STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

S 9
(M 2)

Hemelrijk J, Vaart H R van der

Het gebruik van een- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen.

1950

Overdruk uit Statistica, Jaargang 4, no. 1-2, 1950

Het gebruik van een- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen *)

door J. Hemelrijk en H. R. van der Vaart
Mathematisch Centrum, Amsterdam

Summary

The use of unilateral and bilateral critical regions in the testing of hypotheses.

This paper endeavours to explain in simple terms the principles of the Neyman-Pearson theory.

Let H_0 be the hypothesis to be tested. Then the observations available for testing H_0 are first condensed into a single statistic, x, the distribution of which can be evaluated when H_0 is true. Out of the possible range of values of this statistic a critical region is selected, and H_0 is rejected when x falls in this region, and not rejected when x falls outside. This critical region is chosen so that

(A) *the probability of rejecting H_0 when true has a prescribed upper limit α (or preferably is equal to α);*

(B) *the probability of rejecting H_0 is higher when an alternative hypothesis H_1 is true than when H_0 itself is true; and*

(C) *if possible, the probability of rejecting H_0 is a maximum when any hypothesis h out of a set of alternative hypotheses is true.*

When the set of alternative hypotheses is specified by a single parameter θ, H_0 corresponding to $\theta = \theta_0$, these requirements will, under conditions of a general nature, lead to the use of unilateral or of bilateral tail-errors according to the range of values of θ taken into consideration. If it may be assumed that either $\theta \geqq \theta_0$, or $\theta \leqq \theta_0$, the critical regions must be unilateral, but if θ can be both greater or smaller than θ_0, they have to be bilateral.

The arguments are illustrated by a few simple examples.

§ 1. *Inleiding.*

In de toegepaste statistiek wordt bij het toetsen van hypothesen vaak gebruik gemaakt van het begrip overschrijdingskans. Daarbij bestaat de mogelijkheid deze overschrijdingskans naar één kant (éénzijdige overschrijdingskans) of naar beide kanten van het gemiddelde (tweezijdige overschrijdingskans) te berekenen. Daar de éénzijdige overschrijdingskans vaak tot een resultaat voert, dat men met een tweezijdige niet bereiken kan, is het nuttig, richtlijnen te bezitten, waarmee men kan beslissen, òf, en in welke zin, het ver-

*) Dit artikel geeft tevens de voornaamste inhoud weer van een voordracht gehouden door de heer J. Hemelrijk op 1 April 1950 ter gelegenheid van de eerste vergadering van de 'Mathematisch-Statistische Sectie' van de Vereniging voor Statistiek. Op een toen bij de discussie door de heer Drion gestelde vraag wordt in een naschrift geantwoord.

antwoord is, deze éénzijdige overschrijdingskans te gebruiken. Veelal [1] worden de beweegredenen, die leiden tot het gebruik van éénzijdige overschrijdingskansen slechts terloops vermeld, terwijl een rechtvaardiging soms geheel ontbreekt. Een door J. Neyman en E. S. Pearson [2] ontwikkelde theorie aangaande het toetsen van hypothesen biedt enkele duidelijke criteria (berustende op vergelijking van de waarschijnlijkheidstheoretische overwegingen behorende bij verschillende mogelijke hypothesen), waarmee men in bepaalde gevallen het gebruik van een éénzijdige overschrijdingskans kan rechtvaardigen. De bedoeling van dit stukje is, de principes van deze theorie in eenvoudige vorm uiteen te zetten en aan de hand daarvan enkele veel voorkomende gevallen te bespreken. Daarbij beperken we ons tot het eenvoudigste geval van de theorie van Neyman en Pearson (dat echter de principes van hun theorie volledig recht doet wedervaren) en laten alle verdere theorieën, die in verband staan met het toetsen van hypothesen buiten beschouwing [3]. Wij hopen hiermee de elementaire kanten van de theorie van Neyman en Pearson gemakkelijker toegankelijk te maken voor niet-wiskundigen dan in de oorspronkelijke publicaties, die een meer algemene opzet hebben, het geval is.

§ 2. *De keuze van de hypothese.*

In het volgende zullen wij steeds de te toetsen hypothese met H_0 en de juiste hypothese met H_1 aangeven. Van deze twee is H_1 onbekend (anders is er niets meer te toetsen), terwijl H_0 bekend is; het is echter niet bekend of H_0 juist of onjuist is. Het toetsen van de hypothese H_0 op grond van gegeven waarnemingsmateriaal W kan nu tot tweeërlei resultaat leiden: verwerping van de hypothese (onder een in het volgend punt te behandelen voorbehoud), of niet-verwerping daarvan. Dit laatste houdt echter in het algemeen niet in, dat H_1 de enige hypothese is, die niet verworpen kan worden, zodat men „niet-verwerpen" niet gelijk kan stellen met „aanvaarden". Alleen in zeer bijzondere gevallen, b.v. als er slechts twee elkaar uitsluitende hypothesen mogelijk zijn, zal niet-verwerping van een hypothese gelijkstaan met aanvaarding daarvan. Daar deze situatie zich zelden voordoet, zal men in het algemeen de te toetsen hypothese H_0 zó trachten te kiezen, dat verwerping van H_0 een voor het betreffende onderzoek belangrijkere conclusie is dan niet-verwerping.

