MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 95 (R3)

Cursus:

Elementaire Mathematische Statistiek

Hoofdstuk I

Algemene Inleiding

door

Prof.Dr J. Hemelrijk

1952/53

1952
3^e druk 1955

Cursus Elementaire Mathematische Statistiek

S 95 (l3)

Inhoud

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.

I Algemene Inleiding.

1. Opvatting. Wij zullen de statistiek opvatten als een toepassing van de waarschijnlijkheidsrekening [1]. De whr, op haar beurt, beschouwen wij als een onderdeel der zuivere wiskunde, opgebouwd op een klein aantal axioma's. Alvorens in te gaan op deze axioma's en hun heuristische achtergrond, enige opmerkingen over de aard van de verschijnselen, waarop de statistiek kan worden toegepast.

2. Object der wetenschap. Iedere wetenschap houdt zich bezig met het onderzoek van regelmatigheden (of: wetmatigheden) in de waarneembare werkelijkheid. Deze wetmatigheden worden in formules of woord-formuleringen samengevat. Een duidelijk voorbeeld is de elementaire natuurkunde, zoals die op de middelbare school onderwezen wordt. De opgespoorde regelmatigheden worden vervolgens, zo goed en zo kwaad als dat gaat, met elkaar in verband gebracht en tot een zo klein mogelijk aantal gereduceerd ("grondwetten"); uit deze grondwetten, die men als een soort experimentele axiomas kan beschouwen, kunnen de overige wetten door logische redenatie (deductie) worden afgeleid. Bij eenvoudige wetenschappen, zoals de elementaire natuurkunde en mechanica, gaat dit zeer goed; bij ingewikkeldere, zoals scheikunde gaat het minder goed, terwijl b.v. in de kernphysica het mathematische apparaat zozeer gaat overheersen, dat er van de aanschouwelijkheid van het gebruikte schema (het "model") weinig overblijft. Alle experimentele wetenschappen hebben echter dit gemeen, dat zij steeds weer aan de werkelijkheid worden getoetst. De techniek berust op de resultaten van dit soort van wetenschappelijk onderzoek en is als de toepassing daarvan te beschouwen.

3. Statistische verschijnselen. De meeste wetenschappen houden zich bezig met verschijnselen, die zich onder dezelfde omstandigheden steeds weer op dezelfde wijze voordoen. Of althans: van de bestaande verschillen wordt geabstraheerd; zij worden verwaarloosd. Het ligt in dergelijke omstandigheden zeer voor de hand, om een "oorzaak-gevolg"-denkschema te gebruiken en dit wordt dan ook meestal gedaan. De statistiek houdt zich nu juist bezig met verschijnselen, die bij herhaling van dezelfde omstandigheden een zekere variatie vertonen (wij zullen deze verschijnselen "statistische verschijnselen" noemen). Voorbeelden: de breeksterkte van garen (meet men deze op een aantal plaatsen van een klosje garen - de omstandigheden worden dus zo goed mogelijk gelijk gehouden - dan verschillen de uitkomsten); de

---

[1]) Afkorting: whr.

levensduur van gloeilampen (een partij gloeilampen, in serie gefabriceerd - dus alle zoveel mogelijk op dezelfde wijze - ; toch verschilt de levensduur van lamp tot lamp); aantallen bacterieën in een watermonster (een aantal monsters wordt uit een rivier of bassin genomen - alle op dezelfde wijze - en bevatten verschillende aantallen bacterieën); herhaalde metingen met eenzelfde apparatuur verricht aan eenzelfde object geven in het algemeen onderscheidbaar-verschillende uitkomsten.

