Toets voor de gelijkheid van twee kleine kansen met behulp van even grote steekproeven en het onderscheidingsvermogen van deze toets [1])

door Gerda Klerk-Grobben en H. J. Prins

S u m m a r y

A test for comparing two small unknown probabilities, using samples of equal size, and its power.

For testing the difference between two small proportions a slightly revised form of the sign test is used, giving a better approximation of the significance level α *and a larger power.*

A table of critical values of the test and graphs of its powerfunction are given.

1. Inleiding.

Op bepaalde gebieden van onderzoek doet zich het probleem voor van het vergelijken van kleine waarschijnlijkheden. Wij hebben daarbij b.v. het oog op geologische onderzoekingen, waarbij analyses van bodemlagen worden verricht aan de hand van de soorten stuifmeelkorrels, die daarin voorkomen. Stellen we de fractie van een bepaalde soort stuifmeelkorrels, die in een bepaalde laag voorkomen, gelijk aan θ_1 en die in een andere laag aan θ_2, dan gaat het er om de hypothese

$$H_0 : \theta_1 = \theta_2 \tag{1}$$

te toetsen, op grond van de in twee steekproeven gevonden aantallen. De statistische methoden hiervoor zijn bekend en elders uitvoerig beschreven (zie [1], eerste deel [2])).

In dit rapport wordt een variant van één dezer methoden (en wel een zeer eenvoudige) besproken, welke toepasbaar is wanneer θ_1 en θ_2 klein zijn.

Een der belangrijkste vragen bij dit soort problemen is echter, hoe groot men de omvang van de beide steekproeven moet nemen, om bij gegeven waarden van θ_1 en θ_2, die ongelijk zijn, te bereiken dat men deze ongelijkheid, behoudens een bepaalde kleine waarschijnlijkheid, ook ontdekt. Bij het voorbeeld van de stuifmeelkorrels komt dit dus neer op de vraag hoeveel korrels men uit ieder der twee te vergelijken lagen moet tellen. Deze vraag zal in § 3 nader worden gepreciseerd.

In de volgende §§ worden de toets zelf en grafieken voor de bepaling van

---

[1]) Rapport S 131 van de Statistische Afdeling van het Mathematisch Centrum. De Afdeling staat onder leiding van Prof. Dr D. v a n D a n t z i g.

[2]) Cijfers tussen vierkante haken zijn een verwijzing naar de literatuurlijst aan het einde van dit rapport.

de steekproefgrootte besproken en aan de hand van enkele voorbeelden nader toegelicht. De appendix bevat wiskundige afleidingen.

2. Toets voor $\theta_1 = \theta_2$, als θ_1 en θ_2 klein zijn.

Het waarnemingsmateriaal bestaat uit twee steekproeven, waarvan wij eerst onderstellen, dat ze even groot zijn en wel beide uit een groot aantal, N, exemplaren bestaan. (Dus b.v. twee steekproeven van ieder N stuifmeelkorrels.) In elk der steekproeven wordt geteld, hoeveel exemplaren er in voorkomen van een in beide steekproeven zeldzame soort A. Laten dit er n_1 en n_2 zijn; we stellen

$$n_1 + n_2 = m \tag{2}$$

Bij de toetsing van de hypothese $H_0\,(\theta_1 = \theta_2)$ handelen we nu alsof we m maal met een munt geworpen hebben, waarvan we de zuiverheid willen toetsen, en alsof van deze m worpen n_1 worpen het resultaat „kruis'' hebben gegeven. Voor de toepassing van de toets is een tabel van kritieke waarden voor n_1 bij verschillende onbetrouwbaarheidsdrempels toegevoegd (tabel I). De opgegeven onbetrouwbaarheidsdrempels behoren bij éénzijdige toetsing (zie hiernaast). De in tabel I opgegeven kritieke waarden komen niet precies overeen met die van de gebruikelijke binomiale toets voor de zuiverheid van een munt (de tekentoets). In het onderhavige geval was een kleine vergroting van de kritieke zones mogelijk zonder de onbetrouwbaarheidsdrempel te overschrijden. Details hierover vindt men in de appendix (§ 4). Voor een opmerking over het bepalen van overschrijdingskansen wordt eveneens naar § 4 verwezen.

Leidt de toepassing van deze toets tot verwerping, dan wordt dus de hypothese $H_0\,(\theta_1 = \theta_2)$ verworpen. De grootte van N heeft op de toets geen invloed; de toets is echter slechts van toepassing als N veel groter dan m is.

Een- en tweezijdige toetsing. We spreken van *tweezijdige* toetsing indien het kritieke gebied bestaat uit grote zowel als kleine waarden van de toetsingsgrootheid. In ons geval moet tweezijdige toetsing worden gebruikt, indien we tot verwerping van $H_0\,(\theta_1 = \theta_2)$ willen komen wanneer $\theta_1 < \theta_2$ en ook wanneer $\theta_1 > \theta_2$. Weten we echter reeds van te voren, dat alleen $\theta_1 \leqq \theta_2$ kan optreden, zodat $\theta_1 > \theta_2$ zeker niet juist is, dan kunnen we een *éénzijdige* toets gebruiken; het kritieke gebied zal dan alleen uit kleine waarden van n_1 bestaan. (Is alleen $\theta_1 \geqq \theta_2$ mogelijk, dan wordt het kritieke gebied door grote waarden van n_1 gevormd.) Het gebruik van een éénzijdige toets is ook dan gerechtvaardigd, indien het b.v. voor het onderzoek uitsluitend van belang is na te gaan of $\theta_1 < \theta_2$, terwijl $\theta_1 = \theta_2$ en $\theta_1 > \theta_2$ voor de onderzoeker