§ 3. *Twee soorten fouten; aan de toetsingsmethode te stellen eisen; vorm der conclusies.*

Daar de toetsing van H_0 tot twee resultaten kan leiden, kunnen er ook twee soorten fouten worden gemaakt:

a. De hypothese H_0 kan verworpen worden, terwijl zij juist is (*fout van de eerste soort*). Men kan deze fout niet met zekerheid vermijden (behalve in zeer eenvoudige gevallen, waarin statistische toetsing overbodig is). De theorie van Neyman en Pearson geeft echter aan (zie § 4 en 5) hoe men toetsingsmethoden kan ontwerpen, waarbij de kans op het maken van deze fout bekend is, of tenminste een bovengrens ervan. Deze kans op het maken van een fout van de eerste soort wordt door Neyman en Pearson de ,,*level of significance*'' genoemd, hetgeen man kan vertalen door ,,*onbetrouwbaarheidsdrempel*''*), daar deze kans een maat is voor de onbetrouwbaarheid van de toetsingsmethode. We zullen deze kans aangeven met K_1; precies geformuleerd is K_1 dus de kans, dat de toetsingsmethode bij toetsing van een *juiste* hypothese tot verwerping van deze hypothese zal leiden.

b. De hypothese H_0 wordt niet verworpen, hoewel zij onjuist is (*fout van de tweede soort*). In dit geval is een van H_0 verschillende hypothese juist, maar in het algemeen zullen naast H_0 vele andere hypothesen mogelijk zijn, zodat het onbekend zal zijn, welke hypothese de juiste is. Noemen wij nu K_2 de kans, dat H_0 niet verworpen zal worden, terwijl $H_1 \neq H_0$ de juiste hypothese is, dan is K_2, die we de kans op een fout van de tweede soort kunnen noemen, afhankelijk van H_1, zodat K_2 niet met een enkel getal aangegeven kan worden. We schrijven daarom voor deze kans: $K_2(H_1)$. Men kan nu trachten de toetsingsmethode zo te maken, dat er voor $K_2(H_1)$ een bekende bovengrens < 1 bestaat.

In verband met het volgende is het nog van belang op te merken, dat de kans K_3, dat H_0 verworpen zal worden als $H_1 \neq H_0$ juist is, gelijk is aan $1 - K_2(H_1)$, dus evenals K_2 van de (onbekende) juiste hypothese H_1 afhangt. We schrijven daarom $K_3(H_1)$. Het is duidelijk, dat deze grootheid $K_3(H_1)$, die door Neyman en Pearson de ,,power function'' **) van de toetsingsmethode wordt genoemd, van groot belang is. Immers $K_3(H_1)$ is de kans, dat men, als H_1 juist is, tot de onjuistheid van H_0 zal concluderen.

De grootheden K_1, $K_2(H_1)$ en $K_3(H_1)$ zijn afhankelijk van de te toetsen hypothese en de toetsingsmethode. K_2 en K_3 bovendien van de juiste hypothese H_1. Over de keuze van H_0 is in § 2 reeds één en ander opgemerkt. De toetsingsmethode is een schema voor de statistische bewerking van het waarnemingsmateriaal W. Men zal nu trachten bij het ontwerpen van een toetsingsmethode te voldoen aan de volgende *eisen*:

---

*) De vertalingen zijn ontleend aan de syllabus van de cursus Mathematische Statistiek, door Prof. Dr. D. van Dantzig, Mathematisch Centrum, Amsterdam, 1947—1949.

**) Vert.: ,,onderscheidingsvermogen''.

A. *Men zal trachten de kans K_1 op een fout van de eerste soort, dus op ten onrechte verwerping van H_0, $\leqq \alpha$ (en zo mogelijk $= \alpha$) te maken,* waarbij α een van tevoren bepaald getal tussen 0 en 1 is. Gewoonlijk neemt men voor α de waarde 0,05 of 0,01. De belangrijkste factor bij de keuze van α wordt gewoonlijk gevormd door het overwegen van de gevolgen, die een fout van de eerste soort zou hebben. Tevens heeft men rekening te houden met het feit, dat men in het algemeen α slechts kan verkleinen ten koste van een vergroting van $K_2(H_1)$, de kans op een fout van de tweede soort.

B. *Men zal trachten de kans $K_3(H_1)$, dat H_0 verworpen zal worden, als $H_1 \neq H_0$ juist is, $> \alpha$ te maken.* Dit is een redelijke eis: men zal niet graag een toetsingsmethode gebruiken, waarbij de kans op terecht verwerpen van H_0 kleiner is dan die op ten onrechte verwerpen. Daar H_1 onbekend is houdt deze eis in, dat $K_3(h) > \alpha$ is *voor iedere mogelijke hypothese h met $h \neq H_0$*. Is aan deze eis voldaan, dan noemen Neyman en Pearson de methode ,,unbiased''. *)

In § 5 zal blijken, dat $K_3(H_1)$ tevens de kans is op het ten onrechte verwerpen van hypothese H_1, indien deze hypothese volgens de zelfde methode wordt getoetst. Is dus $K_3(h) > \alpha$ voor iedere $h \neq H_0$, dan zijn we zeker, dat de kans op ten onrechte verwerpen het kleinst is bij de hypothese H_0, zodat het gerechtvaardigd, is, van alle hypothesen juist H_0 in de eerste plaats te verwerpen.

C. *Men zal trachten $K_3(H_1)$ zo groot mogelijk te maken, liefst door $K_3(h)$ maximaal te maken voor iedere $h \neq H_0$ apart.* **)

Indien men geen kans ziet hieraan te voldoen, is het toch in ieder geval duidelijk, dat een toetsingsmethode T_1 beter is dan T_2, als voor beide voldaan is aan eis A (met zelfde α) en B en als $K_3(h)$ bij T_2 voor iedere $h \neq H_0$ kleiner is dan bij T_1.

Met behulp van deze drie eisen kan men nu vaak besluiten, hoe men de statistische bewerking van het materiaal zal uitvoeren. *In het bijzonder is de keuze tussen gebruik van één- en tweezijdige overschrijdingskans in veel gevallen op deze wijze te verantwoorden.*

De *conclusie*, die men tenslotte trekt, kan dus tweeërlei vorm hebben:

I. Op grond van het waarnemingsmateriaal W wordt de hypothese H_0 verworpen; de kans, dat deze conclusie fout is, is $= \alpha$ (resp. $\leqq \alpha$).

II. Het waarnemingsmateriaal W geeft geen voldoende aanleiding tot verwerping van H_0; de kans, dat H_0 toch onjuist is, is onbekend, maar bij

*) Vert.: ,,zuiver''.