Ook indien de omstandigheden minder sterk op elkaar gelijken dan bij de bovenstaande voorbeelden, zodat een verschil in de uitkomst meer voor de hand ligt, wordt de statistiek vaak toegepast. De omstandigheden, of de beschouwde objecten, zijn dan soortgelijk en men ziet van de verschillen af, op enkele na, die men juist wil onderzoeken. Voorbeelden: de geboorten in een stad of land gedurende een bepaalde tijd worden onderscheiden in jongens- en meisjesgeboorten en de verhouding der aantallen wordt aan een onderzoek onderworpen; door het vangen van vissen, voorzien van een merkteken, weer loslaten en later weer vangen, waarbij in de tweede vangst onderscheid gemaakt kan worden tussen "voor het eerst gevangen" en "voor de tweede maal gevangen", kan men de grootte van een vis-bevolking statistisch schatten (ook bij andere dieren dan vissen wordt deze methode toegepast); door op een aantal tijdstippen in een fabriek na te gaan, hoeveel machines van een bepaalde groep op ieder van deze momenten stilstaan, kan men een schatting verkrijgen van het rendement van die machine-groep. Bij het laatste voorbeeld zijn de tijdstippen van verschillende aard: tijdens de schafttijd b.v. zullen er vele (zo niet alle) stilstaan. Bij de keuze der tijdstippen zal men met deze ongelijksoortigheid op later te bespreken wijze rekening dienen te houden; bij het statistische onderzoek wordt dan dat verschil verwaarloosd.

De hier gegeven gebiedsomschrijving is niet volledig; de bedoeling is een indruk te geven van het soort van problemen, waarmee de statistiek zich bezighoudt.

4. Fundamentele statistische verschijnselen. Hoewel dus de verschijnselen, waar de statistiek op toegepast wordt, minder ondubbelzinnig gedetermineerd zijn dan bij andere wetenschappen, berust ook de statistiek op bepaalde herhaalbare en controleerbare verschijnselen. Deze zijn van zeer eenvoudige aard, zoals uit de volgende voorbeelden blijkt.

Werpt men 100 maal met een munt, dan komt er een aantal malen "kruis". Noemen we dit aantal n en herhalen wij het experiment, dan zal n bij deze herhaling gewoonlijk een andere waarde aannemen dan de eerste keer. Maar het komt niet voor, dat n bij de eerste serie gelijk aan 0 is en bij de tweede gelijk aan 100.

Ook in andere dan "laboratoriumomstandigheden" kan men dit constateren: beschouwt men de verhouding van het aantal jongens- tot het aantal meisjesgeboorten in Delft in verschillende jaren, dan zal men zien, dat geen van deze verhoudingen zeer groot of zeer klein is.

Stel men heeft een vaas met een onbekend aantal witte en zwarte ballen erin en men trekt hieruit met teruglegging van de getrokken bal, goed schuddende tussen iedere twee trekkingen. Heeft men nu 10.000maal getrokken en steeds een witte bal verkregen, dan kan men ervan opaan, dat de 10.001^e bal geen zwarte is.

Indien men een pak speelkaarten goed schudt en vervolgens in 4 stapeltjes uitdeelt, komen er niet in ieder stapeltje uitsluitend kaarten van één kleur voor.

Deze uitspraken zijn alle negatief van karakter: bepaalde verschijnselen doen zich niet voor. Weliswaar zijn zij niet onmogelijk, maar zij realiseren zich niet. Een ieder kan dit controleren. De negatieve uitspraken kunnen soms omgezet worden in voorspellingen van positieve vorm, b.v. bij de vaas met ballen: "de 10.001^e bal zal een witte zijn". Het is de taak van de statistiek dit soort voorspellingen te verscherpen tot voorspellingen, waar men meer aan heeft.

5. Experimentele wet der grote getallen. Uit het bovenstaande blijkt reeds, dat ook statistische verschijnselen, ondanks hun betrekkelijk geringe regelmaat, aan bepaalde experimenteel vaststelbare regels gehoorzamen. Het centrale ervaringsfeit, waarop de toepassing der whr berust (let wel: de toepassing, niet de whr zelf; deze is een onderdeel der zuivere wiskunde en berust slechts op een aantal axiomas en op de regels der logische deductie), is van iets ingewikkeldere aard. Dit centrale ervaringsfeit, de experimentele wet der grote getallen, kan a.v. omschreven worden:

Indien men een experiment, dat een aantal verschillende resultaten kan hebben, een aantal (N) malen op dezelfde wijze herhaalt en het aantal malen, dat een bepaald resultaat S optreedt, deelt door N, dan verkrijgt men een tussen 0 en 1 liggend getal, dat het waargenomen frequentiequotiënt [2]) q van S wordt genoemd. Verricht men nu een aantal van deze reeksen experimenten, met toenemende N, dan blijkt gewoonlijk, dat de verschillen tussen de gevonden waarden van q kleiner worden naarmate N groter wordt.

