

STICHTING
MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

A M S T E R D A M

STATISTISCHE AFDELING

LEIDING: PROF. DR D. VAN DANTZIG

ADVISEUR VOOR STATISTISCHE CONSULTATIE: PROF. DR J. HEMELRIJK

Rapport S 258 (C 12)

Syllabus van de cursus

"Enkele elementaire statistische methoden"

door

J. Fabius

Bewerking van het rapport S 120 (C 5)

"Toegepaste Statistiek" door

Prof. Dr J. Hemelrijk en Ph. van Elteren

September 1959

The Mathematical Centre at Amsterdam, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

Hoofdstuk 1.

Inleiding.

1.1. Enige statistische problemen.

Het prototype, of, zoals men meestal zegt, het model, van een grote groep statistische problemen is het volgende: Men heeft een vaas, waarin zich een groot aantal rode en witte knikkers bevinden. Men wil een schatting hebben van het percentage rode knikkers in de vaas.

Een iets andere vorm van ditzelfde probleem hebben we, als men om een of andere reden vermoedt, dat de fractie rode knikkers een bepaalde waarde p bezit, en wenst na te gaan of dit vermoeden juist is.

Voorbeelden:

- a. Een partij goederen bevat een onbekende fractie ondeugdelijke exemplaren. Hoe groot is deze fractie?
- b. Een deel der rijksambtenaren bereikt de pensioengerechtigde leeftijd, de anderen overlijden voordien. Hoe groot is de laatste fractie?
- c. Als men ratten een bepaalde dosis van een zekere stof inspuit, ondervinden sommige ratten schadelijke gevolgen, andere niet. Hoe groot is het percentage dat wel schadelijke gevolgen ondervindt van de injectie?
- d. Op grond van de erfelijkheidswetten van Mendel mag men verwachten dat een fractie p van alle planten die door kruising van 2 planten verkregen kunnen worden, een bepaalde eigenschap bezit. Hoe kan men de juistheid van dit vermoeden controleren?
- e. Als men personen eerst in een goed verlichte kamer, en daarna in een donkere kamer het punt op de muur laat aangeven dat zij recht voor zich zien, dan ligt bij sommige personen het tweede punt rechts, bij anderen links van het eerste punt. Hoe groot is het percentage mensen waarbij het tweede punt rechts van het eerste ligt?

Vele andere voorbeelden van problemen van statistische aard zijn te vinden in het boek : J. HEMELRIJK en D. WABEKE,

Elementaire statistische opgaven met uitgewerkte oplossingen, Centrumreeks, deel 2. (J. Noorduijn en En, Gorinchem, 1957). Ook als oefenstof zijn deze vraagstukken zeer aan te bevelen.

1.2. Terminologie.

In al deze voorbeelden gaat het erom een uitspraak te doen over een groep van objecten. Deze groep van objecten noemt men de populatie die beschouwd wordt. Zo is de populatie in ons model de vaas met knikkers, in voorbeeld a de partij goederen, in voorbeeld b de verzameling van alle rijksambtenaren, in c de verzameling van alle ratten van een bepaalde soort, in d de verzameling van alle planten welke door kruising uit de 2 planten verkregen kunnen worden, en in e de verzameling van alle mensen.

Het aantal objecten waaruit de populatie bestaat heet de omvang van de populatie. De omvang kan onbepaald zijn. Dit is bijv. het geval bij de populaties uit de voorbeelden b, c, d en e.

In het geval van de vaas met knikkers, en ook in voorbeeld a zou men alle rode en witte knikkers uit de vaas, resp. alle deugdelijke en ondeugdelijke exemplaren uit de partij kunnen tellen. We spreken dan van een volledige waarneming. Als deze aantallen echter zeer groot zijn kan dit bezwaarlijk zijn, en in voorbeeld a kan het zijn, dat het middel dan erger is dan de kwaal, nl. als we een destructieve keuringsmethode moeten gebruiken om uit te maken of een exemplaar al dan niet deugdelijk is.

In de andere voorbeelden is een volledige waarneming zelfs in het algemeen onmogelijk!

Het essentiële van de wiskundige statistiek is nu, dat deze ons methoden geeft, die ons in staat stellen uit sommige onvolledige waarnemingen, dat wil zeggen waarnemingen die slechts een (vaak onbekend) deel van de populatie omvatten, conclusies te trekken over de hele populatie.

1.3. Het nemen van een steekproef.

In 1.2. hebben we gewezen op het belang van onvolledige waarnemingen in de statistiek. We zullen daarom dit begrip nader preciseren en voorwaarden opstellen waaraan de onvol-

ledige waarneming moet voldoen opdat we de statistische methode die in het volgende behandeld wordt er op mogen toepassen.

Om een onvolledige waarneming te kunnen doen moeten we eerst een aantal objecten uit de populatie afzonderen waaraan we de waarnemingen gaan verrichten. In het geval van de vaas met knikkers moeten we een aantal knikkers eruit pakken, in voorbeeld a moeten we in gedachten een aantal rijksambtenaren afzonderen etc. etc.

We spreken in dit verband van het nemen van een steekproef uit de populatie. Een steekproef kan bestaan uit een willekeurig aantal n objecten. We noemen dit aantal de omvang van de steekproef.

Een steekproef van de omvang n kan men op 2 verschillende manieren verkrijgen:

- a. Men trekt telkens één object uit de populatie, doet aan dit ene project de waarneming, en voegt het weer bij de populatie. Eénzelfde object kan op deze wijze méér dan eens in de steekproef voorkomen. We spreken van een steekproef met teruglegging.
- b. Men trekt n objecten (al of niet tegelijkertijd) uit de populatie, zonder getrokken objecten weer aan de populatie toe te voegen. Nu spreekt men van een steekproef zonder teruglegging.

Voor de theorie maakt het een groot verschil of men steekproeven met, dan wel zonder teruglegging neemt, tenminste als de omvang der populatie eindig is. Als echter de omvang van de populatie zeer groot is ten opzichte van de omvang van de steekproef wordt het verschil verwaarloosbaar. Aangezien de theorie het gemakkelijkst is voor steekproeven met teruglegging, zullen wij ons in het volgende beperken tot steekproeven met teruglegging of - indien geen teruglegging plaatsvindt - tot populaties wier omvang groot is in vergelijking met de daaruit getrokken steekproef. We behoeven dan dus slechts de theorie van steekproeven met teruglegging te bestuderen. Een dergelijke steekproef ontstaat per definitie door achtereenvolgens n trekkingen uit dezelfde populatie

te nemen. Immers na elke trekking en de bijbehorende waarneming aan het getrokken object wordt dit object weer bij de populatie gevoegd, zodat de populatie bij de volgende trekking onveranderd is.

We zullen nu 2 eisen geven waaraan deze trekkingen moeten voldoen:

1. Elke trekking moet aselect zijn, d.w.z. alle objecten uit de populatie moeten gelijkwaardig zijn bij de trekking, we mogen niet een bepaald object selecteren, we moeten het te trekken object voetstoots (Eng. at random) bepalen.
2. De trekkingen moeten onderling onafhankelijk zijn, d.w.z. de keuze van een object mag niet afhangen van het resultaat van de vorige trekkingen.

1.4. Hulpmiddelen bij het nemen van steekproeven.

Bij het nemen van een steekproef dient men er steeds voor te waken dat aan de in 1.3. genoemde eisen voldaan is. Bij het trekken van knikkers uit een vaas kan men dit proberen te verwezenlijken door voor elke trekking zo goed mogelijk te schudden en dit te herhalen nadat de getrokken knikker is teruggelegd. Indien echter de rode knikkers bijv. iets groter zijn dan de witte zullen de rode knikkers bij het schudden de neiging hebben zich bovenin de vaas op te hopen. Het gevolg hiervan is dan dat de knikkers niet meer gelijkwaardig zijn bij de trekking, men zal te vaak een rode knikker trekken.

Er zijn echter hulpmiddelen die in vele gevallen uitkomst brengen. We gaan dan als volgt te werk: We nummeren de objecten van de populatie in één of andere volgorde. Een menselijke populatie bijv. alfabetisch, bij goederen uit seriefabricage naar volgorde van fabricage etc. etc. Nu bepalen we door loting welke nummers we in de steekproef zullen opnemen. Daartoe kunnen we een vaas met tien gelijke knikkers, genummerd 0, 1, ..., 9 nemen en hieruit met teruglegging trekkingen doen.

Wenst men bijv. een steekproef met omvang 50 uit een populatie met omvang 1000, dan nummeren we de objecten van de populatie van 000 t/m 999. Uit onze vaas met 10 knikkers

doen we nu 50 maal telkens 3 trekkingen met teruglegging. Zo krijgen we 50 groepjes van 3 cijfers, dus 50 nummers onder de 1000. Onze steekproef stellen we nu samen door die 50 objecten uit de populatie te nemen, wier nummers juist overeenkomen met de 50 groepjes van 3 cijfers die we getrokken hebben. Op deze wijze wordt dan een steekproef met teruglegging verkregen, want onder de 50 nummers kunnen gelijke voorkomen. Wenst men een steekproef zonder teruglegging, dan schrappen men nummers, die reeds eerder voorgekomen zijn en zette het lotingsprocédé voort tot 50 verschillende nummers onder 1000 verkregen zijn.

Deze methode heeft nog wel enige bezwaren. Het kost, vooral als het om grote steekproeven gaat, nogal veel tijd om alle nodige lotingen uit te voeren. Er zijn echter tabellen geconstrueerd die de resultaten van grote aantallen trekkingen uit een vaas met tien genummerde knikkers geven. (Men heeft deze getallen niet verkregen door werkelijk knikkers uit een vaas te trekken, doch met behulp van snellere methoden, die daarmee essentieel gelijkwaardig zijn.) Zo bevatten bijv. de "Tables of random sampling numbers" van M.G. KENDALL en B. BABINGTON SMITH (Cambridge University Press 1946) 100.000 trekkingen uit de populatie 0, 1, ... 9.

Wenst men nu weer een steekproef van 50 elementen uit de populatie met omvang 1000, dan hoeft men slechts, te beginnen op een willekeurig punt in de tabel, de 50 volgende groepjes van 3 cijfers te nemen.

Soms is het niet mogelijk deze methode toe te passen, eenvoudig omdat het niet mogelijk is de te onderzoeken populatie te nummeren, omdat niet alle objecten uit de populatie bereikbaar zijn. Als men een steekproef wil nemen uit de lopende productie van een fabriek, kan men geen exemplaren nemen die reeds verzonden zijn of die nog geproduceerd moeten worden.

In dergelijke gevallen zoekt men naar een steekproefmethode waarvan men verwacht dat toch zo goed mogelijk aan de eisen van 1.3. voldaan is. Wij zullen hier niet verder op ingaan, doch wijzen er nogmaals op, dat men zich steeds goed rekenschap dient te geven van de noodzaak aan deze eisen te voldoen. Is aan de eisen slechts gedeeltelijk voldaan, dan

zijn de in deze cursus te bespreken eenvoudige statistische technieken niet zonder voorbehoud van toepassing. In sommige gevallen bestaan er echter gecompliceerde methoden die dan nog wel toepasbaar zijn.

Hoofdstuk 2.

Enige experimenten.

Keren wij thans terug tot het probleem van de vaas met knikkers, waarvan een onbekende fractie rood gekleurd is. Uit deze vaas nemen we nu een steekproef van n knikkers en we noteren het aantal rode knikkers dat wij daarbij aantreffen. Wij geven dit aantal aan met x . Herhalen we dit experiment, dan zullen we in het algemeen een andere waarde voor n vinden. Dit nu is een typisch statistisch verschijnsel: herhaling van eenzelfde experiment naar beste weten op dezelfde wijze uitgevoerd geeft afwijkende uitkomsten. Het aantal rode knikkers in een steekproef van n stuks varieert van steekproef tot steekproef. Een dergelijke variërende grootte noemen we een stochastische grootte (Engels: random variable). Dit woord is afgeleid van het Griekse woord " $\sigma\tau\omicron\chi\alpha\varsigma\tau\omicron\mu\alpha\iota$ " dat "gissen" en ook "mikken" betekent. Wij zullen in het volgende stochastische grootheden steeds onderstrepen: $\underline{x}$ is het aantal rode knikkers in een steekproef van n knikkers uit onze vaas.

De waargenomen waarden van $\underline{x}$ in concrete gevallen geven we aan met $x_1, x_2, \dots, x_n$, zonder onderstreping. In een bepaald geval hebben wij 30 keer een steekproef van 25 uit een vaas met witte en rode knikkers getrokken ($n=25$, $m=30$). Wij vonden voor $\underline{x}$ de volgende waarden (in volgorde van waarneming):

(I) 1 3 0 1 0 1 2 4 2 3 1 4 1 1 0 1 2 3 3 2 2 2 4 3 3 3 2 3 4 1.

Het is duidelijk, dat de in (1) gegeven waargenomen waarden ieder op zichzelf reeds enige informatie verschaffen over de onbekende fractie p van rode knikkers. Als p dicht bij 1 lag, d.w.z. als bijna alle knikkers in de vaas rood waren, zouden we niet zulke lage uitkomsten hebben gekregen. Het is dus aannemelijk dat p klein is, d.w.z. dat er weinig rode knikkers in de vaas zitten.

Om een duidelijk overzicht te krijgen van de resultaten, zien wij van de volgorde daarvan af en vatten we ze samen in een histogram of frequentiediagram (zie fig. 2.1). De samenstelling daarvan is uit de figuur duidelijk:

In horizontale richting zijn de waarden die x heeft aangenomen uitgezet, en in verticale richting is het aantal malen dat x deze waarde heeft aangenomen uitgezet.

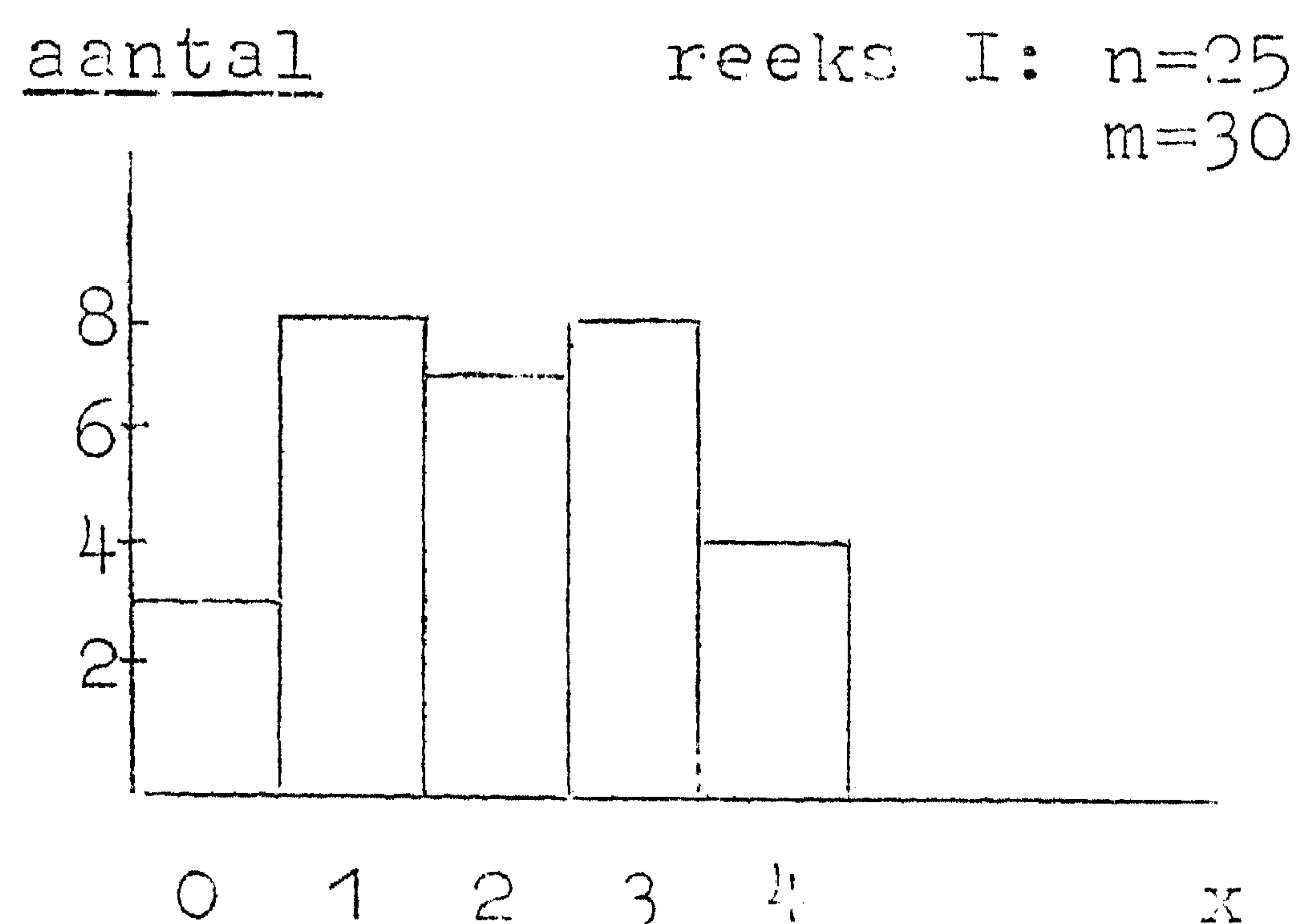


fig. 2.1.

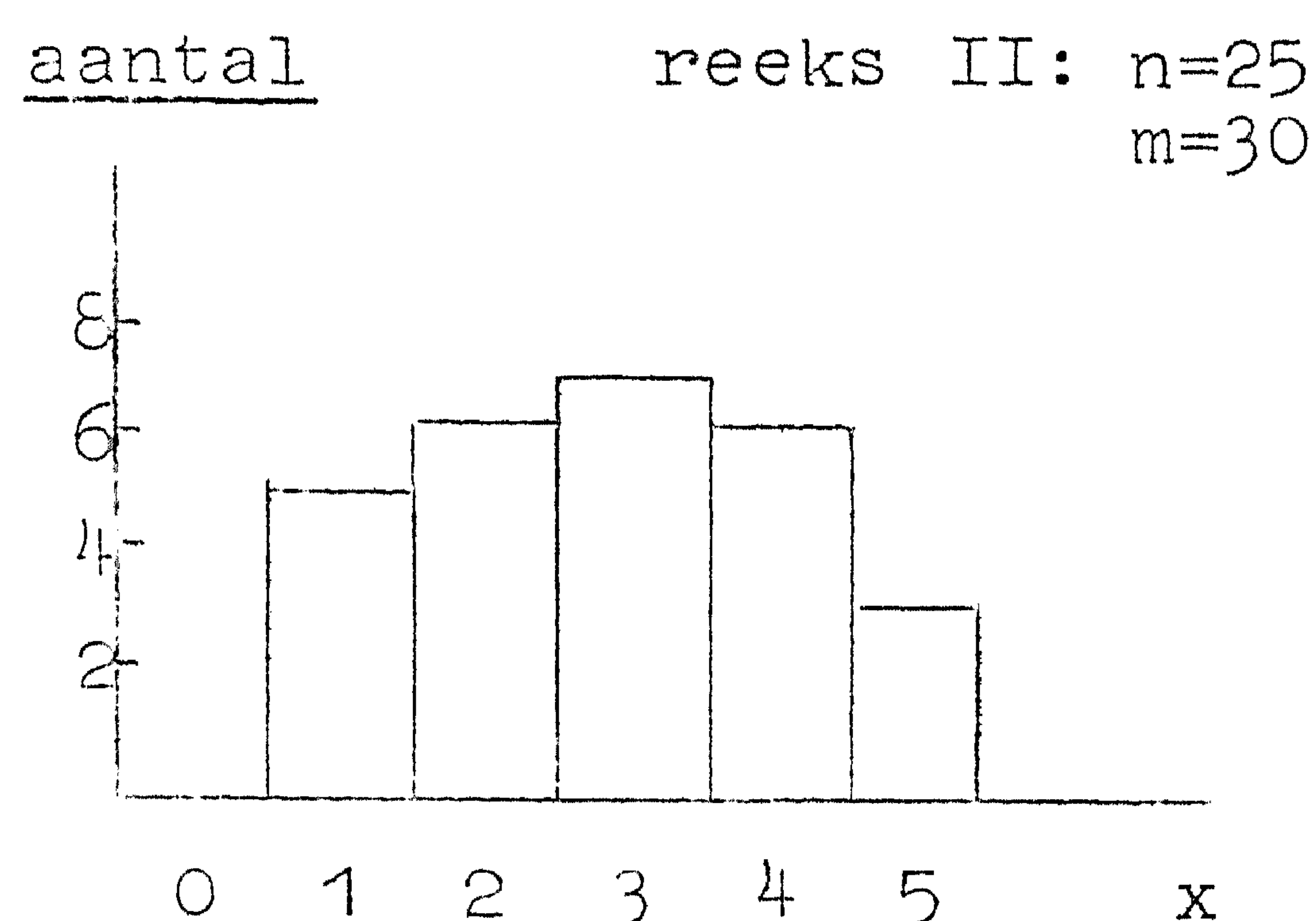


fig. 2.2.

Een dergelijke figuur geeft in één oogopslag een duidelijk overzicht van de waarnemingen. Alleen de volgorde der waarnemingen gaat hierbij verloren.

Herhaling van het experiment gaf de reeks waarnemingen:

(II) 2 2 2 0 3 5 3 5 2 3 4 2 5 3 1 1 4 1 4 0 4 4 1 1 3 3 3 0 4 2.