**) In dat geval heet de toetsingsmethode ,,uniformly most powerful'' (Vert.: gelijkmatig machtigst). Uiteraard wordt steeds onder ,,iedere van H_0 verschillende hypothese'' verstaan: iedere voor het probleem relevante en van H_0 verschillende hypothese.

de gekozen α zo klein mogelijk gemaakt. (Bovendien is deze kans $< 1 - \alpha$ maar dat zegt niet veel, daar α meestal klein is).

De volgende *opmerking* is nog van belang: Ook als K_1 en $K_2(H_1)$ beide bekend zijn (b.v. als er slechts twee hypothesen mogelijk zijn) kan men ze niet optellen om de totale kans op een ,,fout van de eerste of tweede soort'' te berekenen, noch in het algemeen deze totale kans op andere wijze berekenen. De reden hiervan is, dat K_1 en $K_2(H_1)$ berekend worden onder *verschillende* onderstellingen, nl. dat H_0 resp. dat H_1 de juiste hypothese is.

§ 4. *Eerste stap der statistische bewerking.*

Onder aanname van de hypothese H_0 reduceert men het waarnemingsmateriaal W tot één enkele waarneming van een variabele x met een bekende verdeling, waarvan de waarschijnlijkheidsdichtheid (frequentiekromme) gegeven wordt door de vergelijking

$$y = f_{H_0}(x)$$

Hierin kan x b.v. het gemiddelde van een aantel waarnemingen zijn of een andere uit de waarnemingen berekenbare grootheid, waarvan de waarschijnlijkheidsverdeling berekend kan worden, als men de hypothese H_0 als juist aanneemt (vgl. ook de onderstaande voorbeelden). Dit zal niet altijd mogelijk zijn, maar we beperken ons tot die klasse van gevallen, waarin dit wel zo is Het gunstigste geval is dat, waarin we dit voor *iedere* mogelijke hypothese kunnen bereiken, zodat we H_0 uit de verzameling van mogelijke hypothesen naar willekeur kunnen kiezen; ook als dit niet, of veel moeilijker gaat, kan men de methode soms toepassen, maar men wordt dan in zijn keuze van H_0 door deze omstandigheden belemmerd, daar men slechts die H_0's kan toetsen, waarvoor $f_{H_0}(x)$ (numeriek of expliciet) berekend kan worden.

Ter illustratie geven we *twee voorbeelden*:

a. Men werpt N maal kruis-of-munt met een geldstuk, met het doel na te gaan, of de munt ,,onzuiver'' is, en verkrijgt daarbij n maal kruis en $N—n$ maal munt (dit is dus het waarnemingsmateriaal W). W kan beschouwd worden als een serie van N onafhankelijke trekkingen met teruglegging uit een collectie C, waarbij constante kansen p (op kruis) en $1—p$ (op munt) bestaan. Men beschouwt nu, ter uitvoering van bovengenoemde reductie van W tot één waarneming, n als variabele en berekent de kans, die op iedere waarde van n tussen 0 en N bestaat (dit is $f_p(n)$, met n in plaats van x en p in plaats van H_0). W vertegenwoordigt dan één enkele waarneming van deze n; de verdeling van n (een Bernoulli-verdeling) bevat p nog als parameter. Als hypothese H_0 kunnen we nu nemen $p = \frac{1}{2}$; wordt deze verworpen, dan concludeert men dus tot onzuiverheid van de munt.

Opmerking: we verkeren hier in het gunstige geval, dat voor iedere mogelijke hypothese omtrent p de verdeling van n bekend is. De wijze, waarop men tot verwerping kan komen (of tot de conclusie, dat W geen aanleiding tot verwerping van H_0 geeft) wordt in het volgende punt behandeld.

b. Een experiment E, dat een succes of een mislukking kan opleveren, wordt volgens twee methoden M_1 en M_2 uitgevoerd en wel N_1 maal volgens M_1 en N_2 maal volgens M_2. Hierbij worden respectievelijk n_1 en n_2 successen verkregen. Alle $N_1 + N_2$ proeven zijn onafhankelijk van elkaar. Wat valt er statistisch te concluderen aangaande de verhouding van de kansen p_1 resp. p_2 op succes als men M_1 resp. M_2 toepast (m.a.w.: is één van beide methoden beter dan de andere)? We beschouwen nu $v = p_1/p_2$ als een onbekende parameter, die dus iedere positieve waarde kan aannemen (inclusief 0 en oneindig). W bestaat nu uit twee series trekkingen met teruglegging uit collecties C_1 en C_2 met kansen p_1 resp. p_2 op succes en daarbij zijn in totaal $n = n_1 + n_2$ successen opgetreden.
Nemen we nu als hypothese H_0: $v = 1$, dan kan men gemakkelijk voor iedere m_1 tussen 0 en n de kans uitrekenen, dat van deze n successen er m_1 bij methode M_1 en $n - m_1$ bij M_2 zullen optreden. De verdeling van m_1 is dus onder hypothese H_0 bekend en de gevonden n_1 vertegenwoordigt één waarneming van m_1; op deze wijze is het waarnemingsmateriaal weer tot één waarneming van een variabele met bekende verdeling gereduceerd.

Opmerking: in dit geval is de verdeling van m_1 alleen voor $v = 1$ gemakkelijk te berekenen; alleen in dat geval nl. valt de p_1 (die dan $= p_2$ is) bij de berekening weg. Zou men dus een andere waarde voor v als hypothese willen nemen, dan blijft er een parameter in de verdeling van m_1 over, hetgeen de zaak compliceert. We zullen ons hier beperken tot de bespreking van de toetsing van de hypothese $v = 1$. *)

§ 5. *Verdere statistische bewerking; het trekken van een conclusie met behulp van een kritieke zône.*

We onderstellen de bovenstaande reductie van W uitgevoerd, zodat W teruggebracht is tot één waarneming x_0 van een zekere variabele x met, onder hypothese H_0, als waarschijnlijkheidsdichtheid $y = f_{H_0}(x)$ (zie fig. 1). We zoeken nu een gedeelte Z van de x-as uit en spreken af, H_0 dan en slechts dan te verwerpen, als de door W geleverde waarde x_0 van de variabele x in Z ligt. Z heet de *kritieke zône* en moet zó uitgezocht worden, dat aan de in § 3 gestelde eisen A, B, en C voldaan is (voorzover dat mogelijk is). We zullen ons, wat de vorm van Z betreft, beperken tot intervallen en paren van intervallen

*) Problemen van deze aard kan men ook op andere wijze aanpakken. Vgl. [4].