----

2) Afkorting: fq, meervoud fqn.

In tweede instantie kunnen wij statistische verschijnselen definiëren als verschijnselen, die aan de experimentele wet der grote getallen gehoorzamen, of waarvan men aanneemt, dat zij dat zouden doen, als het mogelijk was ze voldoende vaak te herhalen. Deze definitie is reeds iets scherper dan de in § 3 gegevene. Het woord "gewoonlijk" in de bovenstaande formulering van de experimentele wet der grote getallen laat nog een zekere vaagheid bestaan. Een scherpe gebiedsomschrijving valt echter van geen enkele wetenschap te geven. Bovendien is de wet der grote getallen in hoofdzaak een heuristisch principe voor de toepassing der whr. Men kan de bruikbaarheid van iedere statistische methode aan de practijk of door experimenten toetsen, zoals men ook natuurwetenschappelijke wetten steeds door experimenten moet verifiëren.

Fundamenteel voorbeeld van de experimentele wet der grote getallen: wanneer men een aantal (b.v. 20) zoveel mogelijk gelijke kaarten neemt en op één hiervan een f zet en op de overige een g en men trekt hieruit een groot aantal malen een kaart, steeds na goed schudden en met terugleggen van de getrokken kaart, dan zal het fq van het kenmerk f, dicht bij 1/20 liggen.

Dit voorbeeld lijkt triviaal, maar is het niet. Wij zullen later zien, dat vele statistische methoden bestaan uit het trekken van een conclusie (resp. het doen van een voorspelling), waarvan men niet weet of hij juist is (resp. of hij uit zal komen), terwijl in een reeks toepassingen van dergelijke methoden het fq van het aantal onjuiste conclusies (resp. voorspellingen) juist op dezelfde wijze tot stand komt als het fq van het kenmerk f bij het beschreven kaartenexperiment. Dit betekent dus, dat het fq der foute conclusies in een lange reeks ook dicht bij 1/20 ligt. (Wij hebben hier 1/20 genomen; dit is een traditionele waarde, die echter willekeurig is. Men kan ook 1/100 of een andere breuk tussen 0 en 1 nemen.) Een nadere precisering van "dicht bij" volgt later.

6. Frequentiequotiënten. Een andere definitie van statistiek is, dat deze wetenschap zich bezig houdt met de bepaling van fqn en de eigenschappen daarvan op bepaalde collecties [3]. Deze collecties kunnen verzamelingen van soortgelijke objecten zijn (de vissen in een viswater, de door een fabriek gefabriceerde en nog te fabriceren condensatoren, etc.) of ook verzamelingen van experimenten, zoals bij de experimentele wet der grote getallen. Wij leiden daarom enkele eenvoudige eigenschappen van fqn af.

---

[3]) Vaak populaties of universa genoemd.

Wij beschouwen een eindige collectie Λ, met elementen λ, die ieder één of meer der kenmerken A,B,C,... kunnen bezitten. Nu gelden de volgende eigenschappen:

(1) $0 \leqq fq(A) \leqq 1$ voor iedere A

(2) $fq(A)=0$ als A niet voorkomt op Λ

(3) $fq(A)=1$ als iedere $\lambda \epsilon \Lambda$ kenmerk A bezit

(4) $fq(A \vee B)=fq(A)+fq(B)-fq(A \wedge B)$

($\vee$=of, $\wedge$=en, ϵ=behoort tot).

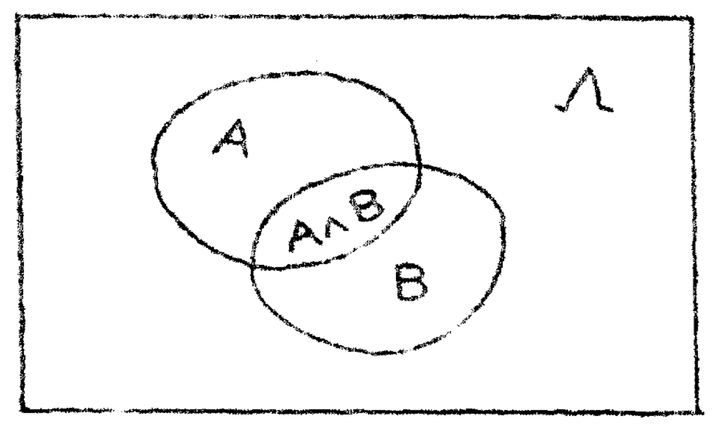


Fig. 1

Schematische voorstelling van collectie Λ met kenmerken A en B.