TABEL I. *Kritieke waarden van de linkseenzijdige toets voor gelijkheid van kleine kansen*

m	onbetrouwbaarheidsdrempel 0,005	0,01	0,025	0,05	m	onbetrouwbaarheidsdrempel 0,005	0,01	0,025	0,05
1	—	—	—	—	51	16	17	18	19
2	—	—	—	—	52	16	17	18	20
3	—	—	—	—	53	17	18	19	20
4	—	—	—	0	54	17	18	19	20
5	—	—	0	0	55	17	18	20	21
6	—	0	0	0	56	18	19	20	21
7	0	0	0	1	57	18	19	21	22
8	0	0	1	1	58	19	20	21	22
9	0	0	1	1	59	19	20	21	23
10	0	1	1	2	60	20	20	22	23
11	1	1	2	2	61	20	21	22	24
12	1	1	2	3	62	20	21	23	24
13	1	2	2	3	63	21	22	23	24
14	2	2	3	3	64	21	22	24	25
15	2	2	3	4	65	22	23	24	25
16	2	3	4	4	66	22	23	25	26
17	3	3	4	5	67	22	23	25	26
18	3	4	4	5	68	23	24	25	27
19	3	4	5	5	69	23	24	26	27
20	4	4	5	6	70	24	25	26	28
21	4	5	5	6	71	24	25	27	28
22	4	5	6	7	72	25	26	27	28
23	5	5	6	7	73	25	26	28	29
24	5	6	7	7	74	25	26	28	29
25	6	6	7	8	75	26	27	28	30
26	6	7	7	8	76	26	27	29	30
27	6	7	8	9	77	27	28	29	31
29	7	7	8	9	78	27	28	30	31
29	7	8	9	9	79	28	29	30	32
30	7	8	9	10	80	28	29	31	32
31	8	8	9	10	81	28	30	31	33
32	8	9	10	11	82	29	30	32	33
33	9	9	10	11	83	29	30	32	33
34	9	10	11	12	84	30	31	32	34
35	9	10	11	12	85	30	31	33	34
36	10	10	12	13	86	31	32	33	35
37	10	11	12	13	87	31	32	34	35
38	11	11	12	13	88	31	33	34	36
39	11	12	13	14	89	32	33	35	36
40	11	12	13	14	90	32	33	35	37
41	12	13	14	15	91	33	34	36	37
42	12	13	14	15	92	33	34	36	38
43	13	13	15	16	93	34	35	37	38
44	13	14	15	16	94	34	35	37	38
45	13	14	15	16	95	34	36	37	39
46	14	15	16	17	96	35	36	38	39
47	14	15	16	17	97	35	37	38	40
48	15	15	17	18	98	36	37	39	40
49	15	16	17	18	99	36	37	39	41
50	15	16	18	19	100	37	38	40	41

(en voor de te nemen beslissing) op hetzelfde neerkomen. Men toetst dan eigenlijk de hypothese $\theta_1 \geqq \theta_2$ tegen $\theta_1 < \theta_2$ en doet dit met een linkeréénzijdig kritiek gebied.

In tabel I zijn linkeréénzijdige kritieke gebieden opgegeven. De rechtséénzijdige kritieke waarden (voor toetsing van $H_0 : \theta_1 \leqq \theta_2$ tegen $\theta_0 > \theta_2$) zijn in deze tabel te vinden door verwisseling van n_1 en n_2 of door de in de tabel opgegeven waarden van m af te trekken.

Ook bij tweezijdige toetsing kan de tabel gebruikt worden. De kritieke zone bestaat dan uit het opgegeven gebied èn het overeenkomstige gebied van hoge waarden. De opgegeven onbetrouwbaarheidsdrempels moeten dan verdubbeld worden.

Voorbeeld. Stel we willen H_0 ($\theta_1 \geqq \theta_2$) toetsen tegen $\theta_1 < \theta_2$ en vinden in de steekproeven respectievelijk $n_1 = 6$ en $n_2 = 14$ exemplaren van de desbetreffende soort. Hierbij is dus $n_1 + n_2 = m = 20$. In tabel I vinden we dan dat bij $m = 20$ en onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ de kritische waarde voor $n_1 = 6$ is. D.w.z. dat, indien de nulhypothese $\theta_1 = \theta_2$ *waar* is, de kans om voor n_1 een waarde $\leqq 6$ te vinden hoogstens 0,05 is; vinden we een dergelijke „onwaarschijnlijke'' waarde van n_1, dan zullen we de nulhypothese verwerpen ten gunste van $\theta_1 < \theta_2$; zoals dus in ons voorbeeld. Kiezen we echter als onbetrouwbaarheidsdrempel $\alpha = 0{,}025$, dan bestaat het kritieke gebied uit waarden $n_1 \leqq 5$; in dit geval zullen we dus op grond van bovenstaande steekproeven niet tot verwerping van de hypothese $\theta_1 = \theta_2$ overgaan. Voor tweezijdige toetsing ($H_0 : \theta_1 = \theta_2$ tegen $\theta_1 < \theta_2$ èn $\theta_1 > \theta_2$) bestaat het kritieke gebied voor $\alpha = 0{,}05$ en $m = 20$ uit waarden voor $n_1 : \leqq 5$ en $\geqq 20 - 5 = 15$ (in tabel I de kolom onder $\alpha = 0{,}025$ gebruiken). Bij tweezijdige toetsing zou de nulhypothese volgens bovenstaande steekproeven en onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ dus *niet* verworpen worden.