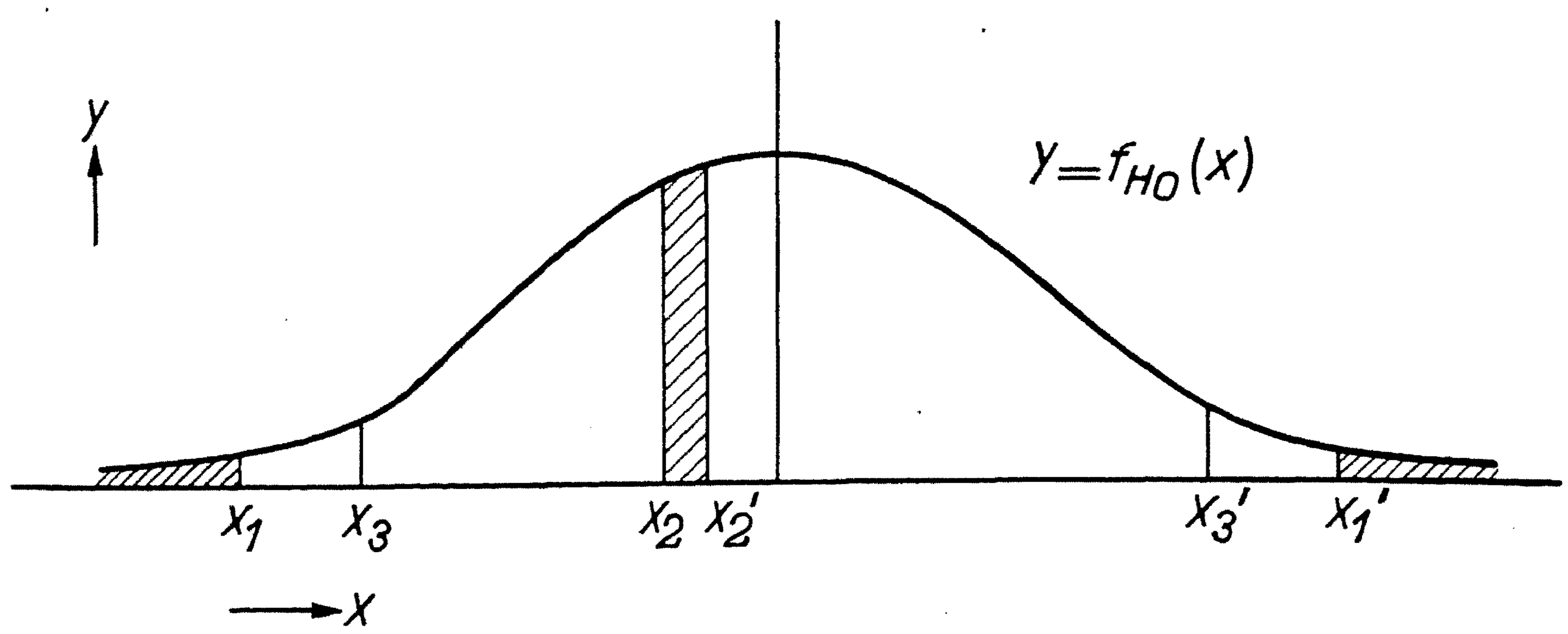

Fig. 1 *)

(waarbij een interval ook oneindige lengte mag hebben). Met de term „*éénzijdige overschrijdingskans*" wordt nu bedoeld het gebruik van een kritieke zône, bestaande uit één interval (eventueel van oneindige lengte) $x \geqq x_3'$ (of $x \leqq x_3$; in beide gevallen zullen we voor het gemak ook van een *éénzijdige kritieke zône* spreken, te onderscheiden in linkse- en rechtse-) en met „*tweezijdige overschrijdingskans*" het gebruik van een uit twee dergelijke intervallen bestaande kritieke zône (dus b.v. bestaande uit de twee intervallen $x \leqq x_1$ en $x \geqq x_1'$; *tweezijdige kritieke zône*). Verder kan men natuurlijk een interval $x_2 \leqq x \leqq x_2'$ als Z nemen, maar dit zal in het algemeen niet aan de gestelde eisen voldoen. We zullen nu nagaan, welke invloed de 3 eisen op de keuze van Z uitoefenen.

Eis A: De kans op ten onrechte verwerpen van H_0 moet $= \alpha$ zijn. Hieraan voldoet iedere Z, waarvoor geldt:

$$\int_Z f_{H_0}(x)\, dx = \alpha; \tag{1}$$

immers het linkerlid, dat gelijk is aan het oppervlak van het tussen Z en het daarbovengelegen deel van de kromme $y = f_{H_0}(x)$, geeft de kans aan, dat x in Z ligt, als H_0 juist is. We kunnen hier dus inderdaad K_1 gelijk aan α maken, daar aan (1) op zeer veel verschillende wijzen voldaan kan worden; o.a. door éénzijdige en tweezijdige kritieke zônes en ook door zônes van het type $x_2 \leqq x \leqq x_2'$. We zullen de integraal in het linkerlid van (1) aangeven met $I_{H_0}(Z)$, zodat aan eis A voldaan is, als

$$K_1 = I_{H_0}(Z) = \alpha \tag{2}$$

is.

*) Voor het gemak nemen we in de tekening een continue en symmetrische verdeling. In werkelijkheid hoeft geen van beide het geval te zijn.

Opmerking: indien de verdeling niet continu is, maar discreet, kan men α niet willekeurig kiezen, maar wel bij iedere Z precies bepalen.