Naast gewone fqn worden voorwaardelijke fqn ingevoerd. Het voorwaardelijke fq van A onder de voorwaarde B is per definitie gelijk aan het quotiënt van het aantal eln [4]) λ, dat de kenmerken A en B bezit en het aantal eln, dat kenmerk B bezit:

(5) $$fq(A|B) \stackrel{def}{=} \frac{fq(A \wedge B)}{fq(B)} . \qquad (fq(B) \neq 0).$$

Men kan ook zeggen: het voorwaardelijke fq van A onder voorwaarde B is het gewone fq van A, indien wij dit berekenen op de deelcollectie $\Lambda(B)$ van Λ, bestaande uit alle eln λ, die kenmerk B bezitten (ga de betekenis hiervan in fig. 1 na).

Twee kenmerken A en B heten onafhankelijk verdeeld over Λ, indien

(6) $$fq(A|B) = fq(A)$$

is. Dan volgt uit (5):

(7) $$fq(A \wedge B) = fq(A)fq(B),$$

en derhalve geldt ook

(8) $$fq(B|A) \stackrel{def}{=} \frac{fq(A \wedge B)}{fq(A)} = fq(B).$$

7. Waarschijnlijkheid. Wij zullen nu het begrip waarschijnlijkheid [5]) van een gebeurtenis invoeren. Een gebeurtenis kan zijn: het verkrijgen van kruis (K) bij het werpen van een munt, het feit, dat een exemplaar in een serieproductie ondeugdelijk is, het feit, dat de meting van een physische grootheid een in een

---

[4]) El=element, eln=elementen.

[5]) Afkorting: wh, meervoud whn.

zeker interval gelegen waarde aanneemt, etc. In het algemeen: de uitkomst van een (al of niet werkelijk verrichte) waarneming. Als illustratief voorbeeld nemen wij het optreden van K bij het werpen met een munt. De intuïtieve achtergrond van het wh-begrip is, dat wij ons een grote serie worpen van deze munt voorstellen, waarin het kenmerk K dan met een zeker fq optreedt. Volgens de experimentele wet der grote getallen (zie § 5) vertoont fq(K) bij lange series een zekere stabiliteit. Men zou nu geneigd zijn het begrip "kans [6]) op K" in te voeren met behulp van een limietovergang tot een oneindig lange serie worpen. Een belangrijke poging in deze richting is gedaan door R.VON MISES [7]), maar daarbij doet zich de grote moeilijkheid voor, dat deze limietovergang geen experimenteel correlaat heeft: reeksen van experimenten zijn altijd eindig. Het verband met de werkelijkheid wordt daardoor problematisch. Wij voeren daarom het begrip "wh" in als een abstract mathematisch begrip, waarvan de eigenschappen axiomatisch worden vastgelegd. Daarbij hebben wij echter een denkbeeldig lange reeks experimenten en het daarbij optredende fq wel in gedachten en wij kiezen de axiomas dan ook analoog aan de in §6 afgeleide eigenschappen van fqn. Het verband van de op die wijze verkregen theorie met de werkelijkheid, dat voor toepassing van de theorie nodig is, moet dan nog gelegd worden. Dit geschiedt in §17.

8. Axiomas. Wij beschouwen dus een experiment E, dat uitkomsten A,B,... kan hebben. Het optreden van een bepaalde uitkomst noemen wij ook wel een gebeurtenis en we geven die met dezelfde letter als de uitkomst aan. Het axiomastelsel der whr kan nu in de volgende vorm gegeven worden [8]) (vgl. § 6, (1),...,(4)).

Ax.1: Ieder der gebeurtenissen A,B,... bezit een wh, d.i. een reëel getal tussen 0 en 1; notatie voor de gebeurtenis A:

(9) $$0 \leqq P[A] \leqq 1.$$

Ax.2: Indien A onmogelijk is, is

(10) $$P[A] = 0.$$

Ax.3: Indien A zeker is, is

(11) $$P[A] = 1.$$

Ax.4:

(12) $$P[A \vee B] = P[A] + P[B] - P[A \wedge B].$$ (optelregel)

----

6) Kans en wh zijn synoniem.