Ook deze waarnemingen hebben we in de vorm van een histogram samengevat (zie fig. 2.2).

We zien dat er ondanks de verschillen toch wel enige overeenkomst is tussen de waarnemingsreeksen (I) en (II): In beide gevallen varieert x slechts van 0 tot hoogstens 5, en de grote meerderheid der gevonden waarden van x is niet kleiner dan 1 en niet groter dan 4. Bij deze experimenten ligt de fractie rode knikkers in een steekproef van 25 stuks dus voornamelijk tussen $\frac{1}{25} = 0,04$ en $\frac{4}{25} = 0,16$.

Een typisch verschijnsel is dat deze statistische regelmaat duidelijker tot uiting komt indien men de waarnemingsreeksen verlengt, dus als men n groter kiest. Om dit te demonstreren zijn twee series van 100 steekproeven van 25 stuks uit dezelfde vaas genomen. De resultaten zijn samengevat in 2 histogrammen (fig. 2.3 en fig. 2.4).

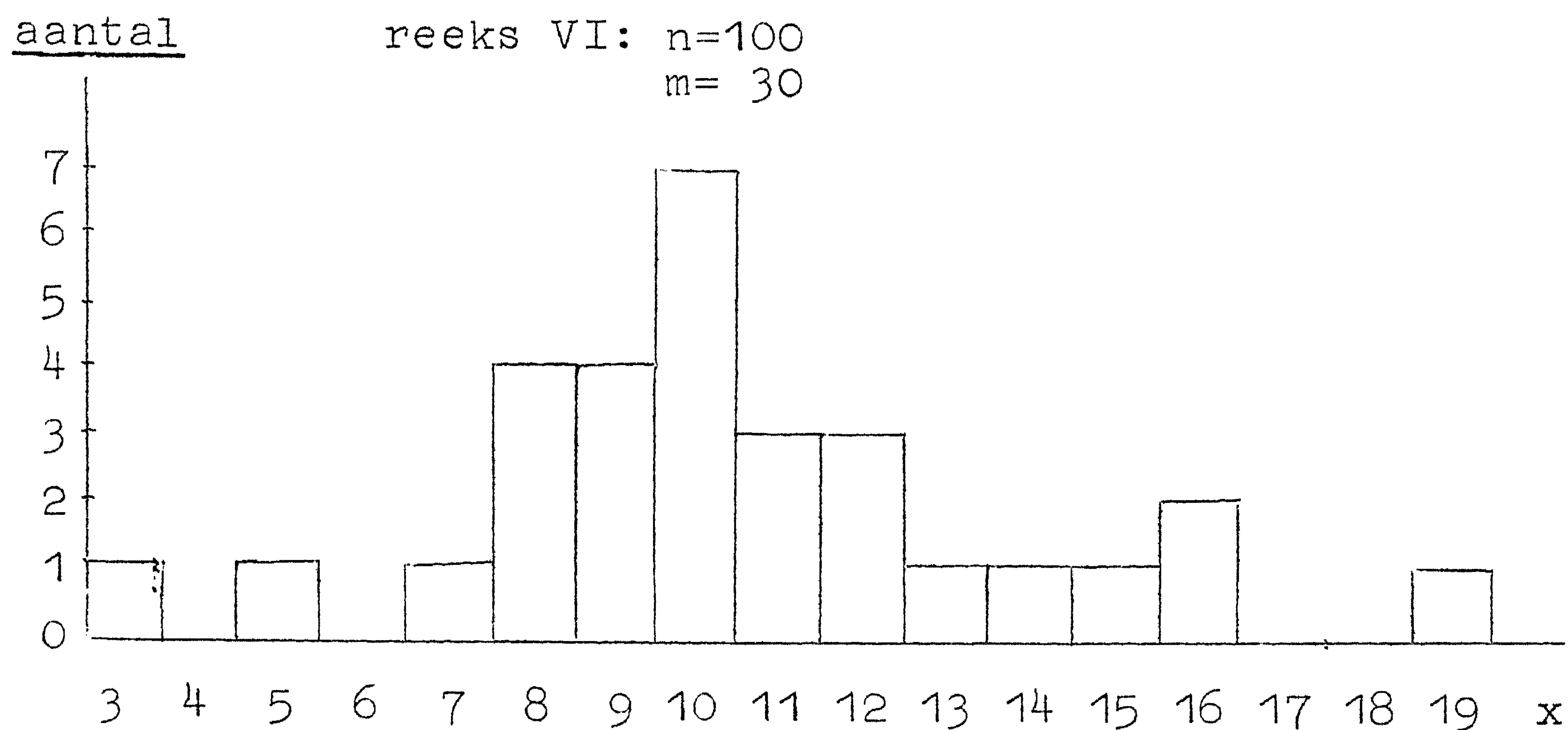


fig. 2.6.

In beide waarnemingsreeksen ligt $\frac{x}{n}$ in de grote meerderheid der gevallen tussen 0,06 en 0,12. Aangezien het voor de hand ligt om de fractie rode knikkers in de steekproef te gebruiken als schatting voor de onbekende fractie rode knikkers in de vaas, kunnen we concluderen dat de nauwkeurigheid van deze schatting blijkbaar toeneemt naarmate we meer trekkingen doen. Dit is een zeer algemeen ervaringsfeit, dat we overal waar aselechte, onderling onafhankelijke trekkingen uit een populatie genomen worden, kunnen constateren. In de statistiek wordt bijna voortdurend gebruik van deze eigenschap van aselechte en onderling onafhankelijke trekkingen gemaakt. In het volgende zullen we zien op welke wijze dit geschiedt en hoe we het bovengenoemde nader kunnen preciseren.

Hoofdstuk 3.

Waarschijnlijkheidsrekening.

3.1. Waarschijnlijkheid.

Zij π een eindige populatie met omvang N , terwijl $N(A)$ objecten van π een bepaald kenmerk A (bijv. een rode kleur) bezitten. Doen we nu een aselecte trekking uit π dan kan het zijn dat we een object met kenmerk A trekken, doch het is ook mogelijk dat dit niet het geval is. "Het vinden van kenmerk A bij een aselecte trekking" is blijkbaar een mogelijke, doch niet zekere, gebeurtenis. Een dergelijke mogelijke gebeurtenis noemen we een eventualiteit.

De kans of waarschijnlijkheid van deze eventualiteit definiëren we als $\frac{N(A)}{N}$, in formule:

$$P \{ A \} \stackrel{\text{def}}{=} \frac{N(A)}{N} \quad (P = \text{probabilitas} = \text{waarschijnlijkheid})$$

Een onmogelijke eventualiteit, d.w.z. het vinden van een kenmerk B dat geen enkel object uit π bezit, krijgt blijkbaar de waarschijnlijkheid 0 en evenzo heeft de zekere eventualiteit, d.w.z. het vinden van een kenmerk C , dat elk object uit π bezit, de waarschijnlijkheid 1.

Als de omvang van π onbepaald of oneindig is, dan is het quotiënt $\frac{N(A)}{N}$ onbepaald, zodat we het kansbegrip niet op dezelfde wijze als bij eindige populaties kunnen invoeren. Toch kan men in die gevallen nog wel spreken over een fractie p van de populatie van objecten die het kenmerk A bezitten. (Wij zullen hier niet ingaan op de streng-mathematische definitie van het begrip fractie in dit verband, daar dit ons ver buiten het kader van deze cursus zou voeren). We definiëren nu:

$$P \{ A \} \stackrel{\text{def}}{=} p.$$

Voorbeelden:

- a. Zij π de vaas met witte en rode knikkers. Als π in totaal N knikkers bevat, waarvan er M rood zijn, dan is de kans dat een aselecte trekking een rode knikker oplevert dus $p \stackrel{\text{def}}{=} \frac{M}{N}$.

- b. Zij π de denkbeeldige verzameling van alle mogelijke worpen met een munt. Als we afzien van de mogelijkheid dat de munt op zijn kant blijft staan, dan is de kans op "kruis" gelijk aan de fractie worpen die de uitslag "kruis" geven; is deze gelijk $\frac{1}{2}$, dan wordt de munt zuiver genoemd. Dan is dus: $P\{K\} = P\{M\} = \frac{1}{2}$.

Het komt voor dat bij elk object uit de populatie π één of meer getallen behoren. Bij een aselechte trekking krijgen we één object uit π , en het (de) bijbehorende getal(len) kunnen we dan waarnemen. Aangezien de waargenomen getallen bij andere trekkingen andere waarden kunnen hebben, hebben we hier stochastische grootheden op de populatie π .

Voorbeelden:

- c. Zij π de denkbeeldige verzameling van alle mogelijke worpen met een dobbelsteen. Bij elk dezer worpen hoort een getal, nl. de uitkomst $\underline{x}$. Dit is een stochastische grootheid. Gooien we één keer met een dobbelsteen, dan kunnen we dat opvatten als een aselechte trekking uit π , en de uitkomst x is dan een waarneming van $\underline{x}$.
- d. Zij π de verzameling van alle recruten van de lichte 1959. Bij elk dezer recruten behoort nu een lengte $\underline{l}$ en een gewicht $\underline{g}$. Dit zijn 2 stochastische grootheden. Trekken we aselechte een recruit, dan zijn zijn lengte l en gewicht g waarnemingen van $\underline{l}$ en $\underline{g}$.
- e. Zij π de denkbeeldige verzameling van alle mogelijke steekproeven van n stuks uit de vaas met knikkers, door aselechte, onderling onafhankelijke trekkingen verkregen. Bij elk dezer steekproeven behoort nu een frequentie $\underline{x}$ (het aantal rode knikkers) en een frequentiequotiënt $\frac{\underline{x}}{n}$.

Als $\underline{x}$ nu een stochastische grootheid op de populatie π is en x een getal is, dan volgt uit het bovenstaande de definitie van de kans dat $\underline{x}$ hoogstens x is: $P\{\underline{x} \leq x\}$. Immers een bepaalde fractie objecten uit π heeft het kenmerk " $\underline{x} \leq x$ ", zodat $P\{\underline{x} \leq x\}$ per definitie gelijk is aan deze fractie. We zeggen dat $\underline{x}$ een waarschijnlijkheidsverdeling heeft,

welke gegeven wordt door de verdelingsfunctie $F(x) \stackrel{\text{def}}{=} P\{\underline{x} \leq x\}$.

3.2. Het rekenen met waarschijnlijkheden.

De waarschijnlijkheidsrekening houdt zich bezig met het berekenen van onbekende waarschijnlijkheden uit bekende waarschijnlijkheden. Fundamenteel zijn daarbij een aantal rekenregels die wij hier zonder bewijs geven. In het geval van een eindige populatie met aselechte trekking zijn ze echter gemakkelijk te verifiëren.

Als A en B twee eventualiteiten zijn, dan is

$$(3.2.1) \quad P\{A \text{ of/en } B\} = P\{A\} + P\{B\} - P\{A \text{ en } B\}.$$

Voorbeeld: Zij bij het gooien met een zuivere dobbelsteen A het gooien van een even aantal ogen en B het gooien van een zes. We hebben dan $P\{A \text{ of/en } B\} = \frac{1}{2} + \frac{1}{6} - \frac{1}{6} = \frac{1}{2}$.

Een bijzonder geval krijgen we als A en B elkaar uitsluiten, d.w.z. als A en B niet tegelijkertijd op kunnen treden. Dan is immers $P\{A \text{ en } B\} = 0$, zodat we hebben:

Als A en B elkaar uitsluitende eventualiteiten zijn, dan is

$$(3.2.2) \quad P\{A \text{ of } B\} = P\{A\} + P\{B\}.$$

Voorbeeld: We trekken een speelkaart uit een goed geschud spel kaarten. Als A het trekken van een schoppenkaart is en B het trekken van ruitenboer is, dan is

$$P\{A \text{ of } B\} = \frac{13}{52} + \frac{1}{52} = \frac{14}{52}.$$

(3.2.2) kan gegeneraliseerd worden als volgt:

Als $A_1, A_2, \dots, A_K$ elkaar uitsluitende eventualiteiten zijn, d.w.z. als geen tweetal dezer eventualiteiten tegelijkertijd kan optreden, dan is

$$(3.2.3) \quad P\{A_1 \text{ of } A_2 \text{ of } \dots \text{ of } A_K\} = P\{A_1\} + P\{A_2\} + \dots + P\{A_K\}.$$

Voorbeeld: Zij bij het gooien met een zuivere dobbelsteen A_1 het gooien van een 1, A_2 het gooien van een 2, A_3 het gooien van een 3 en A_4 het gooien van een 4. Als nu B het gooien van hoogstens 4 is, dan is $P\{B\} = P\{A_1\} + P\{A_2\} + P\{A_3\} + P\{A_4\} = \frac{2}{3}$.

Als A en B twee eventualiteiten zijn die elkaars al dan niet optreden niet beïnvloeden, dan is

$$(3.2.4) \quad P\{A \text{ en } B\} = P\{A\} \cdot P\{B\}.$$

Als (3.2.4) geldt, noemt men A en B stochastisch onafhankelijk.

Voorbeelden:

- a. We gooien tweemaal achtereenvolgens met een munt. We noemen A de eventualiteit dat het resultaat van de eerste worp "kruis" is, en B dat het resultaat van de tweede worp "kruis" is. Als men "eerlijk" werpt, dan wordt de uitkomst van de tweede worp niet door die van de eerste worp beïnvloed, zodat A en B stochastisch onafhankelijk zijn en, als de munt zuiver is,

$$P\{A \text{ en } B\} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

- b. Bij het gooien met een zuivere dobbelsteen noemen we A het gooien van hoogstens 4, en B het gooien van een even aantal ogen. Nu is $P\{A\} = \frac{2}{3}$, $P\{B\} = \frac{1}{2}$ en $P\{A \text{ en } B\} = \frac{1}{3}$. Dit laatste omdat (A en B) het gooien van een 2 of een 4 is. Derhalve is aan (3.2.4) voldaan, zodat A en B stochastisch onafhankelijk zijn.
- c. Zij nu, weer bij het gooien met een zuivere dobbelsteen, A het gooien van hoogstens 3 ogen en B als in voorbeeld b. Dan is nu $P\{A\} = \frac{1}{2}$, $P\{B\} = \frac{1}{2}$ en $P\{A \text{ en } B\} = \frac{1}{6}$, want (A en B) is het gooien van een 2, zodat nu niet voldaan is aan (3.2.4): A en B zijn stochastisch afhankelijk. Dit kunnen we ook wel anders inzien door te bedenken, dat als A optreedt er nog drie - even waarschijnlijke - mogelijke uitkomsten zijn, te weten 1, 2 en 3, waarbij er slechts één is waarbij B optreedt, nl. 2. Treedt A niet op, dan zijn er ook 3 - even waarschijnlijke - mogelijke uitkomsten, waarbij er twee zijn waarbij B optreedt.
- d. Zij π de Nederlandse bevolking op een zeker moment. Als A het bezit van een auto en B het bezit van een rijbewijs is, dan zijn A en B stochastisch afhankelijk. Kiezen we aselekt iemand uit de populatie, en is A vervuld, dan is B immers bijna altijd ook vervuld. Als A niet vervuld is,

zal B echter veel minder vaak vervuld zijn.

3.3. De empirische en de theoretische wet der grote aantallen.

In hoofdstuk 2 maakten wij reeds kennis met wat men wel noemt de empirische wet der grote aantallen. Als men aselechte onafhankelijke trekkingen doet uit een populatie, waarbij men steeds een kans p heeft op een bepaalde uitkomst - die we een succes zullen noemen - dan is het aantal x_n successen dat we in n trekkingen vinden een stochastische grootheid, evenals het frequentiequotiënt $\underline{f}_n = \frac{x_n}{n}$, en naarmate we meer trekkingen nemen ligt $\underline{f}_n$ meestal dichter bij p .

Met behulp van de waarschijnlijkheidsrekening kunnen we dit vaag geformuleerde ervaringsfeit scherper formuleren: Men kan nl. bewijzen dat voor elk positief getal ε , hoe klein ook gekozen

$$(3.3.1) \quad \lim_{n \rightarrow \infty} P \{ |\underline{f}_n - p| > \varepsilon \} = 0.$$

Dit wil zeggen dat wij een willekeurig klein intervalletje $(p - \varepsilon, p + \varepsilon)$ kunnen beschouwen, en dat de kans dat de fractie successen hier buiten zal vallen willekeurig klein gemaakt kan worden door n groot genoeg te kiezen.

Aangezien een kleine kans wegens de definitie van het kansbegrip bij een zeldzame gebeurtenis behoort, stelt deze stelling ons in staat voorspellingen te doen die slechts zeer zelden niet uitkomen, zoals wij in het volgende zullen zien.

Hoofdstuk 4.

Het schatten van een onbekende kans.

4.1. Betrouwbaarheidsgrenzen.

We beschouwen nu weer problemen van het in 1.1 behandelde type: Gegeven is een populatie π , zodanig dat er een (onbekende) kans p is dat aselechte trekking een object met een bepaald kenmerk - we zullen zo'n object voortaan een succes noemen - oplevert. Gevraagd wordt p zo goed mogelijk te schatten op grond van een steekproefexperiment.

Nemen we een steekproef van n stuks uit π , en vinden we daarin x successen, dan kunnen we zeggen dat p wel in de buurt van $\frac{x}{n}$ zal liggen. We gebruiken dan $\frac{x}{n}$ als schatting voor p . Nemen we nog eens een steekproef, dan zullen we in het algemeen een ander aantal successen, en dus ook een andere schatting voor p vinden: Onze schatting is een waarneming van een stochastische grootte en als zodanig aan steekproefvariatie onderhevig. We zagen al in hoofdstuk 2 dat deze steekproefvariatie kleiner wordt als we de steekproefgrootte n groter kiezen. De schatting voor p wordt dus nauwkeuriger naarmate n groter gekozen wordt. Een exacte quantitative uitspraak over de nauwkeurigheid van de schatting kunnen we echter niet doen, behalve dan dat $0 \leq p \leq 1$ moet zijn.

Men zou daarom liever een bovengrens en een ondergrens voor p willen hebben, aangezien twee zulke grenzen ons meer vertellen dan één schatting met onbekende nauwkeurigheid. Aan dit verlangen nu komen de zogenaamde betrouwbaarheidsgrenzen tegemoet. Dit zijn twee grenzen, een bovengrens $p^*(x)$ en een ondergrens $p_*(x)$, voor p , die men berekent uit het waargenomen aantal successen x in een steekproef. Daar ze van x afhangen, hebben ook deze grenzen een steekproefvariantie. Het essentiële is nu, dat men er een getal α bij opgeeft, de eenzijdige onbetrouwbaarheidsdrempel, die een bovengrens is voor de kans dat $p > p^*(x)$ en ook voor de kans dat $p < p_*(x)$. Op grond van een steekproef waarin we x successen vinden, komen we zo tot de uitspraak

$$(4.1.1) \quad p > p_*(x)$$

met onbetrouwbaarheidsdrempel α , hetgeen wil zeggen dat deze uitspraak fout kan zijn, doch dat de kans dat hij fout is hoogstens α is, m.a.w. als we een lange reeks experimenten doen, en na elk experiment een conclusie van de vorm (4.1.1) trekken, dan zal een fractie, die hoogstens ongeveer α bedraagt, van deze conclusies fout zijn.

Geheel analoog kunnen we, na een steekproef genomen te hebben waarin we x successen hebben gevonden, een uitspraak van de volgende vorm doen:

$$(4.1.2) \quad p < p^*(n), \text{ onbetrouwbaarheidsdrempel } \alpha.$$

Uiteraard kunnen we (4.1.1) en (4.1.2) ook combineren tot

$$(4.1.3) \quad p_*(x) < p < p^*(x), \text{ onbetrouwbaarheidsdrempel } \underline{2\alpha}.$$

4.2. Bepaling der betrouwbaarheidsgrenzen.

Voor sommige waarden van n en α bestaan er tabellen die de betrouwbaarheidsgrenzen als functie van x geven. Wij geven hier twee van zulke tabellen.

Tabel 4.I.

