrijdingskans 0,40.

Wil men als alternatieve hypothesen alleen een stijging (resp. alleen een daling) beschouwen dan neme men een linkséénzijdige (resp. een rechtséénzijdige) kritieke zone voor U' .

Wil men nu de hypothese, dat voor ieder der juryleden de kans op goedkeuring constant is in de tijd, toetsen tegen de alternatieve hypothesen H_2 , dan berekent men, voor ieder der juryleden, de grootheid $(U - \mu)^2/\sigma^2$. De som van deze grootheden bezit, onder de te toetsen hypothese, bij benadering een χ^2 -verdeling met een aantal vrijheidsgraden dat gelijk is aan het aantal juryleden. De kritieke zone bestaat uit grote waarden van de toetsingsgrootheid.

In het voorbeeld vinden wij voor de toetsingsgrootheid 1,94 en in een tabel van de χ^2 -verdeling met zes vrijheidsgraden vinden wij voor de overschrijdingskans 0,93.

Men kan deze methoden om een aantal toetsen tegen verloop van kansen te combineren ook toepassen als de grootheden $n_i > 1$ zijn. In de volgende paragraaf zullen wij de methoden algemeen formuleren en in par. 5.3 zullen wij beschrijven hoe men de exacte verdelingen van $\underline{U}'$ en $\underline{\chi}^2$, voor kleine aantallen waarnemingen, kan berekenen.

5.2. Algemene formulering.

De in par. 5.1 beschreven methode om een aantal toetsen voor verloop van kansen te combineren kan als volgt algemeen geformuleerd worden.

Gegeven h tabellen van de vorm van tabel 8 (zie par. 3). Iedere tabel bestaat uit k onafhankelijke reeksen van onafhankelijke experimenten, waarbij ieder experiment één der mogelijke uitkomsten succes of mislukking heeft. De i^e reeks van de v^e tabel bestaat uit $n_{i,v}$ experimenten met $\underline{a}_{i,v}$ successen en $\underline{b}_{i,v}$ mislukkingen; $p_{i,v}$ is de kans op succes voor ieder experiment van de i^e reeks ($i = 1, 2, \dots, k$) en

$$(5.1) \quad \underline{t}_{1,v} = \sum_i \underline{a}_{i,v}, \quad \underline{t}_{2,v} = \sum_i \underline{b}_{i,v}, \quad n_v = \sum_i n_{i,v} \quad (v = 1, 2, \dots, h).$$

De hypothese, die wij op grond van deze waarnemingen willen toetsen luidt:

$$(5.2) \quad H_0: p_{1,v} = p_{2,v} = \dots = p_{k,v} \text{ voor iedere } v.$$

Om deze hypothese te toetsen tegen de in par. 5.1 genoemde alternatieven H_1 gebruiken wij de grootheid

$$(5.3) \quad \underline{W}' = \sum_v \underline{W}_v,$$

waarin

$$(5.4) \quad \underline{W}_v = \sum_i (k + 1 - 2i) \frac{a_{i,v}}{n_{i,v}} \quad (\text{zie (3.3)}).$$

Voor deze grootheid $\underline{W}'$ geldt:

$$(5.5) \quad E(\underline{W}' | t_{1,1}, t_{1,2}, \dots, t_{1,h}, H_0) = \sum_v E(\underline{W}_v | t_{1,v}, H_0) = 0 \quad (\text{zie (3.4)}),$$

$$(5.6) \quad \sigma^2(\underline{W}' | t_{1,1}, t_{1,2}, \dots, t_{1,h}, H_0) = \sum_v \sigma^2(\underline{W}_v | t_{1,v}, H_0),$$

waarin

$$(5.7) \quad \sigma^2(\underline{W}_v | t_{1,v}, H_0) = \frac{t_{1,v} t_{2,v}}{n_v (n_{v-1})} \sum_i \frac{(k + 1 - 2i)^2}{n_{i,v}} \quad (\text{zie (3.5)}).$$

Als de hypothese H_0 juist is dan is $\underline{W}'$ bij benadering normaal verdeeld met gemiddelde 0 en variantie volgens (5.6) en (5.7).

De tweezijdige toets is asymptotisch onderscheidend ten opzichte van de klasse van alternatieve hypothesen

$$(5.8) \quad \lim_{N \rightarrow \infty} \frac{1}{h k^2} \sum_v \sum_{i < j} (p_{i,v} - p_{j,v}) \neq 0,$$

waarin

$$N = \sum_v n_v.$$

Wil men als alternatieve hypothesen alleen een stijging (resp. alleen een daling) beschouwen, dan neme men een linkséénzijdige (resp. een rechts-éénzijdige) kritieke zone voor $\underline{W}'$.