Steekproeven van ongelijke omvang. Tot nu toe hebben we ondersteld dat de steekproeven beide uit evenveel exemplaren (N) bestaan. Is dit echter niet het geval, bestaan dus de steekproeven uit N_1 resp. N_2 exemplaren met $N_1 \neq N_2$, dan kan als N_1 en N_2 sterk verschillen, de beschreven toets niet meer op n_1 en n_2 worden toegepast. (De toets berust nl. op een binomiale verdeling met gelijke kansen, terwijl nu een binominale verdeling met kansen $N_1/(N_1 + N_2)$ en $N_2/(N_1 + N_2)$ gebruikt behoort te worden, of een benadering hiervan; zie appendix, § 4.) Zijn N_1 en N_2 echter groot en verschillen ze slechts weinig (b.v. $N_1 = 1000$ en $N_2 = 985$), dan zal de fout, die gemaakt wordt wanneer we bovenstaande toets toch toepassen op n_1 en n_2, slechts gering zijn. Op het geval van ongelijke steekproeven gaan we, behoudens nog enkele opmerkingen in de appendix, verder niet in.

Enkele opmerkingen.

1) De in deze § beschreven toets kan alleen gebruikt worden indien de kansen θ_1 en θ_2 klein zijn (b.v. $< 0{,}1$) en N groot is. In andere gevallen is de in de appendix beschreven benadering van de binomiale verdelingen (met kansen θ_1 en $1 - \theta_1$ resp. θ_2 en $1 - \theta_2$) door een Poisson-verdeling niet goed meer en kunnen beter de gebruikelijke methoden der dubbele dichotomie (2×2-tabel) gebruikt worden (zie [1]).

2) Ook bij meer dan twee steekproeven kunnen methoden, zij het minder eenvoudige, voor de toetsing van de gelijkheid van kleine kansen worden ontwikkeld (zie b.v. [2]).

3. De keuze van N.

In deze § behandelen we de vraag naar de grootte van N. We behandelen hierbij alleen het geval $N_1 = N_2 = N$, daar de in § 2 beschreven toets alleen hiervoor geldt en daar dit bovendien tot de minste waarnemingen leidt. Is $N_1 \neq N_2$ dan moet het totale aantal waarnemingen, $N_1 + N_2$, groter zijn dan wanneer $N_1 = N_2$ om dezelfde doeltreffendheid te bereiken en dit wordt sterker naarmate N_1 en N_2 meer verschillen. Men zal dus zo mogelijk even grote steekproeven nemen.

In principe berust de keuze van N in problemen als het onderhavige op de wens, de ongelijkheid van θ_1 en θ_2 met minstens een bepaalde kans te ontdekken, indien θ_1 en θ_2 bepaalde van elkaar verschillende waarden bezitten. Deze kans, β, om tot verwerping van de getoetste hypothese, $\theta_1 = \theta_2$, te komen, wordt het *onderscheidingsvermogen* van de toets genoemd en het gaat dus eigenlijk om kennis van dit onderscheidingsvermogen.

Voor twee gevallen, nl. voor de *eenzijdige* toets met onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ en voor de *eenzijdige* toets met $\alpha = 0{,}025$ is het onderscheidingsvermogen berekend als functie van de alternatieve hypothesen. Voor *tweezijdige* toetsing is dit het onderscheidingsvermogen bij respectievelijk $\alpha = 0{,}10$ en $\alpha = 0{,}05$. Bij andere onbetrouwbaarheidsdrempels verlopen de berekeningen volkomen analoog; grotere waarden van α zullen tot een groter en kleinere waarden van α tot een kleiner onderscheidingsvermogen bij dezelfde alternatieve hypothese leiden.

Het blijkt dat het onderscheidingsvermogen β bepaald wordt door de grootheden:

(3) $$\mu_1 \stackrel{def}{=} N\theta_1 \text{ en } \mu_2 \stackrel{def}{=} N\theta_2$$

waarin N de (nog te kiezen) omvang van de steekproeven is. Verder blijkt het nuttig de grootheid:

(4) $$k \stackrel{def}{=} \frac{\theta_2}{\theta_1} = \frac{\mu_2}{\mu_1}$$

in te voeren. De in de appendix (§ 5) beschreven berekeningen voeren tot de in de figuren 1*a* en 1*b* samengevatte numerieke resultaten. In deze figuren is het onderscheidingsvermogen β bij gegeven k uitgezet als functie van μ_1.

Het gebruik van de figuren voor de bepaling van N is het eenvoudigste te beschrijven aan de hand van een voorbeeld.

Voorbeeld. We onderstellen dat vroegere ervaring of een vooronderzoek geleerd hebben, dat de fracties θ_1 en θ_2 van de orde 0,01 zijn en dat men nu met kans 0,90 wil kunnen onderscheiden tussen fracties $\theta_1 = 0{,}02$ en $\theta_2 = 0{,}005$. Is al van te voren bekend welke van de twee steekproeven bij de grootste fractie zal horen, indien deze ongelijk zijn (dus welke bodemlaag in dat geval de grootste fractie stuifmeelkorrels van de soort A bezit), dan toetsen we *eenzijdig* (zie § 2). Toetsen we met $\alpha = 0{,}05$, dan moet dus figuur 1*b* met

$$k = \frac{\theta_2}{\theta_1} = \frac{0{,}005}{0{,}02} = \tfrac{1}{4}$$

gebruikt worden. Uit deze figuur lezen we dan af, dat om een onderscheidingsvermogen $\beta = 0{,}90$ te bereiken bij $k = \frac{1}{4}$, μ_1 minstens ongeveer 18 moet bedragen.

Volgens (3) moet dan

$$N \geqq \frac{\mu_1}{\theta_1} \approx \frac{18}{0{,}02} = 900$$

zijn.