Betrouwbaarheidsgrenzen voor p . $n=25$

(onbetrouwbaarheidsdrempel voor beide grenzen apart 0,05)

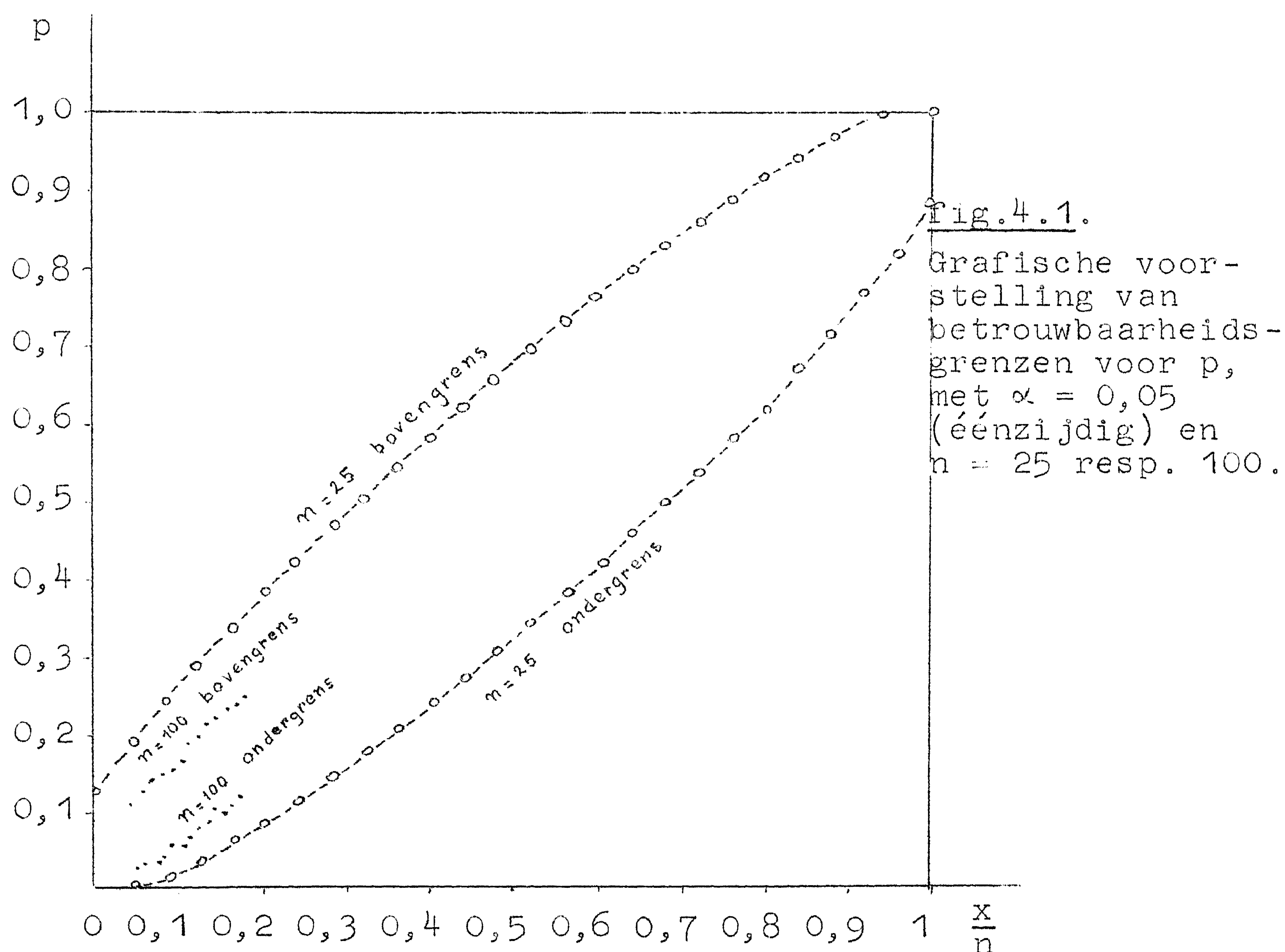
x	ondergrens voor p	bovengrens voor p	x	ondergrens voor p	bovengrens voor p
0	0,00	0,11	13	0,34	0,70
1	0,00	0,18	14	0,38	0,73
2	0,01	0,23	15	0,42	0,76
3	0,03	0,28	16	0,46	0,80
4	0,06	0,33	17	0,50	0,83
5	0,08	0,38	18	0,54	0,86
6	0,11	0,42	19	0,58	0,89
7	0,14	0,46	20	0,62	0,92
8	0,17	0,50	21	0,67	0,94
9	0,20	0,54	22	0,72	0,97
10	0,24	0,58	23	0,77	0,99
11	0,27	0,62	24	0,82	1,00
12	0,30	0,66	25	0,89	1,00

Tabel 4.II.

Betrouwbaarheidsgrenzen voor p . $n=100$
(onbetrouwbaarheidsdrempel voor beide grenzen apart 0,05)

x	ondergrens voor p	bovengrens voor p	x	ondergrens voor p	bovengrens voor p
5	0,02	0,10	12	0,07	0,19
6	0,03	0,12	13	0,08	0,20
7	0,03	0,13	14	0,09	0,21
8	0,04	0,14	15	0,09	0,22
9	0,05	0,15	16	0,10	0,23
10	0,05	0,16	17	0,11	0,24
11	0,06	0,18			

Volledige tabellen voor vele waarden van n en verschillende waarden van α zijn niet beschikbaar. Wel zijn er echter grafieken waaruit de betrouwbaarheidsgrenzen afgelezen kunnen worden. Ter illustratie hebben wij de tabellen 4.I en 4.II in beeld gebracht in fig. 4.1.



Op de horizontale as is $\frac{x}{n}$ uitgezet om een grootheid te krijgen die voor alle n van 0 tot 1 loopt. Hebben we bijv. $x=3$ en $n=25$, dan lezen we in verticale richting af bij $\frac{x}{n} = \frac{3}{25} = 0,12$ dat de grenzen voor p 0,03 en 0,28 zijn. Het aflezen gaat iets minder vlot dan uit een tabel, maar men kan op deze wijze een groot aantal tabellen voor een aantal waarden van n in één grafiek samenvatten. Bovendien kan men gemakkelijk op het oog interpoleren tussen krommen, die bij verschillende waarden van n behoren, om voor een niet opgenomen waarde van n toch uit de grafiek de betrouwbaarheids-grenzen te bepalen.

In de grafiek zijn eigenlijk alleen de ronde stippen van betekenis, daar alleen de daarbij behorende waarden van $\frac{x}{n}$ voor kunnen komen. Wij hebben daarom deze punten niet door een lijn vervangen, maar alleen, voor de samenhang van de figuur, door een onderbroken lijn verbonden. Voor $n=100$ was dit niet nodig, omdat in dat geval de punten veel dichter bij elkaar liggen. Het feit dat de lijn niet geheel vloeiend is - voor $n=100$ is dit vooral duidelijk - wordt vooral door afronding veroorzaakt.

Deze grafieken vindt men, veel uitgebreider, in het ook overigens zeer aanbevelenswaardige boek over statistiek van W.J. DIXON en F.J. MASSEY, Introduction to Statistical Analysis, 2nd edition, McGraw-Hill Book Company, New York-Toronto-Londen, 1957, op blz. 413-416. Wij zullen dit boek voortaan kortweg aangeven als "DIXON en MASSEY", zonder verdere aanduiding. DIXON en MASSEY geven niet de onbetrouwbaarheidsdrempel op, maar de z.g. betrouwbaarheidscoëfficiënt ("confidence coefficient"). Noemen we de eenzijdige onbetrouwbaarheidsdrempel weer α , dan is deze betrouwbaarheidscoëfficiënt $1-2\alpha$, nl. de kans dat p tussen de betrouwbaarheidsgrenzen ligt. De grafiek op blz. 414 (betrouwbaarheidscoëfficiënt 0,90) komt dus overeen met fig. 4.1. De grafieken zelf geven zij als vloeiende lijnen. Deze lijnen geven niet de exacte grenzen, die wij in de tabellen 4.I en 4.II en in fig. 4.1 hebben gegeven, maar benaderingen daarvan. Vergelijking van onze tabellen met de grafieken van Dixon en Massey toont aan dat deze benadering zeer goed is voor niet

te kleine n en in het bijzonder als $\frac{x}{n}$ niet te dicht bij 0 of 1 ligt.

4.3. Benadering der betrouwbaarheidsgrenzen.

Als men niet de beschikking heeft over de juiste tabellen of grafieken, dient men de betrouwbaarheidsgrenzen zelf te berekenen. Daar exacte berekening, zonder gebruikmaking van uitgebreide tabellen, die meestal niet voorhanden zijn, zeer ingewikkeld en tijdrovend is, behandelen wij alleen enkele benaderingsmethoden. In de praktijk zijn deze meestal voldoende nauwkeurig om zinvolle resultaten te geven.

Als n niet te klein is en x en $n-x$ beide niet kleiner dan 5 zijn, kan men de bovengrens voor p berekenen uit

$$(4.3.1) \quad p^*(x) = \frac{x + \frac{1}{2} + \frac{1}{2}\zeta^2 + \zeta \sqrt{\frac{(x+\frac{1}{2})(n-x-\frac{1}{2})}{n}} + \frac{1}{4}\zeta^2}{n + \zeta^2}$$

en de ondergrens uit

$$(4.3.2) \quad p^*(x) = \frac{x - \frac{1}{2} + \frac{1}{2}\zeta^2 - \zeta \sqrt{\frac{(x-\frac{1}{2})(n-x+\frac{1}{2})}{n}} + \frac{1}{4}\zeta^2}{n + \zeta^2}.$$

In deze formules is ζ een constante, die afhangt van de gekozen onbetrouwbaarheidsdrempel. De waarden die we voor verschillende waarden van α voor ζ moeten substitueren worden gegeven in tabel 4.III.

Tabel 4.III.

Waarden voor de constante in de formules (4.3.1) en (4.3.2) voor verschillende waarden van α .

α (éénzijdig)	ζ	ζ^2
0,10	1,282	1,642
0,05	1,645	2,706
0,025	1,960	3,842
0,01	2,326	5,412
0,005	2,576	6,635

Uiteraard kunnen we weer een tweezijdig begremsd betrouwbaarheidsinterval krijgen door (4.3.1) en (4.3.2) tegelijkertijd te gebruiken. De onbetrouwbaarheidsdrempel is dan 2α . Als n groot is, terwijl $\frac{x}{n}$ niet teveel van $\frac{1}{2}$ afwijkt, dan kunnen we zonder bezwaar bepaalde termen in (4.3.1) en (4.3.2) verwaarlozen, zodat we eenvoudiger uitdrukkingen krijgen voor $p^*(x)$ en $p_*(x)$:

$$(4.3.3) \quad p^*(x) = \frac{1}{n} \left\{ x + \sqrt{\frac{x(n-x)}{n}} \right\}$$

$$(4.3.4) \quad p_*(x) = \frac{1}{n} \left\{ x - \sqrt{\frac{x(n-x)}{n}} \right\}.$$

In twijfelgevallen kan men altijd nagaan of het gebruik van de formules (4.3.3) en (4.3.4) gerechtvaardigd is, door de uitkomsten te vergelijken met de uitkomsten die (4.3.1) en (4.3.2) geven.

Voor kleine waarden van x of $(n-x)$ is de boven beschreven benaderingsmethode niet erg nauwkeurig. Voor dit geval is er echter een andere benaderingsmethode, die goede resultaten geeft als n groot is. Deze resultaten zijn samengevat in tabel 4.IV.

Tabel 4.IV.

Betrouwbaarheidsgrenzen voor $n.p$ voor grote n en $x \leq 10$ (α éénszijdig).

α x	0,005		0,025		0,05		0,10	
	ondergrens	bovengrens	ondergrens	bovengrens	ondergrens	bovengrens	ondergrens	bovengrens
0	0	5,3	0	3,7	0	3,0	0	2,3
1	0,005	7,4	0,025	5,6	0,05	4,7	0,11	3,9
2	0,1	9,3	0,2	7,2	0,4	6,3	0,5	5,3
3	0,3	11,0	0,6	8,8	0,8	7,8	1,1	6,7
4	0,6	12,6	1,0	10,2	1,4	9,2	1,7	8,0
5	1,0	14,1	1,6	11,7	2,0	10,5	2,4	9,3
6	1,5	15,6	2,2	13,1	2,6	11,8	3,1	10,5
7	2,0	17,1	2,8	14,4	3,3	13,1	3,9	11,8
8	2,5	18,5	3,4	15,8	4,0	14,4	4,7	13,0
9	3,1	20,0	4,0	17,1	4,7	15,7	5,4	14,2
10	3,7	21,3	4,7	18,4	5,4	17,0	6,2	15,4

Men gebruikt deze tabel als volgt: Heeft men x successen gevonden in een steekproef van n stuks, dan leest men in de tabel betrouwbaarheidsgrenzen (bij de gekozen α) voor $n \cdot p$ af. Deling door n geeft de gezochte grenzen voor p .

De aangegeven waarden van α zijn weer eenzijdig en gelden dus voor ieder van de grenzen afzonderlijk. Ook hier geldt, dat de bij een tweezijdig begreind interval behorende onbetrouwbaarheidsdrempel 2α is.

Tabel 4.IV is ingericht voor grote n en $x \leq 10$. Toch kunnen we deze tabel ook gebruiken als x zeer groot is, d.w.z. als $n-x$ klein is en n groot. We doen dan alsof boven de eerste kolom van de tabel niet x maar $n-x$ staat en lezen net als boven de grenzen af die bij $n-x$ behoren. Dit zijn nu grenzen voor $n - np$. Deling door n geeft ons grenzen voor $1-p$, waaruit we gemakkelijk de grenzen voor p zelf kunnen vinden.

In het geval van kleine n kunnen we tabel 4.IV niet gebruiken, en ook de formules (4.3.1) en (4.3.2) geven dan geen goede benadering meer. In dit laatste geval kunnen alleen exacte tabellen of grafieken nog uitkomst brengen.

4.4. Voorbeelden.

- a. Een partij goederen bevat een onbekende fractie p defecten. Om een indruk te krijgen van de grootte van deze fractie p heeft men een steekproef van 100 stuks genomen en geconstateerd dat daarin 7 defecten voorkwamen. Op grond van tabel 4.II of van tabel 4.IV trekken we de conclusie dat $p \leq 0,13$ met onbetrouwbaarheidsdrempel $\alpha = 0,05$. Wil dit nu zeggen dat p inderdaad niet meer dan 0,13 bedraagt? Neen, want als p bijv. 0,20 was, dan hadden we toevallig toch wel een steekproef van 100 met maar 7 defecten kunnen treffen. Wel kunnen we het volgende zeggen: Als we de bovenomschreven methode bij routine keuringen van partijen gebruiken, dan zal de gevonden bovengrens hoogstens in ongeveer 1 op de 20 gevallen fout zijn.
- b. In zeker land, dat een 2-partijen systeem heeft, wil men door middel van een enquête een voorspelling doen over de uitslag der op handen zijnde verkiezingen. Men heeft daar-

toe uit het kiezersregister op aselechte wijze 250 namen getrokken en vervolgens aan de 250 bijbehorende personen gevraagd of zij op partij A dan wel op partij B zouden stemmen. Van de ondervraagden antwoordden 99 personen dat zij op partij A zouden stemmen en 113 personen dat zij op partij B zouden stemmen, terwijl de overige 38 personen nog geen beslissing hadden genomen of het antwoord weigerden, of niet bereikbaar bleken.

Als wij aannemen dat het al dan niet krijgen van een antwoord van een kiezer en zijn partijkeuze onderling onafhankelijk zijn, dan is de fractie personen die op partij A gaat stemmen in de populatie van alle kiezers van wie we een bevredigend antwoord zouden krijgen, indien ze op de bovenomschreven wijze zouden worden ondervraagd, even groot als in de populatie van alle kiezers die ten slotte hun stem uitbrengen. Voor de eerste fractie nu kunnen we betrouwbaarheidsgrenzen opstellen. Als betrouwbaarheidsdrempel wordt gekozen $\alpha = 0,05$ (tweezijdig). Gevonden zijn 99 personen die op partij A gaan stemmen in een steekproef van 212 personen, d.w.z. $\frac{99}{212} = 47\%$.

Uit de tabel in DIXON en MASSEY (blz. 415) volgen de betrouwbaarheidsgrenzen: 40% en 53%.

We kunnen deze grenzen ook berekenen met de formules (4.3.1) en (4.3.2):

$$n = 212$$

$$x = 99$$

$$z = 1,96 \text{ (want } \alpha = 0,05 \text{ tweezijdig!).}$$

We krijgen

$$\begin{aligned} p^* &= \frac{99,50 + 1,92 + 1,96 \sqrt{\frac{99,5 \cdot 112,5}{212} + 0,96}}{212 + 3,84} = \\ &= \frac{101,42 + 1,96 \sqrt{53,76}}{215,84} = 0,536 \end{aligned}$$

$$p_{*} = \frac{98,5 + 1,92 - 1,96 \sqrt{\frac{98,5 \cdot 113,5}{212} + 0,96}}{212 + 3,84} =$$

$$= \frac{100,42 - 1,96 \sqrt{52,73}}{215,84} = 0,400.$$

Zelfs bij deze niet zeer grote n ($n=212$) maakt het weinig verschil meer of we de formules (4.3.1) en (4.3.2) dan wel (4.3.3) en (4.3.4) gebruiken. We kunnen dit constateren door ook met de laatste formules de grenzen te berekenen. We vinden dan:

$$p^{*} = \frac{99}{212} + \frac{1,96}{212} \sqrt{\frac{99 \cdot 113}{212}} = 0,467 + \frac{1,96}{212} \sqrt{52,8} = 52,9\%$$

$$p_{*} = \frac{99}{212} - \frac{1,96}{212} \sqrt{52,8} = 40,5\%.$$

Onze conclusie luidt dus dat de fractie der A-stemmers tussen 40% en 53% ligt, met een onbetrouwbaarheidsdrempel $\alpha = 0,05$. Op grond van deze gegevens, en met de gekozen α , kunnen we dus niet voorspellen welke partij de verkiezing zal winnen, daar het percentage der A-stemmers zowel groter als kleiner dan 50% kan zijn. Wil men toch een dergelijke voorspelling doen, dan zal men meer personen moeten ondervragen. Een probleem is nu hoe groot n moet zijn opdat men inderdaad met een onbetrouwbaarheidsdrempel $\alpha = 0,05$ kan voorspellen welke partij zal winnen. In zijn algemeenheid is op deze vraag geen antwoord te geven. Immers, als de fractie der A-stemmers precies 50% is, zal men, hoe groot n ook gekozen wordt, een kans van minstens 95% hebben dat het betrouwbaarheidsinterval het getal 0,50 bevat, in welk geval een voorspelling niet mogelijk is. Men zal zijn eisen dus moeten matigen en moeten vragen hoe groot n minstens moet zijn opdat we nog een voorspelling kunnen doen als de fractie der A-stemmers bijv. niet groter dan 47% is. Ook nu is de vraag niet te beantwoorden want met een fractie A-stemmers van 47% is het nog mogelijk, hoe groot n ook gekozen wordt, dat we in de steekproef een fractie A-stemmers vinden vlak bij 50%,

zodat het betrouwbaarheidsinterval het getal 0,50 bevat, en een voorspelling niet mogelijk is. Het is pas mogelijk de vraag te beantwoorden als hij als volgt gesteld wordt: "Hoe groot moet n minstens gekozen worden opdat er een kans van minstens bijv. 90% is dat we met een onbetrouwbaarheidsdrempel 0,05 (tweezijdig) kunnen voorspellen welke partij de verkiezing zal winnen, als het percentage der A-stemmers hoogstens 47% bedraagt. De wijze waarop n berekend kan worden zal hier niet besproken worden. Het blijkt dat $n \approx 2750$ moet zijn, dus rekening houdend met het feit dat men mag verwachten dat 15% van de ondervraagden geen antwoord geeft, moet men 3500 personen ondervragen.

Hoofdstuk 5.

De tekentoets.

5.1. Verband met betrouwbaarheidsgrenzen.

In hoofdstuk 4 hebben we gezien, hoe men op grond van een steekproef betrouwbaarheidsgrenzen kan construeren voor de fractie objecten van een populatie, die een bepaald kenmerk bezitten. In de praktijk komt het vaak voor, dat men een dergelijke intervalschatting niet nodig heeft, doch alleen wil weten of genoemde fractie $\frac{1}{2}$ kan zijn of duidelijk van $\frac{1}{2}$ afwijkt. In voorbeeld b van paragraaf 4.4. is het vooral van belang om te weten of de fractie A-stemmers kleiner of groter dan $\frac{1}{2}$ is, dan wel niet duidelijk van $\frac{1}{2}$ verschilt.

We kunnen dit probleem ook anders formuleren: Als men aselekt een kiezer trekt uit de populatie van alle kiezers, dan heeft men een kans p dat deze voor partij A zal stemmen. Deze kans p is per definitie gelijk aan de fractie A-stemmers in de populatie. Iemand die wil voorspellen welke partij (op dit moment) de overwinning zou behalen, zal belang hebben bij een methode, waarmee hij kan nagaan of $p \approx \frac{1}{2}$, $p > \frac{1}{2}$ of $p < \frac{1}{2}$ is, terwijl de exacte waarde van p hem minder interesseert.

Een ander voorbeeld is het volgende: Met behulp van een hier niet verder te beschrijven apparaat wordt de doorslagspanning van een aantal oliemonsters bepaald. Iedere bepaling geschiedt in duplo en het interesseert ons of wellicht door veranderingen in het monster tengevolge van de elektrische stroom de tweede bepaling systematisch afwijkt van de eerste. Als er geen verandering optreedt zal de kans dat de tweede bepaling hoger uitvalt dan de eerste gelijk aan $\frac{1}{2}$ moeten zijn.

(Zowel in dit als in het voorafgaande voorbeeld, laten we de mogelijkheid van onbesliste gevallen (blanco stemmen of gelijke bepalingen) voorlopig buiten beschouwing.)

Wij hebben nu een methode nodig om aan de hand van een reeks waarnemingen na te gaan of de kans op het optreden van een waarneming, waarbij de tweede bepaling groter is dan de

eerste, al dan niet gelijk is aan $\frac{1}{2}$.

Uiteraard kan de statistiek ons geen methode leveren, welke ons in staat zou stellen een zekere uitspraak te doen. Wel is er een methode, waarmee men slechts zelden tot de conclusie komt dat p niet gelijk is aan $\frac{1}{2}$ als dit in feite wel het geval is. Een dergelijke methode noemt men een toets voor de hypothese $p=\frac{1}{2}$. De kans, dat men de hypothese $p=\frac{1}{2}$ verworpt, terwijl zij juist is, wordt de kans op een fout van de eerste soort, of ook wel de onbetrouwbaarheid van de toets genoemd.