Eis B: de kans $K_3(H_1)$, dat H_0 verworpen wordt, als H_1 juist is, moet $> \alpha$ zijn. Geven we een van H_0 verschillende hypothese aan met h, en in het bijzonder de juiste hypothese met H_1, dan is eis B:

$$K_3\,(H_1) = \int_z f_{H_1}(x)\,dx = I_{H_1}(Z) > \alpha. \tag{3}$$

Daar we niet weten, welke van H_0 verschillende hypothese de juiste is, moeten we eisen, dat geldt:

$$I_h\,(Z) > \alpha \text{ voor iedere } h \neq H_0\ ^{*}). \tag{4}$$

Op dit punt wordt het, daar nu alle van H_0 verschillende hypothesen h in de redenering betrokken zijn, noodzakelijk op een speciaal geval over te gaan, wat de vorm van f_h voor de verschillende h's betreft. We zullen ons verder beperken tot het veel voorkomende geval, waarbij de verschillende hypothesen, zoals bij de genoemde voorbeelden het geval was, aangeduid kunnen worden door aan een parameter, die we, terwille van de algemeenheid, θ zullen noemen, verschillende waarden toe te kennen. De hypothese H_0 zij $\theta = \theta_0$ en de bij een willekeurige waarde θ behorende waarschijnlijkheidsdichtheid zij $y = f_\theta\,(x)$. Verder onderstellen we, dat voor iedere $\theta < \theta_0$ de bijbehorende $f_\theta(x)$ voor kleine x boven $f_{\theta_0}\,(x)$ verloopt en er één snijpunt mee heeft en voor $\theta > \theta_0$ voor grote x boven $f_{\theta_0}\,(x)$ verloopt er en één snijpunt mee heeft (zoals in fig. 2 aangegeven).

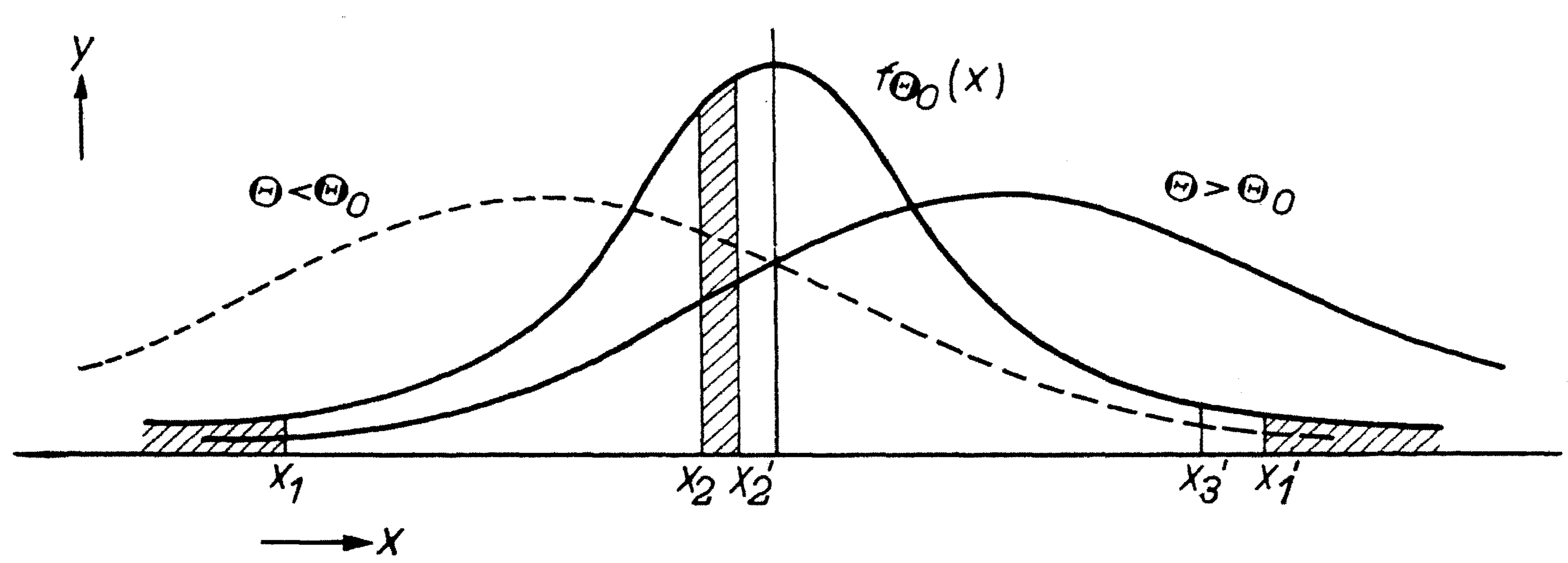

Fig. 2

Aan dit schema voldoen onze beide voorbeelden (met slechts het, voor het betoog niet wezenlijke, verschil, dat de verdelingen in die gevallen discreet zijn, terwijl x bovendien slechts eindig veel waarden aan kan nemen); voor

*) $I_h(Z)$ opgevat als functie van h moet dus voor $h = H_0$ minimaal zijn.

het eerste voorbeeld is gemakkelijk in te zien, dat men een dergelijk verloop van de verdelingen krijgt (θ is daar p, $\theta_0 = \frac{1}{2}$ en x is n); bij het tweede (met $\theta = v$ en x is m_1) is alleen de $f_{\theta_0}(x)$ met $\theta_0 = 1$ bekend, maar men kan toch gemakkelijk inzien, dat voor $v < 1$ de kleine waarden van m_1 en voor $v > 1$ de grote waarden van m_1 grotere waarschijnlijkheden bezitten dan bij $v = 1$, zodat het door fig. 2 gegeven schema ook voor het tweede voorbeeld geldt.