7) R.van Mises, Wahrscheinlichkeitsrechnung, Leipzig-Wien 1931, N.Y. 1943.

8) De axiomas zijn, in de hier gegeven vorm, niet onafhankelijk. Wij zullen ons daarom niet bekommeren. Voor een zorgvuldigere opzet raadplege men A.KOLMOGOROFF, Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin 1933,N.Y.1946.De axiomatische opzet der whr is van Kolmogoroff afkomstig.

De definitie van een voorwaardelijke wh is analoog aan (8)

(13) $$P[A|B] \stackrel{\text{def}}{=} \frac{P[A \wedge B]}{P[B]} \qquad (P[B] \neq 0).$$

De gebeurtenissen A en B zijn stochastisch onafhankelijk indien

(14) $$P[A|B] = P[A]$$

is. Dan is ook (vgl. (7) en (8) en de afleiding daarvan):

(15) $$P[A \wedge B] = P[A] . P[B]$$

en

(16) $$P[B|A] = P[B].$$

Formule (15) heet de vermenigvuldigingsregel voor onafhankelijke gebeurtenissen; de algemene vorm van de vermenigvuldigingsregel volgt uit (13):

(17) $$P[A \wedge B] = P[A|B] . P[B] = P[B|A] . P[A].$$

Indien de gebeurtenissen A en B elkaar uitsluiten, is $P[A \wedge B]=0$ en gaat (12) over in

(18) $$P[A \vee B] = P[A] + P[B].$$

Hieruit volgt door inductie gemakkelijk, dat voor een eindig aantal elkaar uitsluitende gebeurtenissen $A_1,\ldots,A_k$ geldt:

(19) $$P[A_1 \vee \ldots \vee A_k] = \sum_{i=1}^{k} P[A_i].$$

De geldigheid van deze formule voor ∞ veel elkaar uitsluitende gebeurtenissen volgt niet uit de tot nu toe gegeven axiomas en wordt als 5^e axioma toegevoegd.

Ax.5: Indien $A_1, A_2, \ldots$ een aftelbare rij elkaar uitsluitende gebeurtenissen is, en A de gebeurtenis voorstelt, dat één van deze rij gebeurtenissen optreedt (notatie $A= \bigcup_{i=1}^{\infty} A_i$), dan geldt:

(20) $$P[A] = P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i].$$

De axiomas hebben steeds betrekking op een verzameling gebeurtenissen, die op kunnen treden als resultaten van een bepaald experiment- of waarnemingencomplex E. Deze verzameling is van geval tot geval verschillend en wordt, indien er een stelsel whn op gedefinieerd is, dat aan de bovenstaande axiomas voldoet, een wh-veld genoemd (wij gebruiken voor een wh-veld het symbool Γ). Het wh-veld is het mathematische model van het beschouwde experiment.

Voorbeeld. Bij het werpen van een munt bestaat Γ uit 2 elementen, de gebeurtenissen "kruis" (K) en "munt" (M). Wij geven de whn daarvan in de regel aan met p en q (volgens ax.4 is dus $p+q=1$). Indien $p=q=\frac{1}{2}$ is wordt de munt "zuiver" genoemd. Een voorbeeld van een onzuivere munt is een punaise.

Opmerking: Ax.2 en 3 kunnen, indien Γ oneindig is, niet worden omgekeerd. Gebeurtenissen met wh 0 zijn dan niet onmogelijk, met wh 1 niet noodzakelijk.

9. Optel- en vermenigvuldigingsregel.

Stelling 1: Voor n willekeurige gebeurtenissen $A_1,\dots,A_n$ van Γ geldt:

(21) $$P[\bigcup_{i=1}^{n} A_i] = \sum_i P[A_i] - \sum_{i<j} P[A_i \wedge A_j] + \sum_{i<j<k} P[A_i \wedge A_j \wedge A_k] - \dots + (-1)^{n-1} P[\bigcap_{i=1}^{n} A_i],$$

waarin $\bigcup\limits_{i=1}^{n} = A_1 \vee A_2 \vee \dots \vee A_n$ en $\bigcap\limits_{i=1}^{n} A_i = A_1 \wedge A_2 \wedge \dots \wedge A_n$ is.