Wij kunnen een eenvoudige toets voor de hypothese $p=\frac{1}{2}$ afleiden uit de methode der betrouwbaarheidsgrenzen. Zij p de kans dat een aselechte trekking uit de beschouwde populatie π een bepaald kenmerk, dat we weer een succes noemen, oplevert. We nemen nu een steekproef ter grootte n en noemen het aantal successen daarin x . Vervolgens bepalen we de beide betrouwbaarheidsgrenzen behorend bij een onbetrouwbaarheidsdrempel α (tweezijdig) voor p . Als nu het getal $\frac{1}{2}$ buiten deze grenzen ligt verwerpen we de hypothese $p=\frac{1}{2}$. De definitie van de betrouwbaarheidsgrenzen is immers zodanig, dat deze grenzen stochastische grootheden zijn op de populatie van alle mogelijke steekproeven ter grootte n uit onze populatie π , met de eigenschap dat hoogstens een fractie α dezer steekproeven aanleiding geeft tot een betrouwbaarheidsinterval dat p niet bevat. Hieruit volgt dat, als $p=\frac{1}{2}$ is, de kans dat het betrouwbaarheidsinterval, verkregen op grond van een aselechte steekproef van n , het getal $\frac{1}{2}$ niet bevat hoogstens α is. Het bovenomschreven voorschrift impliceert derhalve dat de kans dat we de hypothese $p=\frac{1}{2}$ verwerpen hoogstens α is als in werkelijkheid wel $p=\frac{1}{2}$ is. De onbetrouwbaarheid van de toets is dan hoogstens α , of korter: de onbetrouwbaarheidsdrempel van de toets is α .

Deze toets voor de hypothese $p=\frac{1}{2}$ wordt de (tweezijdige) tekentoets genoemd. Deze naam houdt verband met het feit dat de toets vaak wordt toegepast op verschillen van paren waarnemingen, waarbij men onderzoeken wil of deze verschillen systematisch van 0 afwijken of niet. Een "succes" is dan bijv. een positief verschil en een "mislukking" een negatief

verschil, zodat men bij toepassing van de tekentoets alleen de tekens van de verschillen in aanmerking neemt. Uit het voorgaande volgt, dat de tweezijdige tekentoets ons in staat stelt op grond van een steekproef een keuze te doen tussen de volgende drie mogelijke uitspraken:

- A: $p=\frac{1}{2}$ wordt verworpen ten gunste van $p < \frac{1}{2}$
- B: $p=\frac{1}{2}$ wordt niet verworpen
- C: $p=\frac{1}{2}$ wordt verworpen ten gunste van $p > \frac{1}{2}$.

We besluiten tot A als $\frac{x}{n}$ zo klein is, dat beide betrouwbaarheidsgrenzen kleiner dan $\frac{1}{2}$ zijn, tot B als $\frac{x}{n}$ zo dicht bij $\frac{1}{2}$ ligt, dat de grenzen ter weerszijden van $\frac{1}{2}$ liggen, en tot C als $\frac{x}{n}$ zo groot is, dat beide grenzen groter dan $\frac{1}{2}$ zijn. We kunnen dit schematisch weergeven, zoals in fig. 5.1 gedaan is.

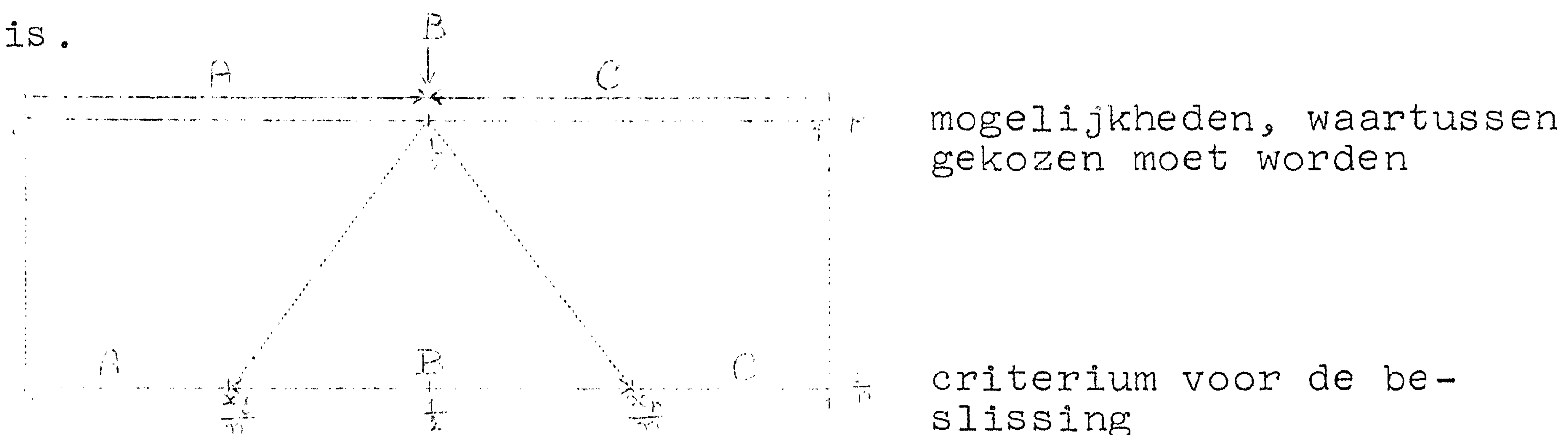


fig. 5.1. Beslissingsschema van de tweezijdige tekentoets.

In het onderste gedeelte van de figuur komt een gebied B voor, dat correspondeert met het punt B in het bovenste gedeelte; men dient hierbij wel te bedenken, dat de conclusie $p=\frac{1}{2}$ niet gerechtvaardigd is, als men een waarde $\frac{x}{n}$ in het gebied B vindt. Als $\frac{x}{n}$ in B ligt weten we slechts dat de betrouwbaarheidsgrenzen ter weerszijden van $\frac{1}{2}$ liggen. Alle andere waarden die tussen deze grenzen liggen zijn echter redelijkerwijs ook mogelijk. Om deze reden hebben wij conclusie B geformuleerd als: " $p=\frac{1}{2}$ wordt niet verworpen".

Als wij in ons voorbeeld van de oliemonsters, bij $n=25$, 5 monsters vinden waarbij de doorslagspanning bij de tweede bepaling lager, en 20 monsters waarbij deze spanning hoger is dan bij de eerste bepaling, dan zijn de betrouwbaarheidsgrenzen met onbetrouwbaarheidsdrempel 0,10 (tweezijdig) volgens tabel 4.I 0,08 en 0,38. Beide grenzen zijn lager dan

$\frac{1}{2}$; we besluiten dus dat de kans, dat de tweede bepaling lager uitvalt dan de eerste, kleiner is dan $\frac{1}{2}$. We zeggen: De door-slagspanning bij de tweede bepaling is systematisch hoger dan bij de eerste bepaling. Dit betekent dus niet dat dit altijd zo is, maar wel dat dit vaker wel dan niet het geval is.

5.2. Kritieke waarden en overschrijdingskansen.

Het gebied dat gevormd wordt door de gebieden A en C in het onderste deel van fig. 5.1 noemt men de kritieke zone of het kritieke gebied. De getallen x_1 en x_r , die de grenzen zijn tussen de gebieden A en B resp. B en C noemt men de linker resp. rechter kritieke waarde. Deze getallen zijn dus zodanig, dat men besluit tot $p < \frac{1}{2}$ als $x \leq x_1$ en tot $p > \frac{1}{2}$ als $x \geq x_r$, terwijl $p = \frac{1}{2}$ niet verworpen kan worden als $x_1 < x < x_r$ is. In tabel 5.I zijn de linker kritieke waarden getabelleerd voor de waarden 1, 2, 3, ..., 100 van n en de onbetrouwbaarheidsdrempels 0,01; 0,02; 0,05 en 0,10.

Aangezien de linker en rechter kritieke waarden symmetrisch ten opzichte van $\frac{1}{2}n$ liggen, kunnen we de rechter uit de linker kritieke waarde vinden: $x_r = n - x_1$.

Tabel 5.I geeft geen uitkomst wanneer $n > 100$ is. In dat geval kunnen we de kritieke waarden echter berekenen uit de formules (4.3.1) en (4.3.2). De linker kritieke waarde is nl. een waarde van x, zodanig, dat de bovengrens van het bijbehorende betrouwbaarheidsinterval juist $\frac{1}{2}$ is.

We stellen dus

$$p^*(x) = \frac{x + \frac{1}{2} + \frac{1}{2}\xi^2 + \xi \sqrt{\frac{(x + \frac{1}{2})(n - x - \frac{1}{2})}{n}} + \frac{1}{4}\xi^2}{n + \xi^2} = \frac{1}{2}$$

en lossen hieruit x op. We krijgen twee oplossingen, nl. $\frac{1}{2}n \pm \frac{1}{2}\sqrt{\lambda n}$. We weten echter dat $x_1 < \frac{1}{2}n$ moet zijn, zodat alleen de kleinste der twee oplossingen bruikbaar is. We krijgen daarom:

$$(5.2.1) \quad x_1 = \frac{1}{2}n - \frac{1}{2}\xi^1 \sqrt{n} - \frac{1}{2}.$$

Evenzo geldt:

$$(5.2.2) \quad x_r = \frac{1}{2}n + \frac{1}{2}\xi^1 \sqrt{n} + \frac{1}{2} \quad ^1).$$

¹⁾ De term $-\frac{1}{2}$, resp. $+\frac{1}{2}$ in (5.2.1) resp. (5.2.2) noemt men de continuïteitscorrectie. Als n groot is kunnen we deze termen zonder bezwaar verwaarlozen.

Tabel 5.I. Linker kritieke waarden van de tweezijdige tekentoets.

n	Onbetrouwbaarheidsdrempel				n	Onbetrouwbaarheidsdrempel			
	0,01	0,02	0,05	0,10		0,01	0,02	0,05	0,10
1	-	-	-	-	51	15	16	18	19
2	-	-	-	-	52	16	17	18	19
3	-	-	-	-	53	16	17	18	20
4	-	-	-	-	54	17	18	19	20
5	-	-	-	0	55	17	18	19	20
6	-	-	0	0	56	17	18	20	21
7	-	0	0	0	57	18	19	20	21
8	0	0	0	1	58	18	19	21	22
9	0	0	1	1	59	19	20	21	22
10	0	0	1	1	60	19	20	21	23
11	0	1	1	2	61	20	20	22	23
12	1	1	2	2	62	20	21	22	24
13	1	1	2	3	63	20	21	23	24
14	1	2	2	3	64	21	22	23	24
15	2	2	3	3	65	21	22	24	25
16	2	2	3	4	66	22	23	24	25
17	2	3	4	4	67	22	23	25	26
18	3	3	4	5	68	22	23	25	26
19	3	4	4	5	69	23	24	25	27
20	3	4	5	5	70	23	24	26	27
21	4	4	5	6	71	24	25	26	28
22	4	5	5	6	72	24	25	27	28
23	4	5	6	7	73	25	26	27	28
24	5	5	6	7	74	25	26	28	29
25	5	6	7	7	75	25	26	28	29
26	6	6	7	8	76	26	27	28	30
27	6	7	7	8	77	26	27	29	30
28	6	7	8	9	78	27	28	29	31
29	7	7	8	9	79	27	28	30	31
30	7	8	9	10	80	28	29	30	32
31	7	8	9	10	81	28	29	31	32
32	8	8	9	10	82	28	30	31	33
33	8	9	10	11	83	29	30	32	33
34	9	9	10	11	84	29	30	32	33
35	9	10	11	12	85	30	31	32	34
36	9	10	11	12	86	30	31	33	34
37	10	10	12	13	87	31	32	33	35
38	10	11	12	13	88	31	32	34	35
39	11	11	12	13	89	31	33	34	36
40	11	12	13	14	90	32	33	35	36
41	11	12	13	14	91	32	33	35	37
42	12	13	14	15	92	33	34	36	37
43	12	13	14	15	93	33	34	36	38
44	13	13	15	16	94	34	35	37	38
45	13	14	15	16	95	34	35	37	38
46	13	14	15	16	96	34	36	37	39
47	14	15	16	17	97	35	36	38	39
48	14	15	16	17	98	35	37	38	40
49	15	15	17	18	99	36	37	39	40
50	15	16	17	18	100	36	37	39	41

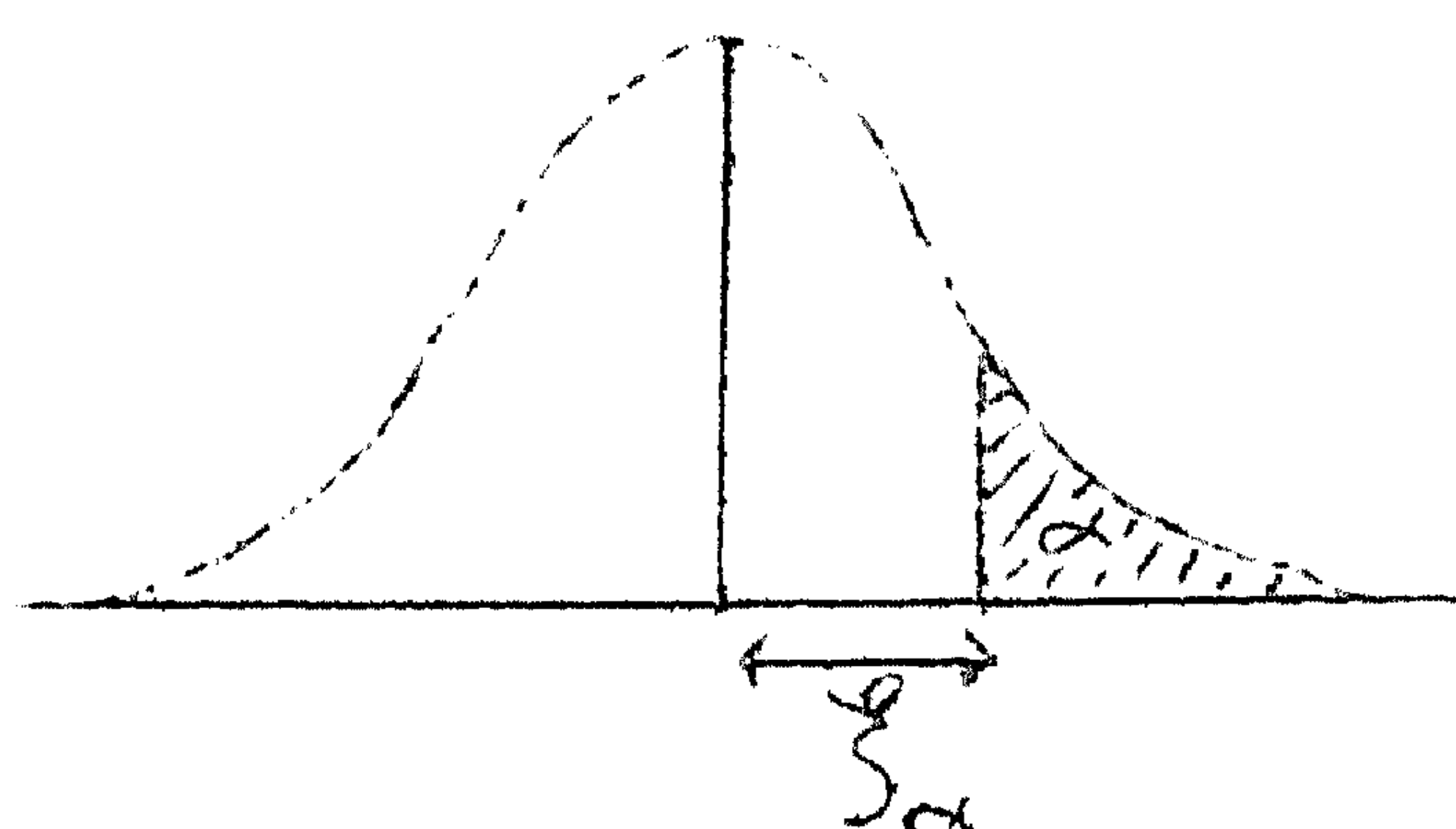
De waarden van χ^2 die men moet gebruiken vindt men wederom in tabel 4.III. Aangezien daar de éézijdige onbetrouwbaarheidsdrempels zijn gegeven, zullen we de waarden in de kolom α bij gebruik voor de tweezijdige tekentoets met twee moeten vermenigvuldigen. Met behulp van tabel 4.III kunnen we derhalve de kritieke waarden voor de tweezijdige tekentoets vinden voor grote n en bij de onbetrouwbaarheidsdrempels 0,01; 0,02; 0,05; 0,10 en 0,20.

Zoals wij boven zagen, is de onbetrouwbaarheidsdrempel een bovengrens voor de onbetrouwbaarheid, dus een bovengrens voor de kans dat de hypothese $p=\frac{1}{2}$ ten onrechte verworpen wordt. Nu wordt volgens de tweezijdige tekentoets deze hypothese verworpen als $\underline{x}$ ¹⁾ één der waarden x_1 of x_r of verder van $\frac{1}{2}n$ gelegen waarden aanneemt, dus als, vanuit het punt $\frac{1}{2}n$ gezien, de kritieke waarden bereikt of overschreden worden. Men noemt nu de kans dat $\underline{x}$ een zekere waarde x , of een verder van $\frac{1}{2}n$ gelegen waarde aanneemt, als $p=\frac{1}{2}$ is, de tweezijdige overschrijdingskans van x . Uit deze definitie volgt, dat bij toetsing met de onbetrouwbaarheidsdrempel α de tweezijdige overschrijdingskans van x_1 of x_r hoogstens α bedraagt, en van verder van $\frac{1}{2}n$ gelegen waarden nog minder. We kunnen dus zeggen, dat we de hypothese $p=\frac{1}{2}$ verwerpen als de bij de gevonden waarde van $\underline{x}$ behorende tweezijdige overschrijdingskans k niet groter is dan α . Als $k \leq \alpha$ is, dan zullen wij $p=\frac{1}{2}$ met des te meer stelligheid verwerpen naarmate k kleiner is. Om deze reden geeft men bij toepassing van een toets dikwijls niet alleen de conclusie, doch ook de overschrijdingskans van het gevonden resultaat. Met behulp van tabel 5.I kunnen we enige grenzen voor de overschrijdingskans vinden. Hebben we bijv. bij $n=40$ $x=13$ gevonden, dan weten we dat $0,02 < k \leq 0,05$ is. Immers, de kritieke waarde bij $\alpha = 0,02$ is 12, zodat $k > 0,02$ is, en de kritieke waarde bij $\alpha = 0,05$ is 13, zodat $k \leq 0,05$ is. Wegens de bovengenoemde symmetrie van de kritieke waarden van de tekentoets geldt hetzelfde voor $x = 40-13 = 27$ bij $n=40$. Verder kan k voor niet te kleine waarden van n benaderd worden met behulp van tabel 5.II.

¹⁾ Daar de letter x hier een stochastische grootheid voorstelt, is hij onderstreept.

Tabel 5.II. Eenzijdige overschrijdingskans van ξ_α voor een $N(0,1)$ verdeelde grootheid.

Waarden van $\alpha \cdot 10^4$ voor $\xi_\alpha = 0,00(0,01)3,49$



met $\frac{1}{\sqrt{2\pi}} \int_{\xi_\alpha}^{\infty} e^{-\frac{x^2}{2}} dx = \alpha$

ξ_α	0	1	2	3	4	5	6	7	8	9
0,0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0,7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1,5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1,6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1,8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1,9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2,0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2,2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2,3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2,4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2,7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2,8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2,9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3,0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3,1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3,2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
3,3	0005	0005	0005	0004	0004	0004	0004	0004	0004	0003
3,4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0002

Een 5 betekent, dat bij afronding naar het dichtstbijgelegen oneven cijfer afgerond moet worden. Een 5 wordt naar het dichtstbijzijnde even cijfer afgerond.

Men berekent dan de grootheid $\mathcal{Z}_\alpha$ uit

$$(5.2.3) \quad \mathcal{Z}_\alpha = \frac{|x - \frac{1}{2}n| - \frac{1}{2}}{\frac{1}{2}\sqrt{n}} \quad 1).$$

Vervolgens zoekt men in tabel 5.II de bijbehorende waarde van α op. De tweezijdige overschrijdingskans k is dan gelijk aan 2α .

In deze tabel heeft α dus niet de betekenis van een onbetrouwbaarheidsdrempel.

Hebben we bijv. in een steekproef van 60 gevonden $x=40$, dan is $\mathcal{Z}_\alpha = \frac{10 - \frac{1}{2}}{\frac{1}{2}\sqrt{60}} = 2,45$.

We vinden in tabel 5.II: $\alpha = 0,0071$, dus $k = 2\alpha = 0,014$. Deze uitkomst is in overeenstemming met tabel 5.I, waar we vinden dat $0,01 < k < 0,02$. (Men bedenke daarbij dat de overschrijdingskans van $x=40$ gelijk is aan die van $x=20$ bij deze waarde van n).

5.3. Eénzijdige tekentoets.

In 6.1 hebben we gezien dat de tweezijdige tekentoets een statistische methode is om een verantwoorde keuze te maken tussen de drie mogelijkheden

$$A: p < \frac{1}{2}$$

$$B: p = \frac{1}{2} \text{ (eigenlijk } p \approx \frac{1}{2})$$

$$C: p > \frac{1}{2}.$$

1)

$$|x - \frac{1}{2}n| = \begin{cases} x - \frac{1}{2}n & \text{als } x > \frac{1}{2}n \\ \frac{1}{2}n - x & \text{als } x \leq \frac{1}{2}n. \end{cases}$$

De term $-\frac{1}{2}$ in de teller van (5.2.3) heet de continuïteitscorrectie, en kan bij grote waarden van n verwaarloosd worden.