Is niet van te voren bekend bij welke laag de grootste fractie korrels van de bepaalde soort behoort, dan moet er *tweezijdig* getoetst worden. Voor dezelfde onbetrouwbaarheidsdrempel $\alpha = 0{,}05$, moeten we nu figuur 1*a* gebruiken. Bij $k = \frac{1}{4}$ vinden we dan bij $\beta = 0{,}90 : \mu_1 \approx 23$ dus

$$N \geqq \frac{\mu_1}{\theta_1} \approx \frac{23}{0{,}02} = 1150.$$

Opmerkingen.

1) De figuren bezitten geen grote nauwkeurigheid, daar de krommen een-

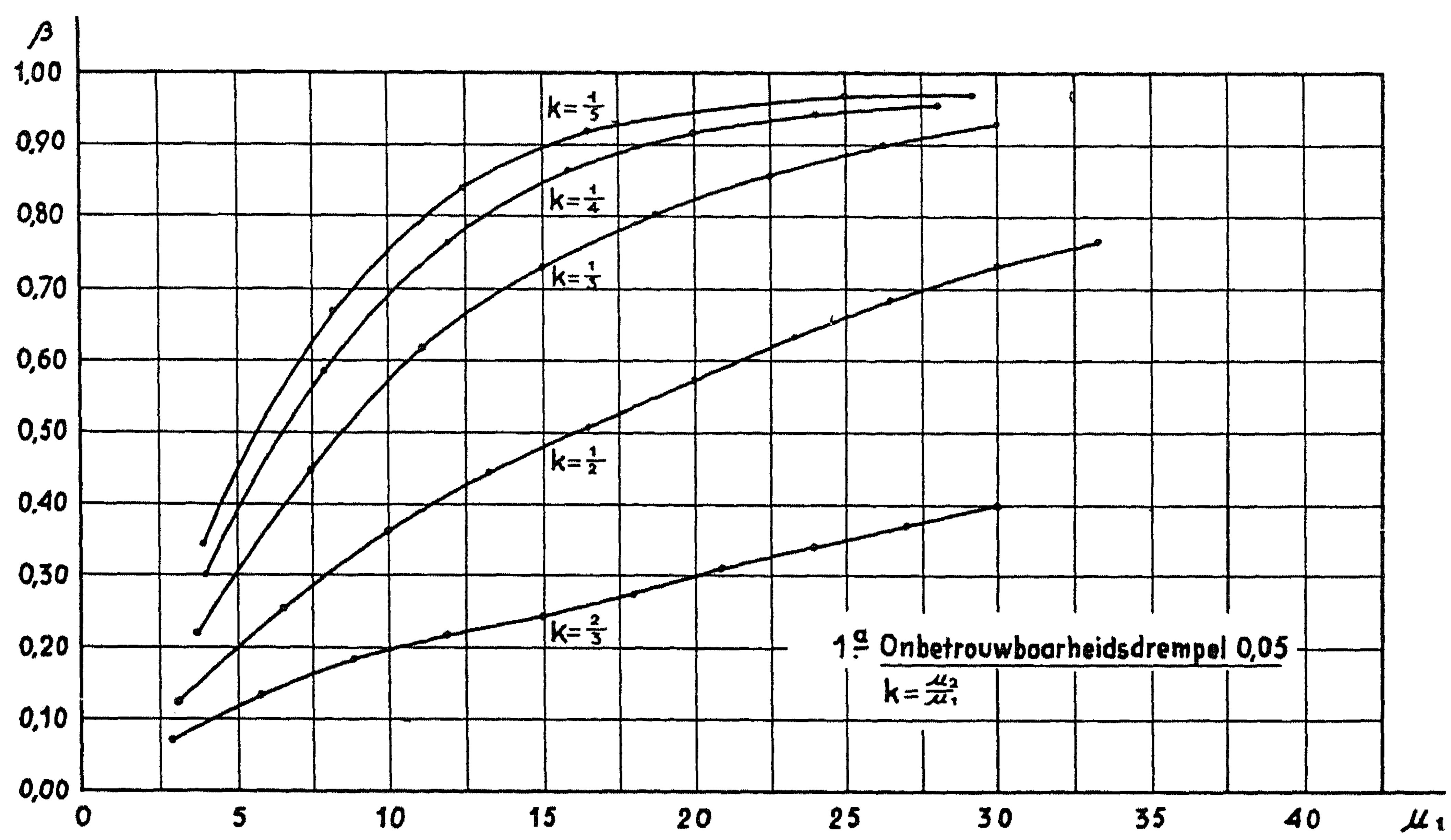
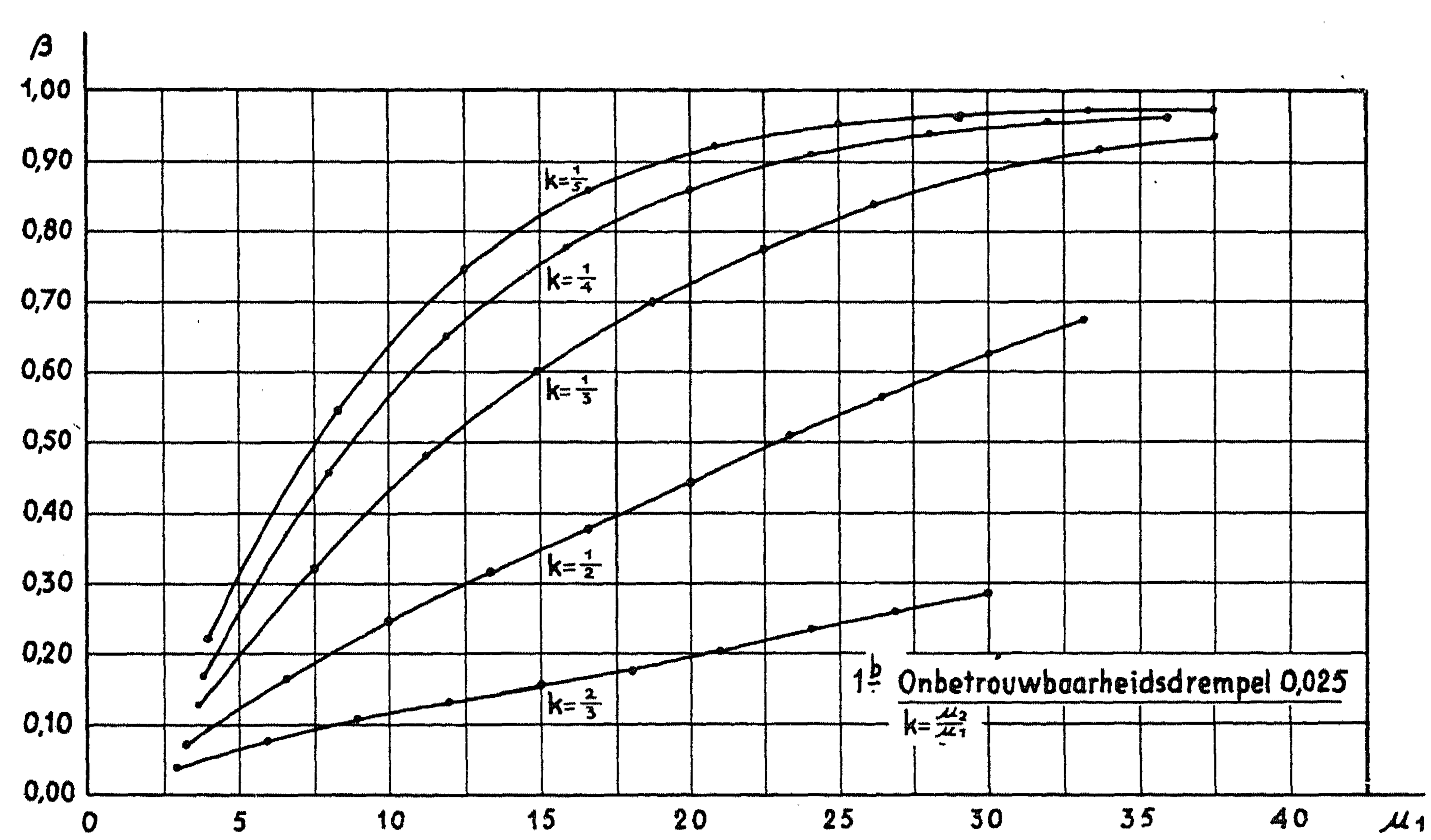