Bewijs door volledige inductie. Voor n=2 gaat deze stelling over in ax.4. Stel (21) geldt voor n=h, dan is voor n=h+1:

$$P[\bigcup_{i=1}^{h+1} A_i] = P[\bigcup_{i=1}^{h} A_i \vee A_{h+1}] = P[\bigcup_{i=1}^{h} A_i] + P[A_{h+1}] - P[\bigcup_{i=1}^{h} A_i \wedge A_{h+1}] =$$

$$= \sum_{i=1}^{h} P[A_i] - \sum_{i<j}^{h} P[A_i \wedge A_j] + \dots + (-1)^{h-1} P[\bigcap_{i=1}^{h} A_i] + P[A_{h+1}] +$$

$$- P[\bigcup_{i=1}^{h} (A_i \wedge A_{h+1})] = \sum_{i=1}^{h+1} P[A_i] - \sum_{i<j}^{h} P[A_i \wedge A_j] + \dots + (-1)^{h-1} P[\bigcap_{i=1}^{h} A_i] +$$

$$- \sum_{i=1}^{h} P[A_i \wedge A_{h+1}] + \sum_{i<j}^{h} P[A_i \wedge A_{h+1} \wedge A_j] + \dots +$$

$$+ (-1)^{h} P[\bigcap_{i=1}^{h} (A_i \wedge A_{h+1})] = \sum_{i=1}^{h+1} P[A_i] - \sum_{i<j}^{h+1} P[A_i \wedge A_j] + \dots +$$

$$+ (-1)^{h} P[\bigcap_{i=1}^{h+1} A_i].$$ q.e.d.

Stelling 2: Voor n willekeurige gebeurtenissen $A_1,\dots,A_n$ van Γ geldt:

(22) $$P[\bigcap_{i=1}^{n} A_i] = P[A_1] . P[A_2 | A_1] . P[A_3 | A_1 \wedge A_2] \dots P[A_n | \bigcap_{i=1}^{n-1} A_i].$$

Deze stelling volgt door volledige inductie uit definitie (13). De voorwaarden A_1, $A_1 \wedge A_2$, enz., worden ondersteld positieve whn te bezitten. Indien dit niet het geval is, is het linkerlid gelijk aan 0.

Deze stellingen worden aangeduid als de algemene optel- en vermenigvuldigingsregel.

10. Stochastische grootheden. De kenmerken op een wh-veld kunnen, behalve qualitatief (kruis of munt, kleuren, deugdelijk of ondeugdelijk, etc.) ook quantitatief zijn (leeftijd, lengte, sterkte, levensduur, doorslagspanning). Indien een grootheid $\underline{x}$ voor de verschillende eln van een wh-veld Γ verschillende waarden aanneemt, bezit deze grootheid een wh-verdeling. Hiermee wordt bedoeld, dat voor ieder reëel getal a de kans $P[\underline{x} \leqq a]$, dat $\underline{x}$ een waarde $\leqq a$ aanneemt, op Γ gedefinieerd is. Dit wordt aangegeven door onderstreping van het symbool, dat voor die grootheid wordt gebruikt en de grootheid wordt dan een stochastische grootheid genoemd. Hetzelfde symbool, zonder onderstreping, wordt dan gebruikt, om waarden aan te geven, die de grootheid kan aannemen of heeft aangenomen, of als coördinaat op een as, waarop het wh-veld wordt afgebeeld. De functie

$$F(x) \overset{\text{def}}{=} P[\underline{x} \leqq x] \tag{23}$$

wordt de verdelingsfunctie van $\underline{x}$ (soms: cummulatieve verdelingsfunctie van $\underline{x}$) genoemd. $F(-\infty)=0$, $F(+\infty)=1$.