Het komt echter vaak voor, dat men slechts wenst te beslissen tussen (A of B) enerzijds en C anderzijds, of tussen A enerzijds en (B of C) anderzijds. In dergelijke gevallen is het efficiënter om, ook bij het opstellen van de toets, rekening te houden met het feit dat men slechts twee beslissingsmogelijkheden nodig heeft. Men dient dan een eenzijdige tekentoets toe te passen.

Er zijn twee éénzijdige tekentoetsen: de rechts-eenzijdige toets, waarbij $p \leq \frac{1}{2}$ getoetst wordt tegen $p > \frac{1}{2}$, en de links-eenzijdige toets, waarbij $p \geq \frac{1}{2}$ getoetst wordt tegen $p < \frac{1}{2}$. De beslissingsschemata voor deze twee toetsen zijn weergegeven in fig. 5.2 en fig. 5.3.

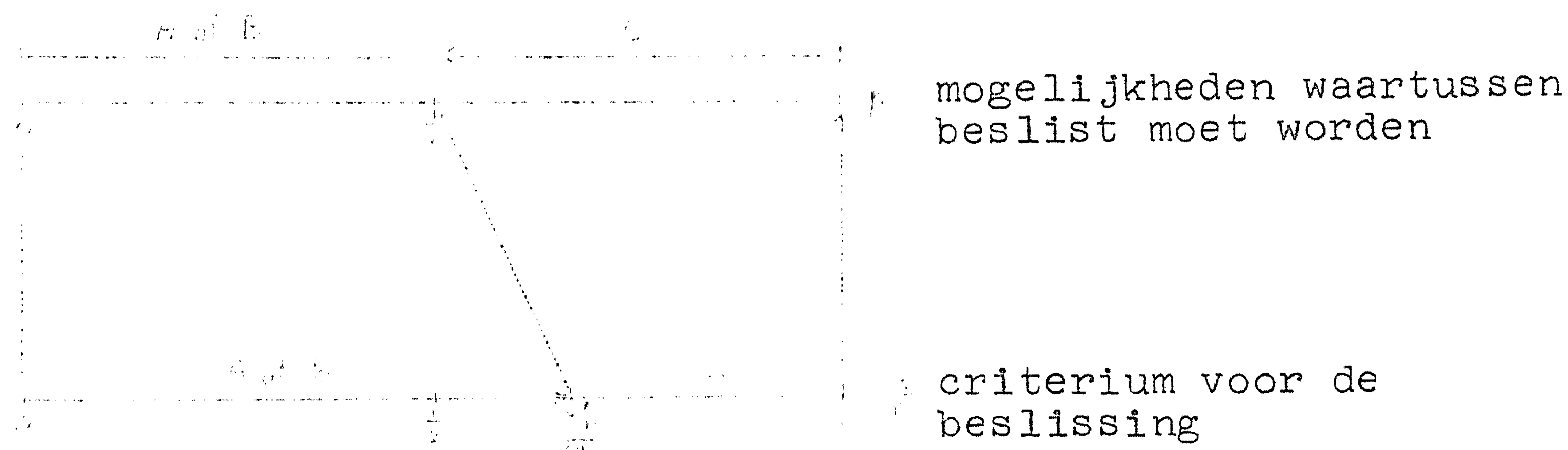


fig. 5.2. Beslissingsschema voor de rechtseenzijdige tekentoets.

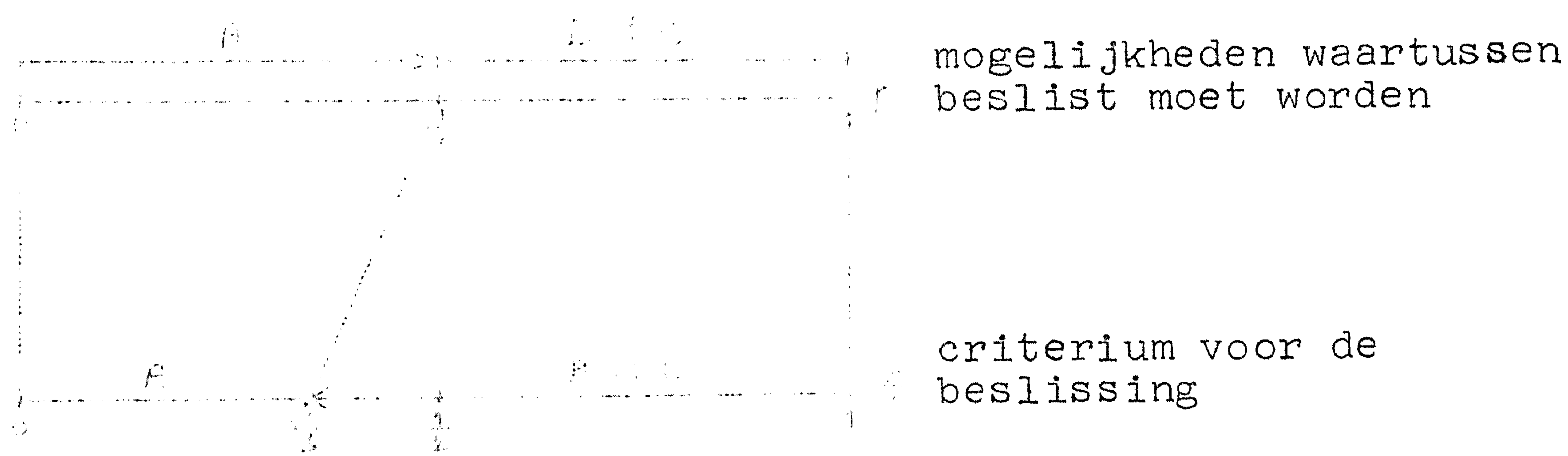


fig. 5.3. Beslissingsschema voor de links-eenzijdige tekentoets.

Bij de rechtseenzijdige toets kiezen we het interval C in het onderste gedeelte van fig. 5.2 zodanig, dat de kans dat $\frac{\bar{x}}{n}$ in dit interval valt, als $p = \frac{1}{2}$ is, hoogstens gelijk is aan de gekozen onbetrouwbaarheidsdrempel α . We bereiken daarmee dat deze kans als $p < \frac{1}{2}$ is zeker kleiner is dan α , daar de kans op veel "successen" bij $p < \frac{1}{2}$ natuurlijk kleiner is dan bij $p = \frac{1}{2}$. De kans dat de hypothese $p < \frac{1}{2}$ ten onrechte wordt verworpen is derhalve hoogstens α . We noemen het interval C nu de rechter kritieke zone en het getal x'_1 , de ondergrens van dit interval, de rechter kritieke waarde.

Geheel analoog is in het onderste deel van fig. 5.3 het interval A zo gekozen, dat bij $p \geq \frac{1}{2}$ de kans, dat $\frac{\bar{x}}{n}$ in A valt, hoogstens gelijk is aan α . A noemt men de linker kritieke zone en x'_r de linker kritieke waarde.

Zoals de tweezijdige tekentoets nauw verband houdt met de theorie der tweezijdig begrensde betrouwbaarheidsintervallen, zo houdt de links- resp. rechts-eenzijdige toets verband met de theorie der rechts- resp. links-eenzijdig begrensde betrouwbaarheidsintervallen.

Bij de links-eenzijdige tekentoets met onbetrouwbaarheidsdrempel α verwerpen we nl. de hypothese $p \geq \frac{1}{2}$ ten gunste van $p < \frac{1}{2}$ als we voor x een waarde vinden die zó klein is, dat de rechter betrouwbaarheidsgrens met dezelfde onbetrouwbaarheidsdrempel (éénzijdig!) kleiner dan $\frac{1}{2}$ is. Hieruit volgt, dat we x'_1 en x'_r voor grote waarden van n op precies dezelfde wijze uit de formules (4.3.1) en (4.3.2) kunnen vinden als x_1 en x_r (zie 5.2). We krijgen dus:

$$(5.3.1) \quad x'_1 = \frac{1}{2} \cdot n - \frac{1}{2} \cdot \xi^2 \cdot \sqrt{n} - \frac{1}{2}$$

en

$$(5.3.2) \quad x'_r = \frac{1}{2} \cdot n + \frac{1}{2} \cdot \xi \cdot \sqrt{n} + \frac{1}{2} \quad 1).$$

1) Ook hier kunnen de termen $+\frac{1}{2}$ en $-\frac{1}{2}$ bij grote n verwaarloosd worden.

De te gebruiken waarden van χ^2 vinden we weer in tabel 4.III, met dien verstande, dat de daar opgegeven waarden voor α nu niet met 2 vermenigvuldigd worden.

Voor $n \leq 100$ kunnen we x'_1 en x'_r uit tabel 5.I vinden op dezelfde wijze als x_1 en x_r , met dien verstande, dat de daar opgegeven waarden van α nu gehalveerd dienen te worden. Voor $n=60$ vinden we derhalve bij $\alpha = 0,025$ uit de tabel $x'_1=21$, dus $x'_r = 60-21 = 39$.

In ditzelfde geval volgt uit (5.3.1) en (5.3.2):

$$\begin{aligned} 30 - \frac{1}{2} \cdot 1,96 \cdot \sqrt{60 - \frac{1}{2}} &= 21,9 \text{ dus } x'_1 = 21 \\ \text{en } 30 + \frac{1}{2} \cdot 1,96 \cdot \sqrt{60 + \frac{1}{2}} &= 38,1 \text{ dus } x'_r = 39. \end{aligned}$$

Bij $n=60$ en $\alpha = 0,025$ is de overeenstemming tussen de (exacte) tabel en de benaderingsformules dus volkomen. Ook bij de eenzijdige toetsen werkt men vaak met het begrip overschrijdingskans. Deze wordt nu nader gespecificeerd als volgt: De rechts-eenzijdige overschrijdingskans $k_r(x)$ van x is per definitie de kans dat $\underline{x}$ de waarde x of een grotere waarde aanneemt, terwijl $p=\frac{1}{2}$ is. In formule geeft men deze kans aan als

$$k_r(x) = P \left\{ \underline{x} \geq x \mid p = \frac{1}{2} \right\}.$$

Als $x < \frac{1}{2}n$ is, dan is $k_r(x) > \frac{1}{2}$; naarmate x groter wordt neemt $k_r(x)$ af. De rechter kritieke waarde x'_r ligt zodanig, dat $k_r(x'_r)$ hoogstens α is, terwijl $k_r(x'_r - 1)$ groter dan α is. Men zal de hypothese $p \leq \frac{1}{2}$ met meer stelligheid verwerpen op grond van een waargenomen x , naarmate deze x groter is, d.w.z. naarmate de rechts-eenzijdige overschrijdingskans $k_r(x)$ kleiner is. Evenzo definieert men de links-eenzijdige overschrijdingskans $k_l(x)$ als de kans dat $\underline{x}$ de waarde x of een kleinere waarde aanneemt, onder de hypothese $p=\frac{1}{2}$. Voor $k_l(x)$ geldt nu:

$$k_1(x) = P \left\{ \underline{x} \leq x \mid p = \frac{1}{2} \right\}$$

$$k_1(x) \geq \frac{1}{2} \text{ als } x \geq \frac{1}{2}n$$

$$k_1(x'_1) \leq \alpha \text{ en } k_1(n'_1 + 1) > \alpha.$$

Ter illustratie zijn de één- en tweezijdige overschrijdingskansen $k_r(6)$, $k_l(6)$ en $k(6)$ van $x=6$ voor $n=8$ in fig. 5.4 weergegeven. Daartoe is voor ieder dezer drie gevallen de waarschijnlijkheidsverdeling van $\underline{x}$ onder de hypothese $p=\frac{1}{2}$ getekend in de vorm van een histogram.

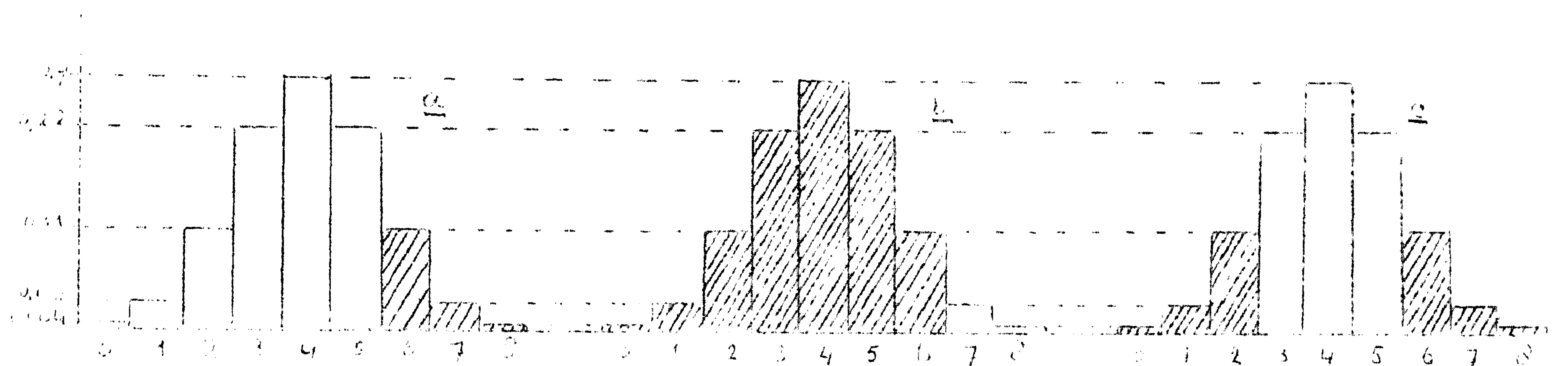


fig. 5.4. Overschrijdingskansen van $x=6$ bij de tekentoets ($n=8$), gegeven door de gearceerde oppervlakken in de figuren:

a rechts-eenzijdig

b links-eenzijdig

c tweezijdig.

We zien dat $k_r(6) = 0,004 + 0,03 + 0,11 = 0,14$

$$k_l(6) = 0,004 + 0,03 + 0,11 + 0,22 + 0,27 + 0,22 + 0,11 = 0,97$$

$$k(6) = 0,004 + 0,03 + 0,11 + 0,11 + 0,03 + 0,004 = 0,29.$$

Ook nu kunnen we uit tabel 5.I enige grenzen voor de eenzijdige overschrijdingskansen vinden. Zij bijv. $n=80$ en $x=31$; x ligt dan in de linker kritieke zone voor $\alpha = 0,05$ (eenzijdig!) maar niet voor $\alpha = 0,025$, dus $0,025 < k_l(31) \leq 0,05$.

Voorts volgt uit de tabel: $P \left\{ \underline{x} \leq 30 \mid p=\frac{1}{2} \right\} \leq 0,025$, zodat

$$P \{ \underline{x} \geq 31 | p = \frac{1}{2} \} = k_1(31) \geq 1 - 0,025 = 0,975.$$

Daar bovendien $P \{ \underline{x} \leq 30 | p = \frac{1}{2} \} > 0,01$ is, hebben we

$$P \{ \underline{x} \geq 31 | p = \frac{1}{2} \} = k_r(31) < 1 - 0,01 = 0,99.$$

We concluderen: $0,975 \leq k_r(31) < 0,99$ voor $n=80$.

Voor niet te kleine waarden van n kunnen we bovendien de eenzijdige overschrijdingskansen vinden met behulp van tabel 5.II. Om $k_r(x)$ te vinden berekent men dan

$$(5.3.3) \quad \mathcal{Z}'_{\alpha} = \frac{x - \frac{1}{2}n - \frac{1}{2}}{\frac{1}{2} \sqrt{n}} \quad 1)$$

en om k_1 te vinden

$$(5.3.4) \quad \mathcal{Z}''_{\alpha} = - \frac{x - \frac{1}{2}n + \frac{1}{2}}{\frac{1}{2} \sqrt{n}} \quad 1).$$

Als nu $\mathcal{Z}'_{\alpha}$ resp. $\mathcal{Z}''_{\alpha}$ positief is, dan is de rechts- resp. links-eenzijdige overschrijdingskans de bijbehorende waarde van α uit tabel 5.II. Als $\mathcal{Z}'_{\alpha}$ resp. $\mathcal{Z}''_{\alpha}$ negatief is, dan bepaalt men de waarde van α die volgens de tabel bij $-\mathcal{Z}'_{\alpha}$ resp. $-\mathcal{Z}''_{\alpha}$ behoort en dan is de rechts- resp. links-eenzijdige overschrijdingskans gelijk aan $1-\alpha$.

Bij $n=80$ en $x=31$ vindt men bijv.

$$\mathcal{Z}'_{\alpha} = \frac{31 - 40 - \frac{1}{2}}{\frac{1}{2} \sqrt{80}} = -2,06 \text{ en } \mathcal{Z}''_{\alpha} = - \frac{31 - 40 - \frac{1}{2}}{\frac{1}{2} \sqrt{80}} = + 1,90.$$

Daar $\mathcal{Z}''_{\alpha}$ positief is lezen we direct in de tabel af:

$k_1(31) = 0,0287$; $\mathcal{Z}'_{\alpha}$ is negatief. Bij $-\mathcal{Z}'_{\alpha} = 2,06$ behoort $\alpha = 0,0197$, zodat $k_r(31) = 1 - 0,0197 = 0,9803$.

Dit is in overeenstemming met de bovengevonden grenzen voor de overschrijdingskansen voor dit geval.

Uit de definities der drie soorten overschrijdingskansen volgt dat de tweezijdige overschrijdingskans van x tweemaal zo groot is als de kleinste der eenzijdige overschrijdingskansen. (Zie bijv. fig. 5.4). Dit heeft tot ge-

¹⁾ De term $-\frac{1}{2}$ resp. $+\frac{1}{2}$ in de teller van (5.3.3) resp. (5.3.4) noemt men ook hier de continuïteitscorrectie. Voor grote n kan men deze term zonder bezwaar weglaten.

volg dat, bij eenzijdige toetsing de betrokken kritieke waarde dichter bij $\frac{1}{2}n$ ligt dan ieder der twee kritieke waarden bij tweezijdige toetsing. Bij toepassing van bijv. de links-eenzijdige toets, dus bij toetsing van $p \geq \frac{1}{2}$ tegen $p < \frac{1}{2}$ komt men daardoor bij grotere waarden van x al tot de conclusie $p < \frac{1}{2}$, dan wanneer men de tweezijdige toets zou toepassen. Men formuleert dit voordeel gewoonlijk als volgt: De links-eenzijdige tekentoets heeft voor $p < \frac{1}{2}$ een groter onderscheidingsvermogen dan de tweezijdige. Eveneens heeft de rechts-eenzijdige tekentoets voor $p > \frac{1}{2}$ een groter onderscheidingsvermogen dan de tweezijdige.

5.4. Onbesliste experimenten.

In dit hoofdstuk hebben wij tot nu toe steeds aangenomen, dat de steekproef waarop we de tekentoets willen toepassen geheel uit "successen" en "mislukkingen" bestaat, m.a.w. dat geen onbesliste experimenten voorkomen. In werkelijkheid zullen echter bijna altijd wel enkele onbesliste experimenten in de steekproef aanwezig zijn, al was het alleen maar tengevolge van meet-onnauwkeurigheden. Als het aantal dezer onbesliste experimenten relatief klein is, kunnen we ze het beste weglaten en daarbij n dienovereenkomstig kleiner nemen. Deze methode is altijd beter dan die, waarbij men de helft van de onbesliste experimenten positief (succes) en de andere helft negatief (mislukking) rekent, een methode die ook wel eens wordt voorgesteld. Bij deze laatste methode is de kans, dat men de getoetste hypothese verwierpt, terwijl deze hypothese niet juist is, kleiner dan bij de eerste methode. Volgens de eerder ingevoerde terminologie kunnen we dus zeggen dat de eerste methode een groter onderscheidingsvermogen heeft dan de laatste.

Als het aantal onbesliste experimenten relatief groot is, dan kan men de tekentoets niet meer toepassen zonder wijzigingen, waarvan de behandeling buiten het bestek van deze cursus valt.

5.5. Voorbeelden.

a. In 5.1 werd gesproken over de bepaling van de doorslag-

spanning van oliemonsters. Hier zullen we iets dieper op dit probleem ingaan.¹⁾ Bij de bepaling van de doorslagspanning gaat men als volgt te werk: Een glazen buisje, aan de beide uiteinden voorzien van een electrode, wordt met olie van de te onderzoeken soort gevuld, waarna men tussen de twee elektroden een elektrische spanning aanlegt. Deze spanning voert men geleidelijk op, totdat doorslag optreedt. De spanning, die op dat moment bereikt was, noemt men dan de doorslagspanning. Men kan deze bepaling van de doorslagspanning met dezelfde olie-vulling herhalen zo vaak men wil. Wij zullen speciaal het geval beschouwen dat men de bepaling twee maal doet bij één vulling. Het is nu van belang om te weten of de tweede bepaling systematisch een hogere of lagere waarde voor de doorslagspanning geeft - wat het gevolg zou kunnen zijn van mogelijke veranderingen in de olie-vulling door de doorslag bij de eerste bepaling - of niet. Om dit te onderzoeken werden 28 vullingen gebruikt. Van ieder dezer vullingen werd op de bovenomschreven wijze twee maal de doorslagspanning bepaald. De bij deze paren van bepalingen gevonden verschillen tussen de tweede en eerste uitkomst zijn in tabel 5.III gegeven. We zullen nu de hypothese toetsen, dat de kans op een positief verschil tussen de doorslagspanningen van de 2e en 1e bepaling bij dezelfde olievulling gelijk is aan de kans dat dit verschil negatief uitvalt. We houden daarbij geen rekening met de mogelijkheid dat we een verschil 0 kunnen vinden, waardoor de te toetsen hypothese luidt $p=\frac{1}{2}$, als we de kans op een positief verschil p noemen. (Als we maar nauwkeurig genoeg meten, zal een verschil 0 inderdaad praktisch niet optreden, zelfs als $p=\frac{1}{2}$.) We passen dus de tweezijdige tekentoets toe. Van de 28 verschillen zijn er 3 gelijk aan nul,

¹⁾ Dit voorbeeld is ontleend aan W.J. YODEN and J.M. CAMERON, Use of statistics to determine precision of test methods, Symposium on Application of Statistics, Am.Soc. for Testing Materials, Philadelphia 1950, pp. 27-34.