We zien nu direct, dat, indien zowel $\theta < \theta_0$ als $\theta > \theta_0$ mogelijke hypothesen zijn, alleen een tweezijdige kritieke zône, zoals die bestaande uit de intervallen $x \leqq x_1$ en $x \geqq x_1'$ aan eis B kan voldoen. Voor iedere andere Z zijn er θ's te vinden met

$$I_\theta(Z) < I_{\theta_0}(Z) = \alpha.$$

In het algemeen, nl. als er geen nadere gegevens omtrent θ zijn, zal men dus in een dergelijk geval van een *tweezijdige kritieke zône* gebruik moeten maken, indien men tenminste niet met eis B in botsing wil komen.

Indien echter *van tevoren gegeven* is, dat b.v. $\theta < \theta_0$ niet kan voorkomen, zodat in fig. 2 krommen van het type van de gestippelde niet voorkomen, is het duidelijk, dat ook een éénzijdige Z, en wel de rechtse (dus $x \geqq x_3'$ met $I_{H_0} = \alpha$) aan eis B voldoet. Ook nu echter kan een tweezijdige Z aan eis B voldoen. Dan beslist de derde eis.

Eis C: van twee toetsingsmethoden is die de beste, waarbij $K_3(h)$ voor iedere van H_0 verschillende hypothese h het grootst is. Men ziet echter direct, dat voor iedere $\theta > \theta_0$ geldt, dat I_h voor de éénzijdige kritieke zône $x \geqq x_3'$ groter is dan voor iedere tweezijdige (met gelijke $\alpha = I_h$). In dit geval is dus, daar $\theta < \theta_0$ niet kan voorkomen, het gebruik van een *éénzijdige kritieke zône* boven dat van een tweezijdige te verkiezen. Er dient echter met nadruk op gewezen te worden, dat tot de onmogelijkheid van $\theta < \theta_0$ *niet* besloten mag worden uit het waarnemingsmateriaal W (tenzij door een statistisch onderzoek met behulp van tweezijdige overschrijdingskansen), maar dat dit een gegeven moet zijn, dat van tevoren op andere gronden reeds vaststond. Algemeen: *indien men de keuze der kritieke zône van het waarnemingsmateriaal laat afhangen, is de kans op het ten onrechte verwerpen van de hypothese niet noodzakelijk meer* $\leqq \alpha$.

§ 6. *Bijzonder geval van gebruik van een éénzijdige kritieke zône.*

Een belangrijk geval, waarbij men, hoewel $\theta < \theta_0$ wèl mogelijk is, toch met recht een eenzijdige kritieke zône kan gebruiken, treedt op, indien men in het in de vorige paragraaf beschreven geval verkeert en als men van te voren weet dat de op grond van de waarde van θ te nemen beslissing voor $\theta < \theta_0$ en $\theta = \theta_0$ dezelfde is, maar voor $\theta > \theta_0$ een andere, dus indien het er slechts om gaat uit te maken of $\theta > \theta_0$ is of niet. Het gebruik van een éénzijdige kritieke zône is dan slechts schijnbaar in strijd met bovenstaande redenering. Immers de hypothesen, waar het nu om gaat zijn van een ander type, dan de vroeger besprokene, nl.

1e $\theta \leqq \theta_0$; dit is de eventueel te verwerpen H_0,
2e $\theta > \theta_0$; dit is de alternatieve hypothese h.

De zaak aldus stellende hebben we slechts twee hypothesen, maar die van ingewikkelder aard zijn, dan de in de vorige punten besprokene. We kunnen echter op deze nieuwe hypothesen in dit speciale geval met succes de criteria van punt 3 voor het opsporen van de gunstigste kritieke zône toepassen:

Is H_0 juist, dan is $\theta \leqq \theta_0$, maar overigens onbekend.
Eis A houdt dus in, dat Z voldoet aan

$$I_\theta(Z) \leqq \alpha \text{ voor iedere } \theta \leqq \theta_0 \text{ *)}, \tag{5}$$

Eis B houdt in, dat Z voldoet aan

$$I_\theta(Z) > \alpha \text{ voor iedere } \theta > \theta_0. \tag{6}$$

Tenslotte wijst eis C die Z als de beste aan, waarvoor het verschil $I_{\theta_1}(Z) - I_{\theta_2}(Z)$ voor iedere θ_1 en θ_2 met $\theta_1 > \theta_0$ en $\theta_2 \leqq \theta_0$, zo groot mogelijk is. Aan de hand van figuur 2 ziet men gemakkelijk in, dat de rechtse kritieke zône $x > x_3'$ aan eis A en B voldoet en door eis C als de beste aangewezen wordt.

De plaats van x_3' wordt dan bepaald door de vergelijking

$$I_{\theta_0}(Z) = \alpha \text{ dus } \int_{x_3}^{\infty} f_{\theta_0}(x)\,dx = \alpha, \tag{7}$$

daar van alle $f_\theta(x)$ met $\theta \leqq \theta_0$ voor positieve x $f_{\theta_0}(x)$ de grootste is, zodat voor iedere rechtse kritieke zône, die slechts positieve x-waarden bevat, van alle $I_\theta(Z)$ met $\theta \leqq \theta_0$, $I_{\theta_0}(Z)$ de grootste is. Is dus aan (7) voldaan, dan ook aan (5), terwijl omgekeerd aan (7) voldaan moet zijn, indien aan (5) voldaan is.

Een dergelijk geval als hier beschreven kan zich b.v. voordoen, indien in het tweede voorbeeld van § 4 M_1 en M_2 twee verschillende methoden (voor genezing van een ziekte of voor fabricage van een product) voorstellen, waarvan M_2, indien men niet voor zo goed als zeker kan aannemen, dat M_1 beter is (dus $p_1 > p_2$ is; $v > 1$), boven M_1 geprefereerd wordt (b.v. om financiële redenen). Immers dan is de te toetsen hypothese $H_0 : M_2$ is minstens even goed als M_1 ($v \leqq 1$) en slechts als deze verworpen kan worden kiest men methode

*) In dit geval kunnen we voor K_1 slechts een bovengrens aangeven.