Als $n_{i,v} = m_v$ voor iedere i ($v = 1, 2, \dots, h$) dan zou men voor iedere v apart de toets van Wilcoxon kunnen toepassen op de steekproeven A en B (zie par. 2.2 en 2.3) en de toetsen kunnen combineren door de toetsingsgrootheden van Wilcoxon op te tellen. Noemen wij deze toetsingsgrootheid voor de v^e tabel $\underline{U}_v$ en is

$$(5.9) \quad \underline{U}' = \sum_v \underline{U}_v$$

dan geldt

$$(5.10) \quad 2 \underline{U}' = \sum_v m_v^2 \underline{W}_v + \sum_v t_{1,v} t_{2,v} \quad (\text{zie (2.10)}).$$

De toets gebaseerd op $\underline{U}'$ is dus alleen dan identiek met die gebaseerd op $\underline{W}'$ als $m_v = m$ voor iedere v . Is dit niet het geval, dan bestaat, bij gebruik van $\underline{U}'$, weer de mogelijkheid het resultaat te beïnvloeden door manipulatie met

de aantallen m_ν . In het voorbeeld van par. 5.1 was $m = 1$, zodat $\underline{U}'$ en $\underline{W}'$ equivalent zijn.

Een continuïteitscorrectie voor de grootheid W' is in het algemeen moeilijk aan te geven. Als $n_{i,\nu} = m$ voor iedere ν en i dan is de afstand tussen de opeenvolgende waarden van W' constant en gelijk aan m . In dit geval nemen men als continuïteitscorrectie dus $\frac{1}{2}m$ (zie het voorbeeld in par. 5.1). Overigens is voor deze correctie moeilijk een regel te geven.

Als men de hypothese H_0 wil toetsen tegen de in par. 5.1 genoemde alternatieven H_2 dan kan men de toetsingsgrootheid

$$(5.11) \quad \chi^2 = \sum_{\nu} \underline{W}_\nu^2 / \sigma_\nu^2$$

gebruiken, waarin

$$(5.12) \quad \sigma_\nu^2 = \sigma^2(\underline{W}_\nu \mid t_{1,\nu}, H_0) \quad (\text{zie (5.7)}).$$

Deze grootheid bezit, als H_0 juist is, bij benadering een χ^2 -verdeling met h vrijheidsgraden. De kritieke zone bestaat uit grote waarden van χ^2 en de toets leidt voor eindige k en h en voldoende grote N tot verwerping der nulhypothese als voor minstens één waarde van ν geldt

$$(5.13) \quad \sum_{i < j} (p_{i,\nu} - p_{j,\nu}) \neq 0.$$

Als $n_{i,\nu} = m_\nu$ voor iedere i ($\nu = 1, 2, \dots, h$) dan kan men de toets van Wilcoxon toepassen op de steekproeven A en B (zie par. 2.2 en 2.3). Noemen wij de toetsingsgrootheid van Wilcoxon voor de ν^e tabel weer $\underline{U}_\nu$ en is

$$(5.14) \quad \sigma^2(\underline{U}_\nu) = \frac{t_{1,\nu} t_{2,\nu} m_\nu^2 (k^2 - 1)}{12 (n_\nu - 1)} \quad (\text{zie (2.4)})$$

dan geldt

$$(5.15) \quad \chi^2 = \sum_{\nu} \underline{W}_\nu^2 / \sigma_\nu^2 = \sum_{\nu} (\underline{U}_\nu - \frac{1}{2} t_{1,\nu} t_{2,\nu})^2 / \sigma^2(\underline{U}_\nu).$$

De toets verandert hier dus, in tegenstelling tot de toets gebaseerd op $\underline{W}'$, niet bij overgang van $\underline{W}_\nu$ op $\underline{U}_\nu$ en omgekeerd.

5.3. De exacte verdelingen van de grootheden $\underline{W}'$ en χ^2 onder de hypothese H_0 .

De exacte verdelingen van $\underline{W}'$ en χ^2 onder de hypothese H_0 kan men berekenen uit de exacte verdelingen van $\underline{W}_1, \underline{W}_2, \dots, \underline{W}_h$. Wij zullen dit aan de hand van een voorbeeld beschrijven:

Stel $h = 2$, $k = 4$ en

$$n_{1,1} = n_{3,1} = 1, \quad n_{2,1} = n_{4,1} = 3,$$

$$t_{1,1} = t_{1,2} = 1, \quad t_{2,1} = t_{2,2} = 7.$$

$$n_{1,2} = n_{4,2} = 1, \quad n_{2,2} = n_{3,2} = 3,$$

De exacte verdeling van $\underline{W}_1$ en die van $\underline{W}_2$ onder de hypothese H_0 berekenen wij zoals in par. 2.4 is aangegeven. Wij vinden dan