5 negatief en 20 positief. Na weglaten van de "nullen" hebben we dus: $n=25$, $x=20$. Uit tabel 5.I volgt, dat dit resultaat een tweezijdige overschrijdingskans van hoogstens 0,01 bezit. We verwerpen de hypothese daarom met onbetrouwbaarheidsdrempel 0,01. Daar we overwegend positieve verschillen vonden concluderen we $p > \frac{1}{2}$, d.w.z.: De uitkomst van de tweede bepaling ligt systematisch hoger dan de eerste. Met een eventuele verklaring van dit verschijnsel houden wij ons uiteraard niet op; dit is een taak voor physici of chemici, niet voor statistici.

Tabel 5.III Verschillen (in K.V.) van waargenomen doorslagspanningen.

olie-vulling no.	verschil tussen 2e en 1e gevonden doorslag- spanning	olie-vulling no.	verschil tussen 2e en 1e gevonden doorslag- spanning
1	3	15	-1
2	2	16	1
3	4	17	2
4	4	18	5
5	1	19	-2
6	5	20	1
7	6	21	3
8	0	22	0
9	-1	23	7
10	1	24	5
11	7	25	1
12	-1	26	-1
13	1	27	0
14	1	28	8

b. In 1930 werden in Amsterdam 6848 jongens en 6374 meisjes geboren. Noemen we de kans, dat een geboorte een jongensgeboorte is, p , dan kunnen we de hypothese $p=\frac{1}{2}$ toetsen: We hebben een steekproef met $n = 6848 + 6374 = 13222$ en $x = 6848$.

Formule (5.2.3) geeft
$$z_{\alpha} = \frac{6848 - 6611 - \frac{1}{2}}{\frac{1}{2} \sqrt{13222}} = \frac{236,5}{57,5} = 4,1.$$

Deze waarde van ξ_α is zo groot, dat zij niet meer in tabel 5.II opgenomen is. De bijbehorende tweezijdige overschrijdingskans is derhalve kleiner dan de tweezijdige overschrijdingskans bij elke ξ_α die wel in tabel 5.II is opgenomen, zodat we vinden: $k < 0,0004$, zijnde kleiner dan alle gebruikelijke onbetrouwbaarheidsdrempels. Op grond van dit resultaat kunnen we dus met grote stelligheid concluderen dat $p \neq \frac{1}{2}$, en wel dat $p > \frac{1}{2}$, d.w.z. dat de kans op een jongensgeboorte groter is dan die op een meisjesgeboorte.

- c. In het bijzonder in Engeland, waar men de thee met veel melk drinkt, treft men mensen aan die beweren te kunnen proeven of de melk dan wel de thee het eerst in het kopje gedaan is. Het probleem is nu hoe een dergelijke bewering statistisch is te toetsen. Wij zullen dit voorbeeld, dat onder de naam "Lady tasting tea" welhaast klassiek geworden is¹⁾, gebruiken om te laten zien hoe men zich hier van de tekentoets bedienen kan. Als proefopzet zou men bijv. 20 kopjes kunnen nemen. In 10 van deze kopjes doet men eerst de melk, bij de andere 10 eerst de thee. Vervolgens laat men de theeproevende dame zeggen wat er volgens haar het eerst ingedaan is. Indien haar bewering, dat zij dit kan proeven, onjuist is, dan raadt zij er dus naar. Bij dit raden kan het echter zijn dat zij geneigd is - om welke reden dan ook - vaker te zeggen, dat bijv. de thee er het eerst ingedaan is. In dat geval is de kans dat zij het juiste antwoord geeft bij het proeven uit een kopje waar inderdaad eerst de thee ingedaan is, groter dan $\frac{1}{2}$, en bij het proeven uit een kopje waar de melk het eerst ingedaan is, kleiner dan $\frac{1}{2}$. We kunnen daarom de tekentoets niet toepassen, want daarvoor is nodig dat bij elk kopje de kans op een juist antwoord $\frac{1}{2}$ is, aangenomen dat de bewering van de dame onjuist is. Door middel van een kunstgreep kunnen we dit echter

¹⁾ Zie bijv.: J. NEYMAN, First course in probability and statistics, N.Y. 1950.
R.A. FISHER, The design of experiments, Edinburgh, 1925.

bewerkstelligen. Daartoe loten we bij elk der 20 kopjes, of we er eerst de thee dan wel eerst de melk indoen, en wel zodanig, dat voor elk kopje de kans dat we er eerst de thee indoen gelijk is aan de kans dat we er eerst de melk indoen. We kunnen dit bereiken door bijv. bij elk kopje met een zuivere munt te werpen en bij "kruis" eerst de melk, bij "munt" eerst de thee in het kopje te doen. Een andere methode is deze: Uit een tabel van aselechte getallen nemen we de eerste 20 cijfers. Is het eerste even, dan doen we in kopje no. 1 eerst de melk, is het eerste oneven, dan eerst de thee, en evenzo geeft het 2e aselechte getal, aan hoe te handelen met kopje no. 2 etc. etc.

Nu is, als de dame haar antwoorden door raden "bepaalt", bij elk kopje de kans op het juiste antwoord $\frac{1}{2}$, zodat nu wel de tekentoets toegepast kan worden. Voeren we het experiment uit, dan krijgen we dus 20 antwoorden. Het aantal juiste antwoorden hieronder noemen we x. We zullen nu de rechter-éénzijdige tekentoets gebruiken. Immers, alleen als de dame relatief veel juiste antwoorden geeft, d.w.z. bij grote x, zullen we kunnen besluiten dat haar bewering juist is. (Anders is het, als we na willen gaan of zij verschil proeft. Het is nl. ook denkbaar dat ze, door verkeerde interpretatie van het smaakverschil, veel onjuiste antwoorden geeft, zodat dan de tweezijdige toets gebruikt moet worden.)

In een concreet geval werden 23 kopjes gebruikt. De loting werd met behulp van aselechte getallen op de bovenbeschreven wijze uitgevoerd. De eerste 23 cijfers op blz. VI van TIPPETT's tabel van aselechte getallen¹⁾ zijn: 1 3 7 0 4 3 6 9 3 8 8 1 8 7 4 2 1 2 2 0 4 1 1, in deze volgorde. De proefopzet die hieruit volgde en de resultaten zijn weergegeven in tabel 5.IV.

¹⁾ Random Sampling Numbers, arranged by L.H.C. TIPPETT, Tracts for Computers No. XV, Cambridge University Press, 1927.

Tabel 5.IV. Proefopzet en resultaten "Lady Tasting Tea".

Kopje No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Wijze van vullen	T	T	T	M	M	T	M	T	T	M	M	T	M	T	M	M	T	M	M	M	M	T	T
Ant- woord	T	M	T	T	M	T	M	T	M	M	M	T	M	M	M	T	M	M	M	M	M	T	T
Goed(+) of fout(-)	+	-	+	-	+	+	+	+	-	+	+	+	+	-	+	-	-	+	+	+	+	+	+

In deze tabel betekent de letter T op de tweede regel, dat eerst de thee in het kopje is gedaan, en op de derde regel, dat de dame denkt dat de thee het eerst in het kopje gedaan is. De betekenis van de letter M is analoog.

We zien dat 17 van de 23 antwoorden juist waren. Volgens tabel 5.I heeft dit resultaat een rechts-éénzijdige overschrijdingskans van hoogstens 0,025. Met onbetrouwbaarheidsdrempel 0,025 kunnen we dus concluderen dat de bewering van de dame, dat wij kan proeven of de melk dan wel de thee eerst in het kopje is gedaan, juist is.

Hoofdstuk 6.

Binomiale toetsen.

6.1. Probleemstelling.

In het vorige hoofdstuk hebben wij gezien hoe men de hypothesen kan toetsen, dat de fractie p der elementen met een bepaald kenmerk in een populatie gelijk is aan $\frac{1}{2}$, hoogstens gelijk is aan $\frac{1}{2}$ of minstens gelijk is aan $\frac{1}{2}$, respectievelijk met behulp van de tweezijdige, de rechts-eenzijdige en de links-eenzijdige tekenttoets. Deze methoden kunnen worden toegepast als men uit de populatie een steekproef, bestaande uit een aantal onafhankelijke aselechte trekkingen, heeft genomen en het aantal x der objecten met het beschouwde kenmerk in deze steekproef bepaald heeft. Men kan echter ook, met behulp van andere methoden, die in dit hoofdstuk beschreven worden, hypothesen van de vorm $p=p_0$, $p \geq p_0$ of $p \leq p_0$ toetsen. Hierbij mag het getal p_0 elke waarde tussen 0 en 1 hebben. Men noemt deze methoden binomiale toetsen, omdat, onder de te toetsen hypothese de toetsingsgrootte $\underline{x}$ een zogenaamde binomiale waarschijnlijkheidsverdeling heeft. De tekenttoets is een bijzonder geval van een binomiale toets, nl. de binomiale toets met $p_0 = \frac{1}{2}$.

6.2. De tweezijdige binomiale toets.

De methode die wij in 5.1. beschreven hebben om de hypothese $p=\frac{1}{2}$ tweezijdig te toetsen, kan geheel analoog gebruikt worden om een hypothese $p=p_0$ ($\neq \frac{1}{2}$) te toetsen, waarbij p_0 een getal tussen 0 en 1 is. Men hoeft daartoe slechts beide betrouwbaarheidsgrenzen voor p op grond van het gegeven waarnemingsmateriaal te bepalen. Als p_0 tussen deze grenzen ligt zal men de hypothese $p=p_0$ niet verwerpen; als geen der twee grenzen groter dan p_0 is zal men $p=p_0$ verwerpen ten gunste van $p < p_0$, en als geen der twee grenzen kleiner is dan p_0 zal men $p=p_0$ ten gunste van $p > p_0$ verwerpen.

Voor waarden van p_0 die niet te veel van $\frac{1}{2}$ afwijken en niet te kleine n kan men wederom de kritieke waarden benade-

ren door x op te lossen uit

$$\frac{x + \frac{1}{2} + \frac{1}{2}\zeta^2 + \zeta \sqrt{\frac{(x+\frac{1}{2})(n-x-\frac{1}{2})}{n}} + \frac{1}{4}\zeta^2}{n + \zeta^2} = p_0.$$

Men vindt dan:

$$(6.2.1) \quad x_l = p_0 n - \zeta \sqrt{n \cdot p_0 (1-p_0)} - \frac{1}{2}$$

en

$$(6.2.2) \quad x_r = p_0 n + \zeta \sqrt{n \cdot p_0 (1-p_0)} + \frac{1}{2}.$$

De te gebruiken waarden van ζ^2 bij verschillende onbetrouwbaarheidsdrempels vindt men in tabel 4.III. Deze tabel geeft eenzijdige onbetrouwbaarheidsdrempels, zodat de daar vermelde waarden van α met twee vermenigvuldigd moeten worden.

Eveneens als p_0 niet te veel van $\frac{1}{2}$ afwijkt en grote waarden van n kan men met behulp van tabel 5.II de tweezijdige overschrijdingskans van x benaderen. Daartoe berekent men

$$(6.2.3) \quad \zeta_\alpha = \frac{|x - p_0 n| - \frac{1}{2}}{\sqrt{n p_0 (1 - p_0)}}$$

en zoekt de bijbehorende waarde van α op in de genoemde tabel. De tweezijdige overschrijdingskans k is dan gelijk aan 2α . Wij vestigen er hier nogmaals de aandacht op, dat α in deze tabel geen onbetrouwbaarheidsdrempel voorstelt.

6.3. Eénzijdige binomiale toetsen.

Geheel analoog aan het in 5.3 beschrevene, zullen we de hypothese $p=p_0$ tweezijdig toetsen als we willen beslissen tussen de mogelijkheden $p < p_0$, $p=p_0$ en $p > p_0$. Als we daarentegen $p \leq p_0$ willen toetsen tegen $p > p_0$, zullen we een rechts-eénzijdige toets prefereren en voor het geval $p \geq p_0$ tegen $p < p_0$ een links-eénzijdige toets. Kritieke waarden en overschrijdingskansen worden op analoge wijze gevonden. De kritieke waarde x_l' van de links-eénzijdige toets is dus gelijk aan het rechterlid van (6.2.1) en de kritieke waarde x_r' van de rechts-eénzijdige toets is gelijk aan het rechter-

lid van (6.2.2). De onbetrouwbaarheidsdrempels zijn nu echter gelijk aan de in tabel 4.III gegeven waarden. De eenzijdige overschrijdingskansen kan men weer met behulp van tabel 5.II benaderen, zoals beschreven is in 5.3. Men moet echter $\mathfrak{z}'_{\alpha}$ en $\mathfrak{z}''_{\alpha}$ niet berekenen uit de formules (5.3.3) en (5.3.4), doch uit

$$(6.3.1) \quad \mathfrak{z}'_{\alpha} = \frac{x - p_0 n - \frac{1}{2}}{\sqrt{n \cdot p_0 (1-p_0)}}$$

en

$$(6.3.2) \quad \mathfrak{z}''_{\alpha} = - \frac{x - p_0 n + \frac{1}{2}}{\sqrt{n \cdot p_0 (1-p_0)}} .$$

Zowel voor de bepaling der kritieke waarden als voor de bepaling van de eenzijdige overschrijdingskansen met behulp van tabel 5.II mag p_0 niet teveel van $\frac{1}{2}$ afwijken en n niet te klein zijn. Is dit wel het geval, dan moeten we teruggrijpen naar exact bepaalde betrouwbaarheidsgrenzen of althans naar benaderende methoden die in dat geval wel goede resultaten geven. Hiervoor verwijzen we naar hoofdstuk 4.

6.4. Onbesliste experimenten.

Ook nu zullen we van toetsen moeten afzien als een groot deel van de steekproef uit onbesliste experimenten bestaat. Het enige wat we, met de ons ten dienste staande statistische methoden kunnen doen, is in dat geval het nemen van een nieuwe steekproef, en daarbij de waarnemingsnauwkeurigheid op te voeren. Dit zal in het algemeen een vermindering van het aantal onbesliste experimenten teweeg brengen, daar men met gevoeliger meetmethoden een klein positief of negatief verschil eerder als zodanig herkent.

Komen er in de steekproef relatief weinig onbesliste experimenten voor, dan laat men deze weg, en men past vervolgens de te gebruiken toets toe op de overblijvende (kleinere) steekproef.

6.5. Voorbeelden.

- a. Bij zijn klassieke experimenten verkreeg MENDEL¹⁾ van 10 planten 355 gele en 123 groene erwten. Op grond van zijn erfelijkheidstheorie verwachtte hij gele en groene erwten in de verhouding 3:1. Zijn de door hem gevonden aantallen hiermee in overeenstemming?

Volgens Mendel's theorie is de kans op een erwt met het kenmerk "geel" gelijk aan $\frac{3}{4}$. We zullen dus de hypothese $p = \frac{3}{4}$ toetsen, en wel tweezijdig, want zowel bij teveel als bij te weinig gele erwten in de steekproef zullen we de hypothese willen verwerpen. Tot onze beschikking staat een steekproef met $n = 355 + 123 = 478$ en $x = 355$. We berekenen de kritieke waarden bij $\alpha = 0,05$ uit de formules (6.2.1 en 6.2.2):

$$x_l = \frac{3}{4} \cdot 478 - 1,96 \cdot \sqrt{478 \cdot \frac{3}{4} \cdot \frac{1}{4}} - \frac{1}{2} = 339,4$$

$$x_r = \frac{3}{4} \cdot 478 + 1,96 \cdot \sqrt{478 \cdot \frac{3}{4} \cdot \frac{1}{4}} + \frac{1}{2} = 377,6.$$

Aangezien $339,4 < 355 < 377,6$ is, kunnen wij de getoetste hypothese niet verwerpen. Uit formule (6.2.3) vinden we voorts:

$$\chi_\alpha = \frac{|355 - \frac{3}{4} \cdot 478| - \frac{1}{2}}{\sqrt{478 \cdot \frac{3}{4} \cdot \frac{1}{4}}} = 0,32.$$

In tabel 5.II vinden we bij $\chi_\alpha = 0,32$ de waarde 0,375 voor α , zodat de tweezijdige overschrijdingskans van 355 gelijk aan 0,75 is. We concluderen op grond hiervan dat de gevonden aantallen groene en gele erwten in goede overeenstemming zijn met de theorie.

- b. Bij een cursus statistiek wenste men ter demonstratie enige steekproefexperimenten uit te voeren. Daartoe werd een vaas gevuld met erwten, waarvan een aantal rood geverfd was. Na telling bleken er 1201 erwten, waaronder 94 rode, in de vaas te zitten. Men nam nu steekproeven

¹⁾ G. MENDEL, Versuche mit Pflanzenhybriden, Verhandlungen Naturforsch. Ver. Brünn, 4(1865).

van 25 stuks, door de vaas in de hand te nemen, goed te schudden, en vervolgens met een plankje, voorzien van 25 uithollingen ter grootte van één erwt, 25 erwten uit de vaas te scheppen. Op deze wijze werden 100 steekproeven van 25 verkregen. In tabel 6.I zijn de resultaten, afgezien van de volgorde, gegeven.

Tabel 6.I. Aantallen rode erwten in 100 steekproeven van 25.

aantal rode erwten	frequentie	aantal rode erwten	frequentie
0	2	4	18
1	21	5	3
2	28	6	4
3	24	7	0

In deze tabel staat in de eerste kolom het aantal rode erwten, en in de tweede kolom het aantal steekproeven waarin dit aantal rode erwten gevonden is. In totaal werden dus $0.2 + 21.1 + 28.2 + 24.3 + 18.4 + 3.5 + 4.6 = 260$ rode erwten gevonden op de 2500 erwten, d.w.z. een fractie van $0,104$. Aangezien in de vaas slechts een fractie $\frac{94}{1201} = 0,0783$ rode erwten voorkomt lijkt dit aantal rode erwten vrij hoog. Er zijn 8 onder de 100 steekproeven, nl. de steekproeven met 5 of 6 rode erwten, die aanleiding geven tot een onderste betrouwbaarheidsgrens met onbetrouwbaarheidsdrempel $0,05$ die groter is dan de werkelijke waarde van $p = 0,0783$. Immers voor $x=5$ resp. $x=6$ bij $n=25$ en $\alpha = 0,05$ éénzijdig geeft tabel 4.I als onderste betrouwbaarheidsgrens $0,08$ resp. $0,11$. Er zijn dus 8 steekproeven van de 100 die leiden tot een foutieve betrouwbaarheidsgrens met $\alpha = 0,05$. Ook dit lijkt wat veel, daar we hoogstens ongeveer 5 steekproeven op de 100 verwachten die tot een dergelijke foutieve conclusie voeren. Dit steekproef-experiment doet derhalve het vermoeden rijzen, dat de trekkingswijze niet aselekt is. Om dit na te gaan mogen we niet een toets toepassen op de gegevens van tabel 6.I, aangezien ons vermoeden hierop juist gebaseerd

is. We hebben daarom dezelfde persoon nogmaals 100 steekproeven van 25 laten trekken op dezelfde wijze als tevoren. Onder de 2500 zo verkregen (en weer teruggelegde) erwten bevonden zich 229 rode erwten. Formule (6.3.1) geeft nu:

$$z'_\alpha = \frac{229 - 0,0783 \cdot 2500 - \frac{1}{2}}{\sqrt{2500 \cdot 0,0783 \cdot 0,9217}} = \frac{32,75}{13,43} = 2,44.$$

In tabel 5.II lezen we nu af, dat de rechtseenzijdige overschrijdingskans van $x=229$ gelijk aan 0,0073 is. Zelfs met onbetrouwbaarheidsdrempel 0,01 zouden we de hypothese $p \leq 0,0783$ dus moeten verwerpen ten gunste van $p > 0,0783$. Dit wijst er dus op, dat de trekkingswijze niet aselekt is, nl. dat om de een of andere reden systematisch teveel rode erwten getrokken worden. De volgende stap is nu het opsporen van de oorzaak van dit verschijnsel. Deze werd in dit geval inderdaad gevonden: De 94 rood geverfde erwten waren gemiddeld groter dan de andere erwten. Bij het schudden zakken de kleine erwten gemakkelijker naar de bodem dan de grote. Het gevolg hiervan is, dat men, vooral als men de pot in de hand en dus onwillekeurig wat scheef houdt, systematisch teveel grote, en dus ook systematisch teveel rode erwten vindt.

Hoofdstuk 7.

De twee-steekproeven-toets van Wilcoxon.

7.1. Een voorbeeld.

a. Echtparen.

Een onderzoeker heeft de lengten gemeten van de mannen en vrouwen van 20 aselekt gekozen echtparen, met de bedoeling om na te gaan of gehuwde mannen in het algemeen groter zijn dan hun vrouwen. De (fictieve) resultaten van dit (fictieve) onderzoek zijn samengevat in tabel 7.I.

Tabel 7.I. Lengten in meters van de mannen en vrouwen van 20 echtparen.