M_1. De kans, dat men methode M_1 ten onrechte zal kiezen is dan $\leqq \alpha$ (waarbij men α kiezen kan), terwijl de kans, dat men ten onrechte M_2 zal kiezen bij de gekozen α zo klein mogelijk is gemaakt (en dus bij het gebruik van éénzijdige kritieke zône kleiner is dan bij het gebruik van tweezijdige).

§ 7. *Samenvatting.*

Samenvattend kan men zeggen, dat, indien $I_\theta(Z)$ voor stijgende θ een monotoon dalende functie is als Z een linkerkritieke-zône is, en een monotoon stijgende als Z een rechter-kritieke zône is, dat dan in het algemeen voor de toetsing van de hypothese $\theta = \theta_0$ een tweezijdige kritieke zône gebruikt moet worden om aan de eisen A en B van § 3 te voldoen, terwijl een éénzijdige kritieke zône aan de eisen A, B en C voldoet, indien men van tevoren weet, dat $\theta < \theta_0$ (of $\theta > \theta_0$) onmogelijk is. Ook indien men de hypothese $\theta \leqq \theta_0$ (of $\theta \geqq \theta_0$) wenst te toetsen, voldoet, indien de genoemde monotonie-voorwaarde vervuld is, slechts een éénzijdige kritieke zône aan de eisen A, B en C.

§ 8. *Voorbeeld.*

Een zeer eenvoudig *voorbeeld*, waarin men ook zonder bovenstaande redenering wel reeds begrijpt, dat men beter een éénzijdige dan een tweezijdige kritieke zône kan gebruiken, is het door fig. 3 weergegevenen [5].

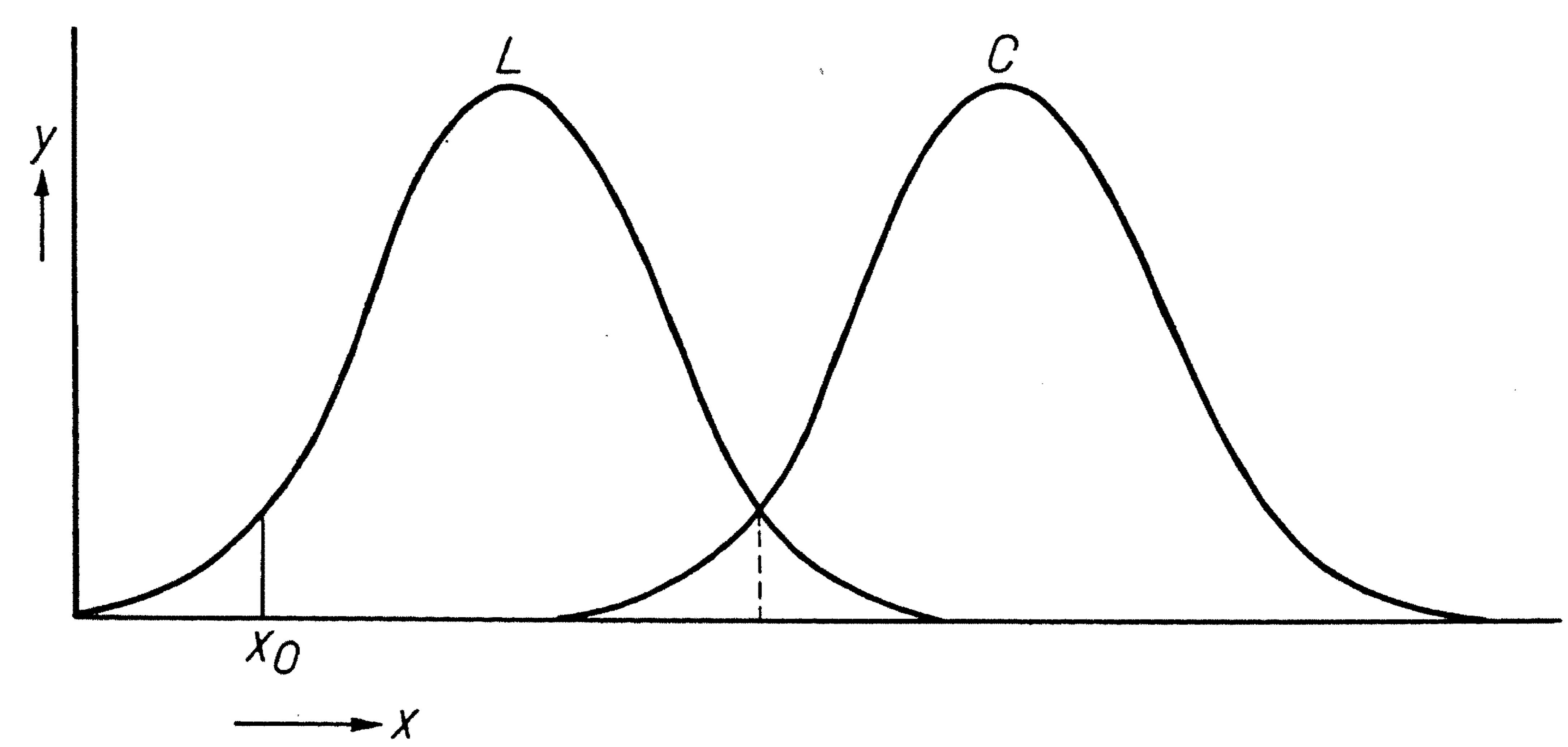

Fig. 3

De frequentieverdeling van de verhouding van kopbreedte tot lengte van het derde antennesegment is bij Cimex lectularius als aangegeven door de kromme L, bij Cimex columbarius als aangegeven door de kromme C. Heeft men nu van een exemplaar van Cimex, waarvan men weet, dat het òf lectularius òf columbarius is, bovengenoemde maat gemeten en wil men de hypothese toetsen, dat het Cimex lectularius is, dan zal men, als x_0 de gemeten waarde is, er wel niet aan denken deze hypothese te verwerpen. Hiermee is in overeenstemming, dat men volgens het bovenstaande de kritieke zône éénzijdig en wel rechts zal moeten kiezen, daar dan de kans, dat x in Z ligt onder de hypothese L gelijk aan $I_L(Z)$ (dus bij geschikte keuze van het linkereindpunt van Z gelijk aan α) is, onder de hypothese C echter de maximale waarde heeft (maximaal, genomen over alle mogelijke Z's met dezelfde α).