Voorbeelden. Een dobbelsteen. Kenmerk $\underline{x}$: het aantal ogen; $\underline{x}$ kan dus de waarden $1,2,\dots,6$ aannemen. Γ bestaat uit 6 eln, met $1,2,\dots,6$ als kenmerken en $p_1,p_2,\dots,p_6$ ($\Sigma p_i=1$) als whn (indien $p_i=1/6$ voor $i=1,\dots,6$ is, wordt de dobbelsteen "zuiver" genoemd). $F(x)$ heeft nu de volgende gedaante:

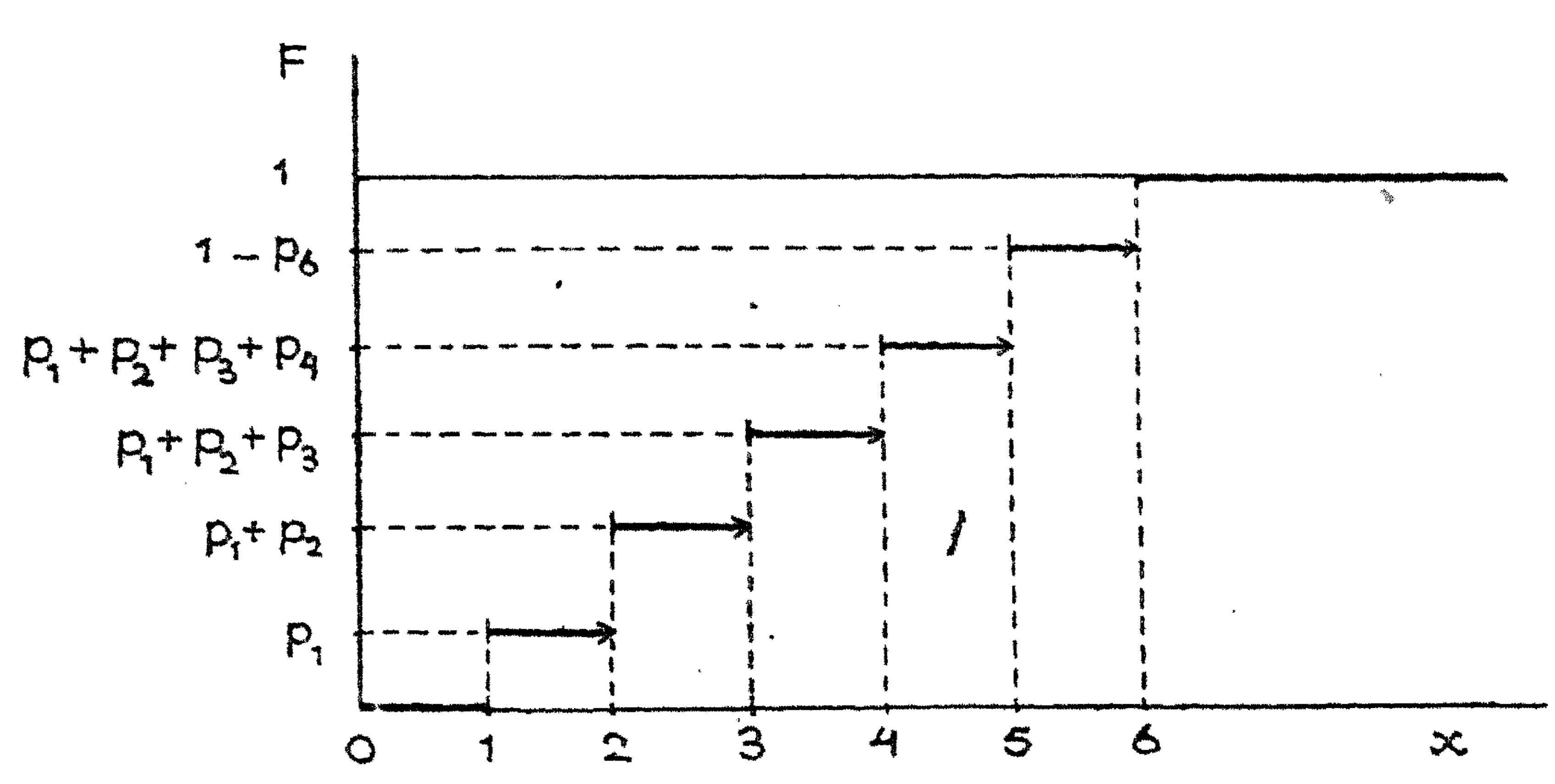


fig. 2: F(x) voor een dobbelsteen.