Nummer van het echtpaar	x_i = lengte van de man	y_i = lengte van de vrouw	teken van $x_i - y_i$
1	1,592	1,633	-
2	1,622	1,663	-
3	1,639	1,531	+
4	1,677	1,612	+
5	1,681	1,557	+
6	1,711	1,718	-
7	1,717	1,645	+
8	1,721	1,741	-
9	1,725	1,675	+
10	1,727	1,631	+
11	1,732	1,647	+
12	1,744	1,615	+
13	1,745	1,657	+
14	1,749	1,649	+
15	1,765	1,674	+
16	1,786	1,692	+
17	1,813	1,742	+
18	1,820	1,828	-
19	1,821	1,743	+
20	1,878	1,890	-

(In de praktijk heeft het uiteraard geen zin lichaamslengten nauwkeuriger dan in hele centimeters op te geven. In dit fictieve geval is dit wel gedaan, om te vermijden dat er gelijke waarnemingen optreden, die tot later te behandelen complicaties aanleiding geven.)

We beschouwen nu de lengte van de man van het i^e

echtpaar als een waarneming x_i van een stochastische grootheid $\underline{x}$, en de lengte van zijn vrouw als een waarneming y_i van een stochastische grootheid $\underline{y}$. We eisen daarbij niet, dat de verschillende $\underline{x}_i$ dezelfde verdeling hebben; we laten dus bijv. toe, dat de verschillende echtparen verschillende bevolkingsgroepen met verschillende lengteverdelingen representeren. De hypothese dat er geen systematisch verschil bestaat tussen de lengten van de mannen en hun vrouwen kunnen we nu als volgt interpreteren:

Voor iedere i zijn $\underline{x}_i$ en $\underline{y}_i$ onafhankelijk en geldt: $P \{ \underline{x}_i > \underline{y}_i \} = P \{ \underline{x}_i < \underline{y}_i \} = \frac{1}{2}$. Deze hypothese kunnen we toetsen met behulp van de tekentoets. We behoeven daartoe slechts het aantal echtparen te tellen met het kenmerk $x_i > y_i$ (de niet-onderstreepte letters x_i en y_i stellen de werkelijk gemeten lengten voor). We vinden in tabel 7.I dat dit aantal 14 is. Daar de kritieke waarden van de tweezijdige tekentoets bij $\alpha = 0,10$ en $n = 20$ 5 en 15 zijn, kunnen we de hypothese, dat er geen systematisch verschil bestaat, niet verwerpen.

b. Mannen en vrouwen.

Wij beschouwen nogmaals tabel 7.I, doch nemen nu aan, dat de 2e kolom de lengten van een twintigtal mannen, aselekt gekozen uit een bepaalde bevolkingsgroep, en de 3e kolom de lengten van een twintigtal vrouwen, eveneens aselekt gekozen uit dezelfde bevolkingsgroep, bevat. De twee lengten op één regel hebben dus in de regel niet meer op één echtpaar betrekking.

We willen nu nagaan of de mannen van de onderzochte bevolkingsgroep in doorsnee groter zijn dan de vrouwen.

We kunnen nu de lengten in de 2e kolom van tabel 7.I als onafhankelijke waarnemingen $x_1, x_2, \dots, x_{20}$ van één stochastische grootheid $\underline{x}$ (de lengte van mannen van de beschouwde bevolkingsgroep) en de lengten $y_1, y_2, \dots, y_{20}$ in de 3e kolom als onafhankelijke waarnemingen van één stochastische grootheid $\underline{y}$ (de lengte van vrouwen van dezelfde bevolkingsgroep) beschouwen, terwijl de x -waarnemin-

gen onafhankelijk zijn van de y -waarnemingen. We willen nu de hypothese toetsen, dat $\underline{x}$ en $\underline{y}$ dezelfde waarschijnlijkheidsverdeling bezitten. Deze hypothese zullen we in het volgende aanduiden als H_0 .

H_0 houdt dus in dat $x_1, x_2, \dots, x_{20}$ en $y_1, y_2, \dots, y_{20}$ onafhankelijke waarnemingen zijn van één stochastische grootheid $\underline{x}$. In gedachten plaatsen we nu de gemeten mannen en vrouwen in een rij, te beginnen met de kortste, daarna de op één na kortste, etc. Indien H_0 juist is, is er geen enkele reden om aan te nemen, dat er voorkeur zou zijn voor een bepaalde volgorde van de mannen en vrouwen in deze rij; men kan dan ook bewijzen dat onder de hypothese H_0 alle mogelijke volgorden even waarschijnlijk zijn. Deze consequentie van H_0 , die we verder H'_0 zullen noemen, zullen we in feite gaan toetsen.

Zou men op het waarnemingsmateriaal de tekentoets willen toepassen, evenals in a, dan zou men daartoe paren van waarnemingen moeten vormen. In a was er één bepaalde wijze gegeven waarop dit moest: De lengte van een man moest vergeleken worden met de lengte van zijn vrouw. Hier is niet een dergelijke indeling in paren gegeven. Er zijn vele verschillende indelingen in paren mogelijk, en geen dezer indelingen verdient enige voorkeur boven de andere, terwijl het resultaat van de tekentoets - toegepast op een indeling in 20 paren - afhangt van de keuze van de indeling. Dit kunnen we gemakkelijk inzien: Passen we de tekentoets toe op de verschillen der lengten van mannen en vrouwen, die in tabel 7.I op dezelfde regel voorkomen, dan vinden we weer $k > 0,10$. Verwisselen we echter $y_8 = 1,741$ en $y_{12} = 1,615$, dan neemt het aantal "plussen" met één toe, zodat we krijgen $k \leq 0,05$.

Wij kunnen deze moeilijkheid oplossen, door de lengte van iedere man te vergelijken met de lengte van iedere vrouw. We hebben nu in totaal 400 paren van waarnemingen; deze zijn uiteraard niet onafhankelijk, zodat we niet de tekentoets met $n = 400$ op de 400 verschillen mogen toepassen. We tellen nu het aantal malen onder deze 400 vergelijkingen tussen een man en een vrouw, dat de man

groter is dan de vrouw. Een andere procedure, die op hetzelfde neerkomt, is de volgende: We rangschikken de mannen en vrouwen in gedachten naar opklimmende grootte (zoals boven aangegeven), en tellen bij iedere man het aantal vrouwen dat vóór hem in de rij staat, en nemen vervolgens de som van deze aantallen. Het resultaat, dat meestal met de letter U wordt aangeduid, is in ons voorbeeld 289; bij de overige 111 vergelijkingen is de vrouw groter dan de man, terwijl er hier geen gevallen voorkomen, waarbij beiden even groot zijn. Wij zullen nu als toetsingsgrootte gebruiken het dubbele van U, dat we aanduiden met W:

$$(7.1.1) \quad W = 2 U.$$

In het volgende zullen we zien hoe wij de overschrijdingskansen van W onder de hypothese H_0 kunnen vinden, en welke conclusies daaruit getrokken kunnen worden.

De statistische methode, die op de bovenomschreven handelwijze berust, noemt men de toets van Wilcoxon, naar de Amerikaanse statisticus F. WILCOXON¹⁾, die deze methode in 1945 introduceerde. De methode werd nader uitgewerkt door H.B. MANN en D.R. WHITNEY²⁾ in 1947. Een op de praktijk gerichte handleiding van de toets is die van Ir Doralien Wabeke en Dr Constance van Eeden³⁾.

7.2. De toetsingsgrootte.

In het algemene geval van de toets van Wilcoxon beschouwen we twee steekproeven:

-
- 1) F. WILCOXON, Individual comparison by ranking methods, Biometrics 1 (1945), p. 80-83.
 - 2) H.B. MANN and D.R. WHITNEY, On a test of whether one of two random variables is stochastically larger than the other; Annals of Math. Statistics 18 (1947), p.50-60.
 - 3) Ir DORALIEN WABEKE en Dr CONSTANCE VAN EEDEN, Handleiding voor de toets van Wilcoxon, Rapport S 176 (M 65) van de statistische afdeling van het Mathematisch Centrum.

$$x_1, x_2, \dots, x_m$$

en

$$y_1, y_2, \dots, y_n$$

waarbij m niet gelijk aan n behoeft te zijn. We berekenen nu de grootheid W , die we krijgen door alle x -waarnemingen te vergelijken met alle y -waarnemingen en de som te nemen van tweemaal het aantal keren dat een x groter is dan een y en éénmaal het aantal keren dat een x gelijk is aan een y . Uit deze definitie volgt dat W minstens gelijk aan 0 en hoogstens gelijk aan $2mn$ is, en dat W deze uiterste waarden alleen aanneemt als alle x -waarnemingen kleiner resp. groter zijn dan alle y -waarnemingen. Ook is gemakkelijk in te zien dat een grote waarde van W met veel grote x -waarnemingen en een kleine waarde van W met weinig grote x -waarnemingen correspondeert. Het ligt daarom voor de hand te besluiten dat de x -waarden systematisch hoger resp. lager liggen dan de y -waarden als de gevonden waarde van W groot resp. klein is. De grootheid W wordt dan bij de toets van Wilcoxon - evenals het aantal successen bij de binomiale toetsen - als criterium voor de te nemen beslissing gebruikt. Men noemt een dergelijke grootheid een toetsingsgrootheid. W is dus de toetsingsgrootheid bij de toets van Wilcoxon. Vaak gebruikt men in plaats van W de grootheid $U = \frac{1}{2}W$. Het gebruik van W heeft echter het voordeel dat we nooit met gebroken getallen behoeven te werken.

Voor de berekening van W uit twee gegeven steekproeven kan men met vrucht van het volgende voorschrift gebruik maken:

1. Rangschik de getallen $y_1, y_2, \dots, y_m$ naar opklimmende grootte in de tweede kolom van een schema, beginnende op de tweede regel en telkens één regel overslaand, met dien verstande dat gelijke waarnemingen op dezelfde regel komen.
2. Noteer in kolom (3) de x -waarden, die niet gelijk zijn aan één der y -waarden, op de regel tussen de y -waarden waar zij tussen liggen, en noteer de x -waarden, die wel gelijk zijn aan een y -waarde, op dezelfde regel als die y -waarde.
3. De getallen die in kolom (1) komen te staan verkrijgt men nu, door voor iedere regel tweemaal het aantal y -waarden

te nemen dat er boven staat, vermeerderd met het aantal y-waarden op dezelfde regel. Op de laatste regel in kolom (1) komt dan het getal $2n$ te staan (contrôle).

4. De bijdrage tot W van iedere regel, waarop één of meer x-waarden voorkomen, wordt in kolom (4) genoteerd. Deze wordt verkregen door het aantal x-waarden in kolom (3) met het getal, dat op dezelfde regel in kolom (1) staat, te vermenigvuldigen.
5. Door optelling van de getallen in kolom (4) vindt men W .
6. In kolom (5) noteert men de aantallen gelijke waarnemingen op iedere regel. Optelling behoort het totale aantal waarnemingen $N = m+n$ te geven.
7. In kolom (6) noteert men de derde machten van de getallen in kolom (5). Men telt deze derde machten bij elkaar op en noemt de aldus verkregen som D .

De kolommen (5) en (6) hebben wij vooralsnog niet nodig, doch in één der volgende paragrafen zal het nut van deze kolommen blijken.

Voorbeeld: Door een wijziging in het productieproces kunnen weerstanden sneller geproduceerd worden. Men heeft nu 2 partijen weerstanden gefabriceerd, één volgens het oude, één volgens het nieuwe procédé, en men wil onderzoeken of de elektrische weerstand voor de beide partijen verschillend is. Hiertoe neemt men van beide partijen een steekproef en meet de weerstanden van de exemplaren in deze steekproeven. De resultaten zijn (in Ohm):

1e steekproef ($\underline{x}$): 147, 150, 150, 153, 148, 144, 155, 151,
146, 151.

2e steekproef ($\underline{y}$): 140, 145, 147, 147, 138, 144, 149, 150.

De berekening van W volgens bovenstaand voorschrift geeft nu het volgende schema:

(1)	(2)	(3)	(4)	(5)	(6)
	waarnemingen van		bijdrage tot W	aantallen gelijken(t)	t^3
	y	x			
0					
1	138			1	1
2					
3	140			1	1
4					
5	144	144	5	2	8
6					
7	145			1	1
8		146	8	1	1
10	147, 147	147	10	3	27
12		148	12	1	1
13	149			1	1
14					
15	150	150, 150	30	3	27
16		151, 151, 153, 155	64	2, 1, 1	8, 1, 1
Totaal	n=8	m=10	W=129	N=18	D=78

7.3. Tweezijdige toetsing.

We zagen in het voorgaande reeds, dat we met de toets van Wilcoxon de hypothese H_0 , dat $x_1, \dots, x_m$ en $y_1, \dots, y_n$ onderling onafhankelijke waarnemingen zijn van eenzelfde stochastische grootte, willen toetsen. Ook zagen we, dat we in feite een andere hypothese H'_0 , die volgt uit H_0 , toetsen. Deze hypothese H'_0 houdt in, dat in de rij bestaande uit m letters x en n letters y, die we krijgen door de waarnemingen naar opklimmende grootte te rangschikken en daarna de x-waarnemingen en de y-waarnemingen door letters x resp. y te vervangen, alle mogelijke volgorden der letters x en y even waarschijnlijk zijn. De toetsingsgrootte $\underline{W}$, die een stochastische grootte is, daar hij afhangt van de waarnemingen die zelf waarden van stochastische grootheden zijn, heeft een waarschijnlijkheidsverdeling die, indien H'_0 juist is, berekend kan worden.

Het blijkt nu, dat waarden van W minder waarschijnlijk zijn naarmate zij verder van mn verwijderd zijn. Als H_0' juist is, en dus ook als H_0 juist is, zullen waarden van W dicht bij 0 of $2mn$ dus weinig voorkomen.

Grote waarden van W komen voor als de meerderheid der letters x achter, en de meerderheid der letters y vóór in de rij der gerangschikte waarnemingen staat. Een dergelijk resultaat zal vaak voorkomen als de stochastische grootheden $\underline{x}$ en $\underline{y}$, waarvan $x_1, \dots, x_m$ resp. $y_1, \dots, y_n$ waarnemingen zijn, niet dezelfde verdeling hebben, doch zodanige verdelingen dat geldt: $P \{ \underline{x} > \underline{y} \} > \frac{1}{2}$.

Kleine waarden van W zal men daarentegen vaak vinden als $P \{ \underline{x} > \underline{y} \} < \frac{1}{2}$ is.

Op grond van deze overwegingen kunnen we nu weer een beslissingsschema opstellen, evenals we dat gedaan hebben voor de tekentoets in hoofdstuk 5. Dit schema is weergegeven in figuur 7.1.

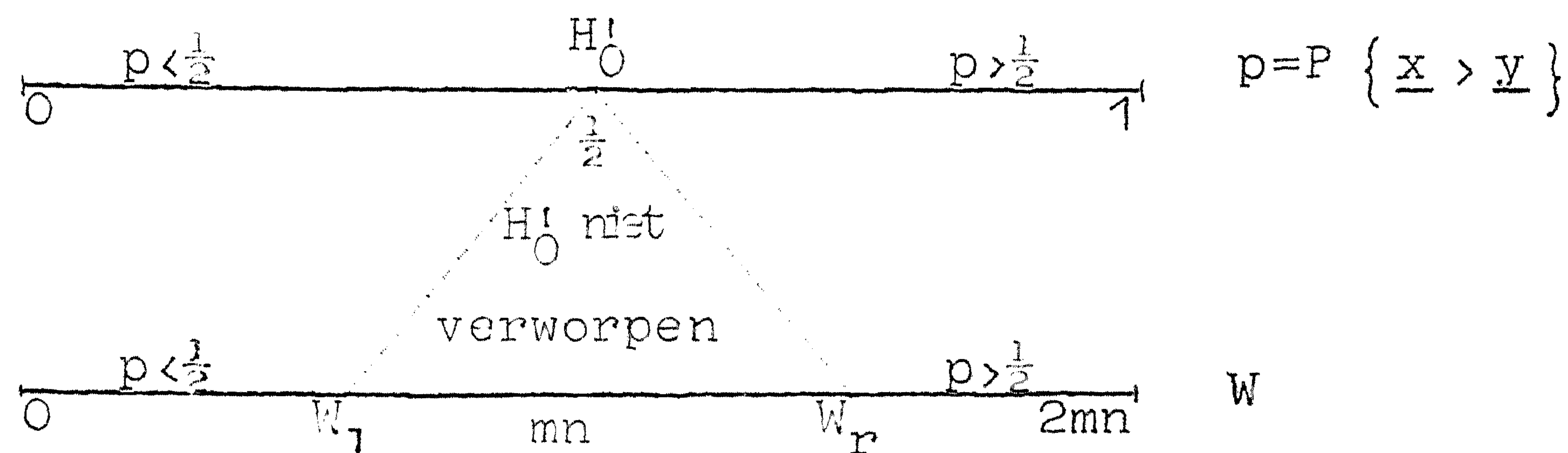


Fig. 7.1. Beslissingsschema voor de tweezijdige toets van Wilcoxon.

Op de bovenste lijn zijn de mogelijkheden (uitgedrukt in p) aangegeven, waartussen men wil beslissen; op de onderste lijn is aangegeven, welke conclusie men op grond van de gevonden waarde van W zal trekken. Daarbij treden weer twee kritieke waarden op voor W , zodanig, dat de kans, dat we een waarde van W vinden, gelijk aan of kleiner dan W_1 of gelijk aan of groter dan W_r hoogstens gelijk is aan de gekozen onbetrouwbaarheidsdrempel α is, indien H_0' juist is. Men concludeert $p < \frac{1}{2}$ als $W \leq W_1$, $p > \frac{1}{2}$ als $W \geq W_r$, terwijl H_0' , en daardoor ook H_0 , niet verworpen kan worden als $W_1 < W < W_r$ is.

Er is nog een kleine complicatie, die zich bij de teken-

toets niet voordeed: De hypothese H_0^1 houdt hier meer in dan $p=\frac{1}{2}$. Het is daarom zeer wel mogelijk dat H_0^1 niet juist is, terwijl toch $p=\frac{1}{2}$. (Dit is bijv. het geval als $\underline{x}$ en $\underline{y}$ beide symmetrisch verdeeld zijn ten opzichte van eenzelfde getal a , doch verschillende spreidingen bezitten.) Als H_0^1 verworpen wordt mag men dus strikt genomen nog niet concluderen dat $p \neq \frac{1}{2}$ is. Men kan echter bewijzen, dat er bij deze toets slechts een kleine kans is, die bovendien bij een toenemend aantal waarnemingen afneemt, om in het kritieke gebied terecht te komen als $p=\frac{1}{2}$ is, ook al is H_0^1 niet vervuld. Hierdoor is het aangegeven beslissingsschema voldoende gerechtvaardigd.

Wij zullen in paragraaf 7.5 aangeven hoe de kritieke waarden bepaald kunnen worden.

Ook bij de tweezijdige toets van Wilcoxon kan men een tweezijdige overschrijdingskans van iedere gevonden waarde W van $\underline{W}$ definiëren. Dit is de kans, dat $\underline{W}$ de waarde W of een minstens even ver van mn verwijderde waarde aanneemt, indien H_0^1 juist is. In formule is bijv. de tweezijdige overschrijdingskans van $W=80$ als $m=10$ en $n=20$:

$$k(80) = P\{\underline{W} \leq 80 \text{ of } \underline{W} \geq 400-80\} = P\{|\underline{W} - 200| \geq 120\}.$$

De wijze van bepalen van deze overschrijdingskansen wordt behandeld in paragraaf 7.6.

7.4. Eenzijdige toetsing.

Men kan de toets van Wilcoxon ook éénzijdig toepassen. Het schema van de rechts-éénzijdige toets is afgebeeld in figuur 7.2.

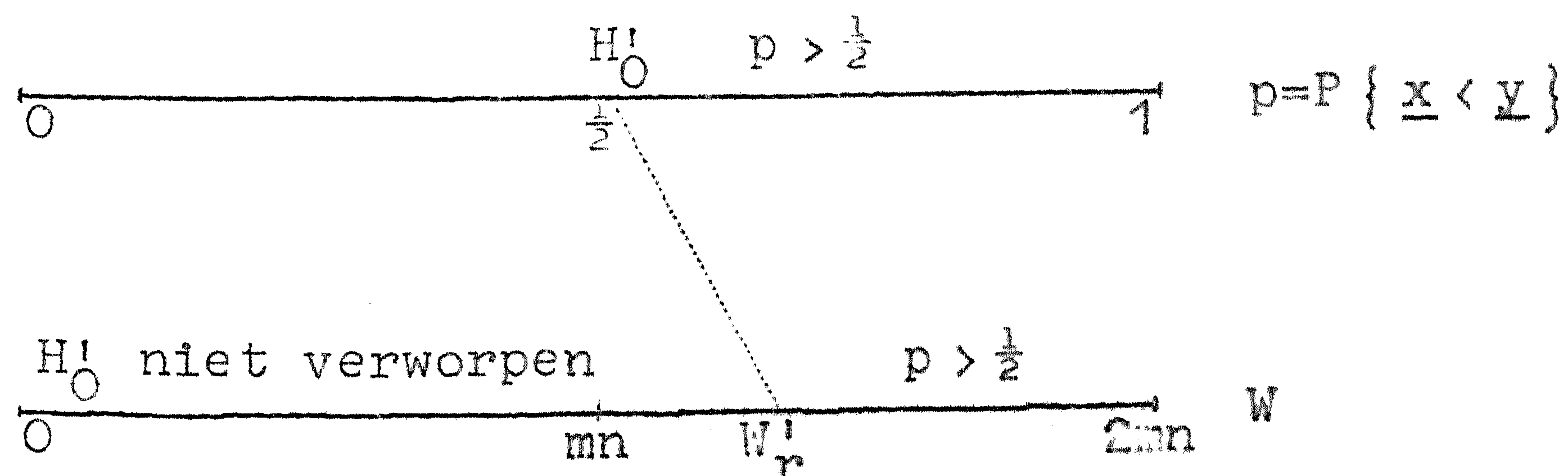


Fig. 7.2. Beslissingsschema van de rechts-éénzijdige toets van Wilcoxon

W'_r is hierin de rechts-éénzijdige kritieke waarde, en is zo gekozen dat de kans onder H'_0 , om een waarde van $\underline{W}$ die minstens gelijk aan W'_r is, hoogstens gelijk is aan de gekozen onbetrouwbaarheidsdrempel.