β
1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
0
5
10
15
20
25
30
35
40
μ_1
k = 4/5
k = 1/4
k = 1/3
k = 1/2
k = 2/3
1ª Onbetrouwbaarheidsdrempel 0,05
k = μ_2/μ_1

Fig. 1a

Fig. 1b


β
1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
0
5
10
15
20
25
30
35
40
μ_1
k = 1/5
k = 1/4
k = 1/3
k = 1/2
k = 2/3
1ᵇ Onbetrouwbaarheidsdrempel 0,025
k = μ_2/μ_1

voudig als vloeiende lijnen door enkele, verdikt weergegeven, berekende punten getrokken zijn. Voor de toepassingen behoeft echter het onderscheidingsvermogen in de regel niet met grote precisie bekend te zijn.

2) De berekening van het onderscheidingsvermogen geldt, evenals de toepasbaarheid van de toets, alleen indien θ_1 en θ_2 klein zijn. Voor grotere θ's is het onderscheidingsvermogen van de toetsing met de methode der dubbele dichotomie benaderd door Prof. Dr D. van Dantzig [3] en P. B. Patnaik [4].

Appendix.

4. Afleiding van de toets.

Is de kans om uit een populatie van exemplaren (b.v. een bodemlaag met stuifmeelkorrels) een exemplaar van een bepaalde soort A te trekken θ, dan wordt de kans, dat men in een steekproef van N exemplaren er n van soort A aantreft, gegeven door de binomiale verdeling:

$$\mathrm{P}\,[\underline{n} = n] = \binom{N}{n} \theta^n (1 - \theta)^{N-n}$$ [1])

Bij vergelijking van twee steekproeven uit verschillende populaties zullen we de kansen op A aangeven met θ_1 resp. θ_2 en de aantallen A-exemplaren in de steekproeven met n_1 resp. n_2. De hypothese H_0, die we willen toetsen, luidt dan: $\theta_1 = \theta_2$.

Het feit dat de kansen θ in de binominale verdeling klein zijn bij de door ons bekeken gevallen, maakt een benadering door middel van Poissonverdelingen mogelijk, te meer daar men grote steekproeven zal moeten nemen om nog een redelijke kans te hebben verschillen tussen twee kleine kansen te ontdekken (zie §§ 3 en 5).

De Poisson-benadering van de binominale verdeling ontstaat door de limiet van de binominale verdeling te nemen voor $N \to \infty$, waarbij tegelijkertijd θ zodanig naar nul gaat, dat $N\theta$ tot een constant getal, μ, nadert. Voor alle gehele waarden van $n \geqq 0$ wordt dan de kans op n A-exemplaren bij benadering gegeven door:

$$\mathrm{P}\,[\underline{n} = n] = \frac{e^{-\mu} \mu^n}{n!} \qquad \text{(Poisson-verdeling).}$$

---

[1]) Een *stochastische grootheid*, d.i. een grootheid, die geen vaste waarde bezit maar aan een waarschijnlijkheidsverdeling onderworpen is, geven we aan door een onderstreepte letter; $\underline{n}$ b.v. is hier het aantal exemplaren A in een te nemen steekproef. Een door $\underline{n}$ aangenomen waarde (in één bepaalde steekproef) geven we aan door n zonder onderstreping. $\mathrm{P}\,[\underline{n} = n]$ betekent: de kans, dat de stochastische grootheid $\underline{n}$ de waarde n aanneemt.