$\underline{W}_1$	$P[\underline{W}_1]$	$\underline{W}_2$	$P[\underline{W}_2]$
-1	$\frac{1}{2}$	-3	$\frac{1}{8}$
$\frac{1}{3}$	$\frac{3}{8}$	$-\frac{1}{3}$	$\frac{3}{8}$
3	$\frac{1}{8}$	$\frac{1}{3}$	$\frac{3}{8}$
		3	$\frac{1}{8}$

De grootheid $\underline{W}' = \underline{W}_1 + \underline{W}_2$ neemt b.v. de waarde -4 aan als $\underline{W}_1 = -1$ en $\underline{W}_2 = -3$ dus

$$P[\underline{W}' = -4] = P[\underline{W}_1 = -1] \cdot P[\underline{W}_2 = -3] = \frac{1}{2} \cdot \frac{1}{8} = 0,0625.$$

De waarde 0 treedt op als $\underline{W}_1 = 3$ en $\underline{W}_2 = -3$ en bovendien als $\underline{W}_1 = \frac{1}{3}$ en $\underline{W}_2 = -\frac{1}{3}$ dus

$$P[\underline{W}' = 0] = P[\underline{W}_1 = 3] \cdot P[\underline{W}_2 = -3] + P[\underline{W}_1 = \frac{1}{3}] \cdot P[\underline{W}_2 = -\frac{1}{3}] = \frac{1}{8} \cdot \frac{1}{8} + \frac{3}{8} \cdot \frac{3}{8} = 0,1563.$$

Op deze wijze kan men voor alle mogelijke waarden van $\underline{W}'$ de kans op die waarde onder de hypothese H_0 berekenen en dit leidt tot

$\underline{W}'$	$P[\underline{W}']$
-4	0,0625
$-2\frac{2}{3}$	0,0469
$-1\frac{1}{3}$	0,1875
$-\frac{2}{3}$	0,1875
0	0,1563
$\frac{2}{3}$	0,1406
2	0,0625
$2\frac{2}{3}$	0,0469
$3\frac{1}{3}$	0,0938
6	0,0156.

Als $h > 2$ is dan berekent men eerst, uit de verdelingen van $\underline{W}_1$ en $\underline{W}_2$, die van $\underline{W}_1 + \underline{W}_2$ en daarna, uit de verdelingen van $\underline{W}_1 + \underline{W}_2$ en $\underline{W}_3$, die van $\underline{W}_1 + \underline{W}_2 + \underline{W}_3$, enz.

De één- en tweezijdige overschrijdingskansen worden gedefinieerd zoals in par. 2.4 is aangegeven.

Om de exacte verdeling van $\chi^2 = (W_1^2/\sigma_1^2) + (W_2^2/\sigma_2^2)$ te vinden berekenen wij eerst, uit de mogelijke waarden van $\underline{W}_1$ en die van $\underline{W}_2$ en met behulp van (5.7) de mogelijke waarden van $\underline{W}_1^2/\sigma_1^2$ en $\underline{W}_2^2/\sigma_2^2$. De exacte verdeling van χ^2 vinden wij dan op analoge wijze als boven is aangegeven voor $\underline{W}' = \underline{W} + \underline{W}_2$. In het voorbeeld vinden wij dan

χ^2	$P[\chi^2]$
$\frac{4}{35}$	0,2813
$\frac{68}{105}$	0,375
$\frac{3^{29}}{35}$	0,0938
$\frac{4^{16}}{35}$	0,125
$\frac{5^{47}}{105}$	0,0938
$\frac{9^9}{35}$	0,0313

De (rechtséénzijdige) overschrijdingskans wordt hier op de gebruikelijke wijze gedefiniëerd als de som van de waarschijnlijkheden van die waarden van χ^2 die niet kleiner zijn dan de gevonden waarde.

Literatuur:

- [1] Van Eeden, Constance, Methoden voor het vergelijken, toetsen en schatten, van onbekende kansen, Statistica 7 (1953), 141-162.
- [2] Van Eeden, Constance en J. Hemelrijk, A test for the equality of probabilities against a class of specified alternative hypotheses, including trend, Proc. Kon. Ned. Akad. v. Wet. A 58 (1955), 191-198 en 301-308.
- [3] Fry, T. C., Probability and its engineering uses, New York (1928).
- [4] Hemelrijk, J., Waarschijnlijkheidsrekening en Statistiek, Vacantiecursus 1950 Mathematisch Centrum.
- [5] Mann H. B. en D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18 (1947), 50-60.
- [6] Statistische tabellen en nomogrammen, uitgegeven onder redactie van de Vereniging voor Statistiek door H. E. Stenfert Kroese N.V., Leiden.
- [7] Wabeke, Ir Doraline en Van Eeden, Constance, Handleiding voor de toets van Wilcoxon, Rapport S 176 (M 65) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam 1955.
- [8] Wilcoxon, F., Individual comparisons by ranking methods, Biometrics 1 (1945), 80-82.

0225 NL