Wij toetsen hier de hypothese H'_0 tegen $p > \frac{1}{2}$, zodat we deze toets zullen gebruiken als we, bijv. op technische gronden, weten, dat p niet kleiner dan $\frac{1}{2}$ kan zijn, of als de mogelijkheid $p < \frac{1}{2}$ ons niet interesseert.

Bij de rechtséénzijdige toets behoort weer de rechts-éénzijdige overschrijdingskans. Dit is de kans, dat $\underline{W}$ onder de hypothese H'_0 de gevonden waarde W of een grotere aanneemt. In formule:

$$k_r(W) = P \{ \underline{W} \geq W | H'_0 \} .$$

We hebben dus weer dat $k_r(W'_r) \leq \alpha$.

Geheel analoog aan de rechtséénzijdige toetsing verloopt de linkséénzijdige toetsing: Nu wordt H'_0 getoetst tegen $p < \frac{1}{2}$. De linkséénzijdige kritieke waarde W'_l is kleiner dan mn en zodanig gekozen, dat onder H'_0 de kans op een waarde van $\underline{W}$, die kleiner dan of gelijk aan W'_l is, hoogstens α bedraagt. De linkséénzijdige overschrijdingskans $k_l(W)$ is gedefinieerd door:

$$k_l(W) = P \{ \underline{W} \leq W | H'_0 \} .$$

Evenals bij de tekentoets is de tweezijdige overschrijdingskans van W gelijk aan tweemaal de kleinste der eenzijdige overschrijdingskansen (en gelijk aan 1 als beide groter dan $\frac{1}{2}$ zijn).

Hierin schuilt het voordeel van de éénzijdige toetsen. Met de rechts-éénzijdige toets zal men reeds H'_0 kunnen verwerpen op grond van een gevonden waarde W , die bij toepassing van de tweezijdige toets nog niet tot verwerping van H'_0 zou leiden. Anders gezegd: De rechts-éénzijdige toets heeft voor $p > \frac{1}{2}$ een groter onderscheidingsvermogen dan de tweezijdige (en de links-éénzijdige) toets. Evenzo heeft de links-éénzijdige toets een groter onderscheidingsvermogen voor $p < \frac{1}{2}$, dan de tweezijdige (en de rechts-éénzijdige toets).

Men dient echter wel te bedenken, dat men niet op grond van de gevonden waarde van W mag beslissen of men één- dan wel tweezijdig wil toetsen. Doet men dit wel, dan heeft dit tot resultaat, dat men in feite met een grotere onbetrouwbaarheidsdrempel werkt.

7.5. Bepaling der kritieke waarden.

In tabel 7.II worden de linker kritieke waarden van de toets van Wilcoxon gegeven voor $2 \leq m \leq n \leq 10$ bij tweezijdige onbetrouwbaarheidsdrempels 0,10; 0,05; 0,02 en 0,01 (éénzijdig: 0,05; 0,025; 0,01 en 0,005). Is in een bepaald geval $m > n$, dan kunnen de letters m en n in de tabel verwisseld gedacht worden. De tabel blijft dan gelden. De overeenkomstige rechter kritieke waarden vinden we door de linker kritieke waarden van $2mn$ af te trekken.

Tabel 7.II. Linker kritieke waarden van de toets van Wilcoxon.

a) $\alpha = 0,10$ tweezijdig (0,05 éénzijdig)

$n \backslash m$	2	3	4	5	6	7	8	9	10
2	-								
3	-	0							
4	-	0	2						
5	0	2	4	8					
6	0	4	6	10	14				
7	0	4	8	12	16	22			
8	2	6	10	16	20	26	30		
9	2	8	12	18	24	30	36	42	
10	2	8	14	22	28	34	40	48	54

b) $\alpha = 0,05$ tweezijdig ($0,025$ eenzijdig)

$n \backslash m$	2	3	4	5	6	7	8	9	10
2	-								
3	-	-							
4	-	-	0						
5	-	0	2	4					
6	-	2	4	6	10				
7	-	2	6	10	12	16			
8	0	4	8	12	16	20	26		
9	0	4	8	14	20	24	30	34	
10	0	6	10	16	22	28	34	40	46

c) $\alpha = 0,02$ tweezijdig ($0,01$ eenzijdig)

$n \backslash m$	2	3	4	5	6	7	8	9	10
2	-								
3	-	-							
4	-	-	-						
5	-	-	0	2					
6	-	-	2	4	6				
7	-	0	2	6	8	12			
8	-	0	4	8	12	14	18		
9	-	2	6	10	14	18	22	28	
10	-	2	6	12	16	22	26	32	38

d) $\alpha = 0,01$ tweezijdig (0,005 eenzijdig)

$n \backslash m$	2	3	4	5	6	7	8	9	10
2	-								
3	-	-							
4	-	-	-						
5	-	-	-	0					
6	-	-	0	2	4				
7	-	-	0	2	6	8			
8	-	-	2	4	8	12	14		
9	-	0	2	6	10	14	18	22	
10	-	0	4	8	12	18	22	26	32

Strikt genomen zijn de in deze tabellen gegeven kritieke waarden alleen juist als er geen gelijke waarnemingen in de steekproeven optreden, dus als de getallen in kolom (5) van het schema in 7.2 alle gelijk aan 1 zijn. Indien er echter weinig gelijke waarnemingen voorkomen in groepjes van 2 of 3, dus als in kolom (5) van het schema naast enkele malen een 2 of een 3 uitsluitend énen voorkomen, kan men zonder bezwaar tabel 7.II gebruiken. De juiste linker kritieke waarden zijn dan in het algemeen iets hoger dan de in de tabel gegeven waarden. Dit heeft tot gevolg dat men dus soms H_0 bij gebruik van de tabel niet zal verwerpen, terwijl dit bij exacte behandeling misschien wel mogelijk zou zijn.

Als m of n of beide groter dan 10 zijn, dan is tabel 7.II niet meer te gebruiken. We moeten dan uitgebreider tabellen raadplegen. Deze zijn opgenomen in "Handleiding voor de toets van Wilcoxon", door Ir Doralien Wabeke en Dr Constance van Eeden (Rapport S 176 (M 65) van de Statistische Afdeling van het Mathematisch Centrum). Indien echter één der getallen m en n groter is dan 10, terwijl het

andere niet kleiner dan 5 is, terwijl bovendien de groepen gelijke waarnemingen niet te veel in grootte verschillen, kunnen we een benaderende methode gebruiken. Daartoe berekenen we de grootheid σ^2 met behulp van de formule

$$(7.5.1) \quad \sigma^2 = \frac{m.n.(N^3-D)}{3.N.(N-1)} .$$

De getallen N en D, die in deze formule optreden, vinden we onderin de kolommen (5) en (6) van het rekenschema uit 7.2. We vinden dan de kritieke waarden voor de tweezijdige toets uit

$$(7.5.2) \quad W_r = mn + \lambda' . \sigma + a$$

en

$$(7.5.3) \quad W_l = mn - \lambda' . \sigma - a .$$

De waarde van λ' die we hierin moeten substitueren vinden we in tabel 4.IV. De daar opgegeven waarden van α moeten we echter verdubbelen. Voor a moeten we de waarde 1 substitueren, tenzij:

- a. de waarnemingen bestaan uit een aantal groepen, ieder bestaande uit eenzelfde aantal t gelijke waarnemingen. (De getallen in kolom (5) van het rekenschema zijn dan alle gelijk aan t). We moeten dan voor a de waarde t nemen.
- b. de waarnemingen uit twee groepen van gelijke waarnemingen bestaan. (In het rekenschema komen dan slechts twee regels voor waarop waarnemingen komen te staan.) In dit geval moeten we $a = \frac{1}{2}N$ nemen.

De kritieke waarden voor de rechts- resp. links-eenzijdige toets vinden we eveneens uit formule (7.5.1) resp. (7.5.2), met dit verschil, dat we nu de waarden van α , die in tabel 4.IV opgegeven worden, onveranderd gebruiken.

7.6. Bepaling der overschrijdingskansen.

De exacte waarden der één- en tweezijdige overschrijdingskansen kan men, als beide steekproeven niet meer dan 10 waarnemingen bevatten, terwijl er geen gelijke waarnemingen optreden, vinden met behulp van tabellen die in de eerder-

genoemde "Handleiding voor de toets van Wilcoxon" zijn opgenomen. Tabel 7.II is aan de hand van deze tabellen samengesteld; deze tabel kan, evenals dat bij de tekentoets het geval was, gebruikt worden om grenzen aan te geven voor de overschrijdingskansen. Zo vindt men voor $m=9$ en $n=6$ voor de rechtséénzijdige overschrijdingskans van $W=94$:

$$0,01 \geq k_r(94) > 0,005.$$

Immers, $W_r = 108 - W_1$ ligt voor $\alpha = 0,01$ bij 94 en voor $\alpha = 0,005$ bij 98.

Voor grotere waarden van m en n kunnen de overschrijdingskansen weer benaderd worden met behulp van tabel 5.II, mits de groepen gelijke waarnemingen niet teveel in grootte verschillen. Ter bepaling van de tweezijdige overschrijdingskansen berekent men

$$(7.6.1) \quad \xi_\alpha = \frac{|W - m \cdot n| - a}{\sigma}$$

waarin de letters σ en a dezelfde betekenis hebben als in 7.5. Men vindt vervolgens in tabel 5.II de bijbehorende waarde van α . De tweezijdige overschrijdingskans van W is dan gelijk aan 2α .

Ter bepaling van de rechts-éénzijdige overschrijdingskansen berekent men

$$(7.6.2) \quad \xi'_\alpha = \frac{W - mn - a}{\sigma}$$

en voor de links-éénzijdige overschrijdingskansen

$$(7.6.3) \quad \xi''_\alpha = \frac{W - mn + a}{\sigma}$$

en handelt hiermee als in 5.3.

Opmerking:

De uitdrukking (7.5.1) voor σ^2 is ook te schrijven als

$$(7.6.4) \quad \sigma^2 = \sigma_u^2 - \frac{A}{100} \cdot (D-N)$$

waarin $\sigma_u^2 = \frac{mn(N+1)}{3}$ gelijk is aan de waarde die σ^2 zou

hebben indien alle waarnemingen verschillend waren, terwijl

$$A = \frac{100 \cdot m \cdot n}{3N \cdot (N-1)} \quad .$$

Indien men nu de term $\frac{A}{100} \cdot (D-N)$, de correctieterm voor gelijke waarnemingen, bij de berekening van σ^2 weglaat, zodat men krijgt: $\sigma^2 = \sigma_u^2$, dan werkt men met een te grote waarde van σ^2 . Dit heeft tot gevolg, dat men dan een te grote waarde voor de overschrijdingskans vindt. Indien nu deze te grote overschrijdingskans reeds tot verwerping van H_0' leidt, dan zal H_0' zeker verworpen worden indien men met de juiste waarde van σ^2 werkt. Men kan dus het rekenwerk bekorten, door eerst te handelen alsof geen gelijke waarnemingen aanwezig zijn, en alleen indien H_0' op deze wijze bij de gekozen onbetrouwbaarheidsdrempel niet verworpen wordt de correcte berekening uit te voeren.

Een verdere beperking van het rekenwerk is mogelijk door gebruik te maken van tabellen welke onder de naam "Auxiliary tables for Wilcoxon's two sample test" by J. ZAALBERG verschenen zijn in Statistica Neerlandica 12 (1958) p.265-273. In deze tabellen kan men zonder enig rekenwerk de waarden van de grootheden σ_u^2 en A uit formule (7.6.4) aflezen voor $2 \leq m \leq 25$ en $11 \leq n \leq 25$.

7.7. Voorbeelden.

a. Men wil de treksterkte van twee partijen metaaldraden vergelijken en neemt daartoe uit partij A een steekproef van $m=8$ draden en uit partij B een van $n=6$ draden. Men bepaalt van deze draden de treksterkte en vindt de volgende uitkomsten:

A: 62,9 68,1 64,6 67,9 65,3 67,4 66,9 67,2
B: 68,2 66,3 65,2 65,4 68,0 67,5.

Onder deze waarnemingen komen geen gelijken voor. We vinden $W=38$. We toetsen tweezijdig en kiezen als onbetrouwbaarheidsdrempel $\alpha = 0,05$. In tabel 7.II vinden wij bij $m=8$ en $n=6$ $W_1=16$, zodat $W_r = 2mn - W_1 = 80$. Aangezien de gevonden waarde van W tussen W_1 en W_r ligt, kunnen we de hypothese H_0 , inhoudende dat de draden van beide partijen - wat hun treksterkte betreft - gelijkwaardig zijn, niet verwerpen.

- b. In het voorbeeld uit 7.2 zullen wij eveneens tweezijdig willen toetsen, en wel met $\alpha = 0,05$. Hoewel er hier enkele gelijke waarnemingen optreden maken wij toch gebruik van tabel 7.II, daar de aantallen gelijken slechts klein zijn. Bij $n=8$ en $m=10$ vinden we $W_1=34$, zodat $W_r=160-34=126$. Daar de gevonden waarde 129 van W groter is dan W_r , besluiten wij dat de waarden van de weerstanden die volgens het oude procédé gefabriceerd zijn systematisch hoger liggen dan de andere.

Hadden we $\alpha=0,02$ gekozen, dan hadden we gevonden $W_r=134$, zodat $W=129$ dan niet in het kritieke gebied had gelegen. We concluderen hieruit dat de tweezijdige overschrijdingskans van $W=129$ groter is dan 0,02. Derhalve is $0,02 < k(129) < 0,05$.

- c. Gemeten is de tijds(in seconden) die personen nodig hebben voor een bepaalde psychologische test. Het waarnemingsmateriaal bestaat uit waarnemingen bij 100 personen van 30-44 jaar ($x_1, x_2, \dots, x_{100}$) en waarnemingen bij 51 personen van 45-60 jaar ($y_1, y_2, \dots, y_{51}$). Men wil onderzoeken of de tijdsduur afhangt van de leeftijd van de proefpersoon. Het waarnemingsmateriaal is geclassificeerd als in tabel 7.III is aangegeven.

Tabel 7.III Tijdsduur (in seconden) van een psychologische test bij personen van 30-44 resp. 45-60 jaar.

Klasse indeling voor s	Aantal waarnemingen	
	van $\underline{x}$	van $\underline{y}$
120-149	7	1
150-179	12	1
180-209	15	7
210-239	21	8
240-269	13	4
270-299	11	11
300-329	11	6
330-359	5	6
360-389	2	2
390-419	3	5

Uit deze tabel ziet men dus bijv. dat 7 personen tussen 30 en 44 jaar voor de test een tijd nodig hadden die tussen 120 en 149 seconden ligt, etc. We hebben nu niet de waarnemingen zelf tot onze beschikking, doch alleen de tijdsintervallen waar zij inliggen. Dit is echter geen bezwaar, omdat bij de toets van Wilcoxon alleen de groottevolgorde der waarnemingen van belang is. Het enige gevolg van de indeling in tijdsintervallen is, dat we de waarnemingen op één regel als een groep gelijken moeten beschouwen, waardoor de toets minder gevoelig wordt. In het rekenschema zetten we nu in de kolommen (2) en (3) **alleen** de aantallen waarnemingen die op elke regel behoren te staan.

Rekenschema:

(1)	(2)	(3)	(4)	(5)	(6)
	Aantal waarnemingen van		Bijdrage tot w	Aantallen gelijken (t)	t^3
	<u>y</u>	<u>x</u>			
0					
1	1	7	7	8	512
2					
3	1	12	36	13	2197
4					
11	7	15	165	22	10648
18					
26	8	21	546	29	24389
34					
38	4	13	494	17	4913
42					
53	11	11	583	22	10648
64					
70	6	11	770	17	4913
76					
82	6	5	410	11	1331
88					
90	2	2	180	4	64
92					
97	5	3	291	8	512
102					
Totaal	n=51	m=100	W=3482	N=151	D=60127

We gaan nu tweezijdig toetsen met $\alpha = 0,01$. We berekenen achtereenvolgens:

$$\sigma_u^2 = \frac{51 \cdot 100 \cdot 152}{3} = 258.400, \text{ dus } \sigma_u = 508,4$$

$$A = \frac{100 \cdot 100 \cdot 51}{3 \cdot 151 \cdot 150} = 7,506$$

$$\sigma^2 = 258.400 - \frac{7,506}{100} \cdot (60127 - 151) = 253898$$

$$\sigma = 503,9.$$

Berekenen we vervolgens $\mathcal{Y}_\alpha$ met gebruikmaking van σ_u en

$$\sigma, \text{ dan vinden we } \mathcal{Y}_\alpha = \frac{|3482-5100| - 1}{508,4} = 3,181$$

$$\text{resp. } \mathcal{Y}_\alpha = \frac{|3482-5100| - 1}{503,9} = 3,209.$$

Het verschil, d.w.z. de invloed van de gelijken, is dus zeer klein.

In tabel 5.II vinden we bij $\mathcal{Y}_\alpha = 3,18$ resp. $3,21$ voor α :
 $\alpha = 0,0007$ resp. $\alpha = 0,0007$, dus $k = 0,0014$. Daar $k < 0,01$ verwerpen we de hypothese dat de leeftijd geen invloed zou hebben op de tijdsduur, en we besluiten dat jongere mensen minder tijd voor de test nodig hebben dan oudere mensen.

Hadden we hier eerst alleen σ_u^2 berekend, zonder D uit te rekenen, dan hadden we deze conclusie ook al kunnen trekken. Het enige wat wij gewonnen hebben door de berekening van D is, dat wij weten dat $k = 0,0014$, terwijl we anders alleen wisten dat $k \leq 0,0014$.

- d. Men wil twee geneesmiddelen A en B vergelijken wat betreft hun uitwerking bij de behandeling van een bepaalde ziekte; 5 patiënten lijdende aan deze ziekte krijgen een injectie met geneesmiddel A, en 10 patiënten een injectie met B. Twee weken daarna wordt voor iedere patiënt de graad van verbetering genoteerd; deze verbeteringsgraad wordt in een cijfer uitgedrukt:

1. geen of geringe verbetering
2. matige verbetering
3. goede verbetering
4. geheel genezen.

De waarnemingen zijn samengevat in tabel 7.IV:

Tabel 7.IV Verbeteringsgraad van patiënten, behandeld met geneesmiddelen A en B.

Verbeterings- graad	Aantal patiënten met de betreffende verbeterings- graad bij geneesmiddel		Totaal
	A	B	
1	1	0	1
2	3	5	8
3	1	2	3
4	0	3	3
Totaal	5	10	15

We hebben nu het geval $m=5$, $n=10$, terwijl de grootten van de groepen gelijke waarnemingen zeer uiteenlopen (1, 8, 3, 3). Hierdoor kunnen we hier noch tabel 7.II, noch tabel 5.II gebruiken, zodat wij in dit geval met de ons ten dienste staande methoden geen verantwoorde conclusie kunnen trekken.

- e. Een textielfabriker wil door toepassing van een prepareermethode A zijn garens sterker maken. Trekproeven, verricht op monsters onbehandelde en met A behandelde garens, geven de volgende resultaten (uitgedrukt in één of andere sterkte-eenheid):

Met A behandelde garens($\underline{x}$)	Onbehandelde garens($\underline{y}$)
11,79	10,34
11,21	11,40
13,20	10,19
12,66	12,10
13,37	11,46

De hypothese die wij willen toetsen luidt: de behandelde en de onbehandelde garens zijn even sterk.

Aangezien de fabrikant de prepareermethode echter alleen wil invoeren indien deze methode de garens sterker maakt, is hij er dus alleen in geïnteresseerd H_0 te verwerpen ten gunste van de alternatieve hypothese dat de met A behandelde garens systematisch sterker zijn dan de niet behandelde. De twee mogelijkheden:

1. de prepareermethode helpt niet

en 2. de prepareermethode schaadt

leiden beide tot dezelfde beslissing: de methode niet in te voeren, zodat zij bij dit probleem, praktisch gezien, equivalent zijn. Wij behoeven deze twee mogelijkheden dus niet van elkaar te onderscheiden. Het linkerdeel van de kritieke zone van een tweezijdige toets zou dus doelloos zijn. Wij toetsen daarom rechts-éénzijdig, en wel met $\alpha = 0,05$.

We vinden $W=42$ en, uit tabel 7.II bij $m=n=5$, $W_1^1=8$, dus $W_r^1=42$. We verwerpen derhalve H_0 ten gunste van de hypothese dat de prepareermethode de garens sterker maakt. Indien wij hier tweezijdig getoetst hadden, ook met $\alpha = 0,05$, dan hadden wij deze conclusie nog niet kunnen trekken.

-.--.-.