Opmerking: als men de hypothese C wil toetsen moet men op analoge gronden de linker-kritieke zône van C gebruiken. De kritieke zônes, behorende bij eenzelfde α, voor hypothesen L resp. C mogen echter in geen geval over elkaar heen vallen, daar men dan, als x in het gemeenschappelijke deel van deze zônes zou vallen, beide hypothesen zou moeten verwerpen. Hierdoor wordt dus in dit geval een grens aan de keuze van α gesteld; kiest men α te groot, dan vallen de bij de hypothesen L en C behorende kritieke zônes over elkaar.

LITTERATUUR:

[1] Zie b.v.; Statistical Research Group, Columbia University, Selected techniques of Statistical Analysis, McGraw Hill, New York 1947, p. 249, 251 en 253.
F. Yates, Contingency tables involving small numbers and the χ^2-test, Suppl. J. roy. stat. Soc. **1** (1934), 231.
M. G. Kendall, The advanced theory of statistics, Griffin and Co., London 1946, Vol. I, p. 304.
S. T. Bok, De gedachtengang van de statistica, Stenfert Kroese, Leiden 1946, p. 17 en 18.

[2] Zie b.v.: J. Neyman, Basic Ideas and some recent results of the theory of testing statistical hypothese, J. roy. stat. Soc. **105** (1942), 292-327, met uitgebreide literatuurlijst.
J. Neyman and E. S. Pearson, Statistical Research Memoirs, I en II, London 1936 en 1938.

[3] Zie b.v.: A. Wald, On the principles of statistical inference, Notre Dame, Indiana 1942

[4] G. A. Barnard, Significance tests for 2 × 2 tables, Biometrika **34** (1947), 123-138.
E. S. Pearson, The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table, Biometrika **34** (1947), 139-167.

[5] C. G. Johnson, Taxonomic characters, variability and relative growth in Cimex lectularius L and Cimex columbarius Jenyns (Heteropt, cimicidae), Trans. roy. entom. Soc. London **89** (1939) 543-568.

Naschrift.

Tijdens de vergadering op 1 April werd bij de discussie de vraag gesteld, of bij de toetsing van de hypothese $p = \frac{1}{2}$ (bij het eerste voorbeeld dus), indien men niet alleen weet, dat $p < \frac{1}{2}$ onmogelijk is, maar ook, dat p zeker niet groter is dan b.v. $^3/_4$, een éénzijdige (rechter-) kritieke zône nog wel aan eis C voldoet. De mogelijkheid werd geopperd, dat in dit geval deze éénzijdige kritieke zône misschien vervangen zou moeten worden door een iets meer naar het midden gelegen interval, van de vorm $x_2 \leqq x \leqq x_2'$. Deze zou dus uit de éénzijdige kritieke zône ontstaan door zoveel waarden aan de uiterste rechterzijde weg te laten, dat er aan de linkerzijde één bij genomen kan worden zonder α te vergroten.

Deze vraag moet, zoals uit de volgende berekening blijkt, ontkennend worden beantwoord. Wij bewijzen nl., dat voor de toetsing van de hypothese $p = \frac{1}{2}$ tegen een hypothese $p = p_0 > \frac{1}{2}$, welke waarde p_0 overigens ook bezit, steeds de rechter-éénzijdige kritieke zône de enige is, die aan eis C voldoet. Dit geldt derhalve ook, als $p = \frac{1}{2}$ getoetst wordt tegen een verzameling van waarden, die groter dan $\frac{1}{2}$ zijn, maar die overigens willekeurig mag zijn.

Het bewijs verloopt als volgt:

De kans, P_x, op x maal kruis bij n worpen is

$$P_x = \binom{n}{x} p^x q^{n-x} \quad \text{met } q = 1 - p.$$

Hieruit volgt:

$$P_x' = \frac{dP_x}{dp} = \binom{n}{x} p^{x-1} q^{n-x-1} (x - np)$$

dus

$$\frac{P_x'}{P_x} = \frac{x - np}{pq}. \tag{I}$$

Voor $x \geqq np$ is dit een monotoon stijgende en positieve functie van x, die zijn maximum bereikt voor $x = n$. De kritieke zône moet zeker geheel rechts van het gemiddelde np liggen. Vervangen wij nu de waarden $x = n$, $x = n - 1, \ldots, x = n - m$ van de éénzijdige kritieke zône door één of meer waarden $x = k$, $x = k + 1, \ldots x = k + h$, die links van de éénzijdige kritieke zône liggen, dan moet, voor $p = \frac{1}{2}$, voldaan zijn aan:

$$\sum_{\nu = n-m}^{n} P_\nu \geqq \sum_{\mu = k}^{k+h} P_\mu \tag{II}$$

daar anders α zou worden vergroot.

Volgens (I) is nu

$$\frac{P_\nu'}{P_\nu} > \frac{P_\mu'}{P_\mu} \quad \text{voor } \nu = n - m, \ldots, n; \ \mu = k, \ldots, k + h.$$

Dus is ook

$$\frac{\sum\limits_{\nu=n-m}^{n} P_\nu'}{\sum\limits_{\nu=n-m}^{n} P_\nu} > \frac{\sum\limits_{\mu=k}^{k+h} P_\mu'}{\sum\limits_{\mu=k}^{k+h} P_\mu}$$

zodat tevens geldt

$$\sum_{\nu=n-m}^{n} P_\nu' > \sum_{\mu=k}^{k+h} P_\mu'.$$

Daar voor $p = \frac{1}{2}$ (II) geldt, geldt (II) voor iedere $p > \frac{1}{2}$ met het groter-teken, waarmee het gestelde bewezen is.

J. H.