In ons geval is dus voor de eerste steekproef:

$$P[\underline{n}_1 = n_1] = \frac{e^{-N_1\theta_1}(N_1\theta_1)^{n_1}}{n_1!} \tag{5}$$

en voor de tweede steekproef:

$$P[\underline{n}_2 = n_2] = \frac{e^{-N_2\theta_2}(N_2\theta_2)^{n_2}}{n_2!}. \tag{6}$$

Daar $\underline{n}_1$ en $\underline{n}_2$ onderling onafhankelijk verdeeld zijn, wordt de kans op een waarnemingsresultaat (n_1, n_2) gegeven door:

$$P[\underline{n}_1 = n_1, \underline{n}_2 = n_2] = P[\underline{n}_1 = n_1] \,.\, P[\underline{n}_2 = n_2].$$

De voorwaardelijke kans op (n_1, n_2) onder de voorwaarde $\underline{n}_1 + \underline{n}_2 = m$ is dan:

$$P[\underline{n}_1 = n_1, \underline{n}_2 = n_2 \mid \underline{n}_1 + \underline{n}_2 = m] = P[\underline{n}_1 = n_1, \underline{n}_2 = m - n_1] = \\ = \frac{P[\underline{n}_1 = n_1]\, P[\underline{n}_2 = n_2]}{P[\underline{n}_1 + \underline{n}_2 = m]}.$$

Nu heeft de som van twee volgens Poisson verdeelde grootheden weer een Poisson-verdeling met als parameter de som van hun parameters, dus:

$$P[\underline{n}_1 + \underline{n}_2 = m] = \frac{e^{-N_1\theta_1 - N_2\theta_2}(N_1\theta_1 + N_2\theta_2)^m}{m!}, \tag{7}$$

zodat:

$$P[\underline{n}_1 = n_1, \underline{n}_2 = m - n_1] = \frac{e^{-N_1\theta_1}(N_1\theta_1{}^{n_1} e^{-N_2\theta_2})\, N_2\theta_2{}^{n_2}\, m!}{n_1! \quad n_2!\, e^{-N_1\theta_1 - N_2\theta_2}(N_1\theta_1 + N_2\theta_2)^m} = \\ = \frac{m!}{n_1!\, n_2!}\left(\frac{N_1\theta_1}{N_1\theta_1 + N_2\theta_2}\right)^{n_1}\left(\frac{N_2\theta_2}{N_1\theta_1 + N_2\theta_2}\right)^{n_2}. \tag{8}$$

Onder de hypothese $H_0\,(\theta_1 = \theta_2)$ gaat dit over in

$$P[\underline{n}_1 = n_1, \underline{n}_2 = m - n_1 \mid H_0] = \binom{m}{n_1}\left(\frac{N_1}{N_1 + N_2}\right)^{n_1}\left(\frac{N_2}{N_1 + N_2}\right)^{n_2} \tag{9}$$

en als we uitgaan van steekproeven met gelijke omvang $(N_1 = N_2 = N)$ dan wordt dit nog vereenvoudigd tot:

$$P[\underline{n}_1 = n_1, \underline{n}_2 = m - n_1 \mid H_0] = \binom{m}{n_1}(\tfrac{1}{2})^m, \tag{10}$$

dus een binomiale verdeling met gelijke kansen.

Bij iedere waarde van m is nu met behulp van deze binomiale verdeling een kritiek gebied voor $\underline{n}_1$ met gegeven onbetrouwbaarheidsdrempel α te bepalen. Is de verdeling van de toetsingsgrootheid, $\underline{n}_1$, discreet (d.w.z. dat $\underline{n}_1$ slechts discrete waarden kan aannemen), dan is het in het algemeen niet mogelijk het kritieke gebied Z_m zo te kiezen, dat de onbetrouwbaarheid precies α wordt. Gewoonlijk wordt dan voor Z_m het grootste gebied gekozen, waarvoor nog geldt:

$$\mathrm{P}\,[\underline{n}_1 \in Z_m | m, H_0] \leqq \alpha. \tag{11}$$

In dit rapport zullen we α de onbetrouwbaarheidsdrempel blijven noemen en voor de waarde van $\mathrm{P}\,[\underline{n}_1 \in Z_m | m, H_0]$ de naam voorwaardelijke (onder voorwaarde $\underline{n}_1 + \underline{n}_2 = m$) onbetrouwbaarheid en het symbool α_{m0} gebruiken. De onvoorwaardelijke onbetrouwbaarheid geven we dan met α_0 aan:

$$\alpha_0 = \sum_{m=0}^{\infty} \alpha_{m0}\, \mathrm{P}\,[\underline{n}_1 + \underline{n}_2 = m | H_0].$$

(De onbetrouwbaarheid α_0 hangt dus ook af van N_1 en N_2, zie (7).) Worden de kritieke gebieden Z_m zo gekozen (oude methode) dat ze aan (11) voldoen, dan blijkt de onvoorwaardelijke onbetrouwbaarheid α_0 belangrijk kleiner dan α te zijn voor de meest voorkomende waarden van m ($m < 100$). Door een iets gewijzigde keuze van de kritieke zones Z_m (nieuwe methode) is het nu mogelijk de onbetrouwbaarheidsdrempel α beter te benaderen (en hierdoor het onderscheidingsvermogen van de toets te vergroten, zie § 5). We kiezen dan nl., bij iedere m, Z_m zo, dat α_{m0} de onbetrouwbaarheidsdrempel α het beste benadert, waarbij nu dus in een aantal gevallen wel $\alpha_{m0} > \alpha$ kan worden. Dit komt er op neer dat aan een aantal kritieke gebieden Z_m één waarde van n_1 wordt toegevoegd (nl. als α_{m0} zeer veel kleiner dan α was). Enkele van deze gevallen zijn in tabel II vermeld.

TABEL II. *Vergelijking van het kritieke gebied, Z_m, voor $\underline{n}_1$ volgens oude en volgens nieuwe methode bepaald* (zie tekst)

m	linkseenzijdige toetsing				tweezijdige toetsing			
	oude methode $\alpha_{m0} \leqq \alpha = 0{,}05$		nieuwe methode $\alpha_{m0} \approx \alpha = 0{,}05$		oude methode $\alpha_{m0} \leqq \alpha = 0{,}05$		nieuwe methode $\alpha_{m0} \approx \alpha = 0{,}05$	
	Z_m	α_{m0}	Z_m	α_{m0}	Z_m	α_{m0}	Z_m	α_{m0}
7	$= 0$	0,008	$\leqq 1$	0,063	$= 0;\ = 7$	0,016	$= 0;\ = 7$	0,016
8	$\leqq 1$	0,035	$\leqq 1$	0,035	$= 0;\ = 8$	0,008	$\leqq 1;\ \geqq 7$	0,070
10	$\leqq 1$	0,011	$\leqq 2$	0,055	$\leqq 1;\ \geqq 9$	0,022	$\leqq 1;\ \geqq 9$	0,022
11	$\leqq 2$	0,033	$\leqq 2$	0,033	$\leqq 1;\ \geqq 10$	0,012	$\leqq 2;\ \geqq 9$	0,065
32	$\leqq 10$	0,025	$\leqq 11$	0,055	$\leqq 9;\ \geqq 23$	0,020	$\leqq 10;\ \geqq 22$	0,050
41	$\leqq 14$	0,030	$\leqq 15$	0,059	$\leqq 13;\ \geqq 28$	0,028	$\leqq 14;\ \geqq 27$	0,060

De kritieke gebieden voor $\underline{n}_1$, welke voor $m = 1$ tot 100 in tabel I zijn opgegeven, zijn volgens dit laatste voorschrift gekozen. Ter vergelijking is in tabel III voor enkele waarden van μ_1 de onbetrouwbaarheid α_0 bij gebruik van beide methodes opgegeven.

TABEL III. *Vergelijking van de onbetrouwbaarheid α_0 bij gebruik van kritieke zones volgens de oude en de nieuwe methode* (zie tekst)

μ_1	linkseenzijdige toetsing			
	$\alpha = 0,025$		$\alpha = 0,05$	
	oude methode	nieuwe methode	oude methode	nieuwe methode
2,5	0,005	0,012	0,012	0,030
5	0,011	0,020	0,023	0,042
7,5	0,014	0,023	0,030	0,047
15	0,015	0,022	0,036	0,046

We zien hierin dat de gewijzigde zones een veel betere benadering van α geven, terwijl de verwachting gerechtvaardigd is, dat de onbetrouwbaarheid α_0 ook voor andere waarden van μ_1 slechts weinig van de onbetrouwbaarheidsdrempel α zal verschillen. Op deze manier is dus bij dezelfde onbetrouwbaarhiedsdrempel het kritieke gebied en daarmee het onderscheidingsvermogen vergroot.

Opmerking. De nieuwe methode voor het bepalen van kritieke gebieden zou ook gebruikt kunnen worden voor het bepalen van overschrijdingskansen, als men de overschrijdingskans, k, bij een gevonden waarde van de toetsingsgrootheid $\underline{n}_1$, definieert als de kleinste waarde van de onbetrouwbaarheidsdrempel waarvoor het bijbehorende kritieke gebied nog de gevonden waarde n_1 bevat. Deze methode voor de berekening van overschrijdingskansen wordt echter vrij bewerkelijk. In de regel zal men kunnen volstaan met het bepalen van de overschrijdingskans volgens de oude methode, dat wil dus zeggen:

$$k = \mathrm{P}\,[\underline{n}_1 \geqq n_1 \mid H_0, m],$$

bij gebruik van rechter éénzijdige kritieke gebieden. De zo bepaalde waarde van k zal in het algemeen iets groter zijn dan men volgens de nieuwe methode zou vinden.

Is $N_1 \neq N_2$ dan geldt voor de toetsingsgrootheid uitdrukking (9); dit is dus een binomiale verdeling met als kansen $N_1/(N_1 + N_2)$ en $N_2/(N_1 + N_2)$, die beide uit het waarnemingsmateriaal bekend zijn. Bij de bepaling van de kritieke gebieden moeten dus tabellen van deze verdelingen gebruikt worden. Zijn deze niet beschikbaar, dan kan, wanneer N_1 en N_2 niet veel uiteenlopen, toch op n_1 en n_2 de beschreven toets worden toegepast. Zijn N_1 en N_2 wel sterk verschillend, is dus één van de kansen $N_1/N_1 + N_2$ en $N_2/N_1 + N_2$ klein, dan kan de verdeling (9) beter met een Poisson-verdeling, zoals we in het begin van deze paragraaf bespraken, worden benaderd. Wij gaan op deze kwesties niet nader in; zij worden uitvoerig besproken in [1].

5. Berekening van het onderscheidingsvermogen.

Zoals reeds in § 3 werd vermeld, is het onderscheidingsvermogen alleen berekend voor het geval $N_1 = N_2 = N$. We zullen beginnen met een wat preciezere definitie:

Het *onderscheidingsvermogen*, β, van een toets ten opzichte van een alternatieve hypothese, welke bepaalde ongelijke waarden aan θ_1 en θ_2 toekent, is de kans op een juiste beslissing. D.w.z. bij *eenzijdige* toetsing is β de kans dat de toetsingsgrootheid een waarde in het kritieke gebied aanneemt, indien de bijbehorende alternatieve hypothese juist is; bij *tweezijdige* toetsing is β de kans dat de toetsingsgrootheid een waarde aanneemt van dat deel van de twee waaruit het kritieke gebied bestaat, dat tot een juiste conclusie omtrent θ_1 en θ_2 leidt. B.v.: is de alternatieve hypothese $\theta_1 = 0{,}02$ en $\theta_2 = 0{,}005$ dan is β de kans, dat n_1 in het rechter deel van het kritieke gebied zal liggen, berekend onder deze hypothese, zodat de toets tot verwerping van H_0 ($\theta_1 = \theta_2$) ten gunste van $\theta_1 > \theta_2$ zal leiden.

Uit deze definities van het onderscheidingsvermogen volgt dat β bij tweezijdige toetsing met onbetrouwbaarheidsdrempel α gelijk is aan β bij eenzijdige toetsing met onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha$.

Het onderscheidingsvermogen zal afhangen van θ_1, θ_2 en N en wel zoals zal blijken van de waarden van $N_1\theta_1$ en $N_2\theta_2$. Bovendien zal het onderscheidingsvermogen afhangen van de waarde van α, die immers de grootte van het kritieke gebied bepaalt: bij een kleinere waarde van α zal een kleinere waarde van β behoren.

De nu volgende afleiding van het onderscheidingsvermogen is ook te vinden bij J. Przyborowski en H. Wilenski [5], die β berekenden voor dezelfde toets, maar gebaseerd op kritieke gebieden welke aan (11) voldoen (zie § 4), waardoor β dus kleiner wordt dan bij de hier gevolgde methode.

Evenals in § 3, definiëren we:

$$\mu_1 = N\theta_1 \; ; \mu_2 = N\theta_2 \; ; k = \frac{\theta_2}{\theta_1} \text{ en } \underline{m} = \underline{n}_1 + \underline{n}_2. \tag{12}$$

De voorwaardelijk kans (voorwaarde $\underline{m} = m$) op een waarnemingsresultaat (n_1, n_2) (zie (8)), gaat hierdoor over in

$$\mathrm{P}\,[\underline{n}_1 = n_1, \underline{n}_2 = m - n_1] = \binom{m}{n_1} \left(\frac{1}{1+k}\right)^{n_1} \left(\frac{k}{1+k}\right)^{m-n_1},$$

een binomiale verdeling met kansen $1/(1 + k)$ en $k/(1 + k)$. Door sommatie van deze kansen $\mathrm{P}\,[\underline{n}_1 = n_1, \underline{n}_2 = m - n_1]$ over het voorwaardelijke kritieke gebied Z_m, vinden we het voorwaardelijk onderscheidingsvermogen β_m. Het onvoorwaardelijke onderscheidingsvermogen β is dan:

$$\beta = \sum_{m=0}^{\infty} \beta_m \,\mathrm{P}\,[\underline{m} = m] \tag{13}$$

waarin volgens (7) en (12):

$$\mathrm{P}\,[\underline{m} = m] = \frac{e^{-(1+k)\mu_1}\{(1 + k)\mu_1\}^m}{m!}$$

geldt (een Poisson-verdeling met parameter $(1 + k)\mu_1$). De grootheden μ_1 en k zijn hierin door de alternatieve hypothese en N bepaald.

Met behulp van tabellen van de Poisson-verdeling en van binomiale verdelingen met ongelijke kansen (zie [6] en [7]) is nu bij gegeven μ_1, k en α het onderscheidingsvermogen numeriek te bepalen. De resultaten zijn verwerkt in de figuren *1a* en *1b*.

De berekeningen werden uitgevoerd voor $\alpha = 0{,}05$ en $\alpha = 0{,}025$ bij éénzijdige toetsing (voor tweezijdige toetsing komt dit dus overeen met $\alpha = 0{,}10$ en $\alpha = 0{,}05$).

De sommatie over m (zie (13)) is niet uitgestrekt tot hele grote en kleine waarden van m, maar zodanig afgebroken dat de som van $\mathrm{P}\,[\underline{m} = m]$ voor alle niet gebruikte waarden van m kleiner dan 0,015 is. Het gevonden onderscheidingsvermogen kan dus hoogstens 0,015 kleiner dan de werkelijke waarde zijn ($\beta_m < 1$ voor alle m), afgezien natuurlijk van fouten, die het gevolg zijn van de benadering van een binomiale verdeling door een Poisson-verdeling, die we in het begin van § 4 uitvoerden.

Opmerking.

Een gebruikelijke definitie van het onderscheidingsvermogen luidt: β is de kans dat H_0 verworpen wordt, indien een gegeven alternatieve hypothese juist is. Dat wil dus zeggen: de kans dat de toetsingsgrootheid in het kritieke gebied valt, berekend onder de alternatieve hypothese. Bij tweezijdige toetsing zou dan ook dat deel van de kritieke zone, dat tot een verkeerde conclusie omtrent de alternatieve hypothese zou leiden, een bijdrage tot het onderscheidingsvermogen leveren. Liever definiëren we dan ook β als de kans om H_0 te verwerpen ten gunste van een juiste beslissing.