Een ander voorbeeld krijgen wij, indien wij bij een munt het kenmerk K door 1 en het kenmerk M door 0 vervangen. De verdelingsfunctie wordt dan als in fig.3:

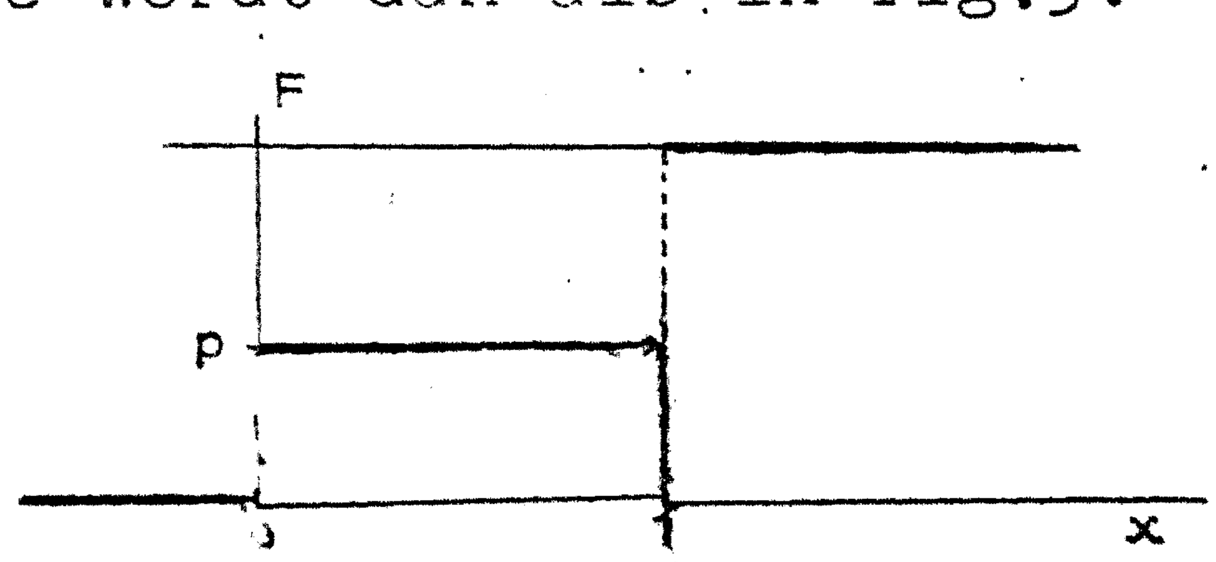


fig. 3: F(x) voor een munt.

Deze laatste stochastische grootheid, die de waarden 1 en 0 met whn p en q ($p+q=1$) aanneemt, wordt een alternatieve stochastische grootheid genoemd. Het bijbehorende wh-veld Γ

heet een alternatief.

Deze voorbeelden lijken wellicht wat triviaal, maar vooral het laatste zal zeer belangrijk blijken te zijn.

De schaal van waarden, die x kan aannemen is, bij deze voorbeelden, discreet. Men spreekt dan van een discrete wh-verdeling en van discreet verdeelde stochastische grootheden. Het aantal verschillende waarden, dat aangenomen kan worden, behoeft niet eindig te zijn; ook aftelbaar oneindig veel waarden worden toegelaten. Indien een stochastische grootheid een continuum van waarden kan aannemen (b.v. lengte, gewicht, treksterkte, etc.) en de wh-verdeling aan de eis voldoet, dat $F(x)$ een continue afgeleide

(24) $$f(x) \overset{\text{def}}{=} \frac{d}{dx} F(x)$$

bezit in alle punten x op hoogstens eindig veel na [9]), dan wordt x een continu verdeelde stochastische grootheid genoemd. De functie $f(x)$ heet de verdelingsdichtheid van x.

Eén der belangrijkste continue wh-verdelingen is de normale verdeling (vaak verdeling van GAUSZ, of van GAUSZ-LAPLACE genoemd; echter afkomstig van DE MOIVRE). De grootheid x is normaal verdeeld, indien

(25) $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \qquad (-\infty < x < +\infty)$$

is, waarin μ en σ parameters zijn. De functies $f(x)$ en $F(x)$ zullen wij in dit geval met $f_N(x)$ en $F_N(x)$ (N=normaal) aangeven. Zij zijn in fig. 4 geschetst.

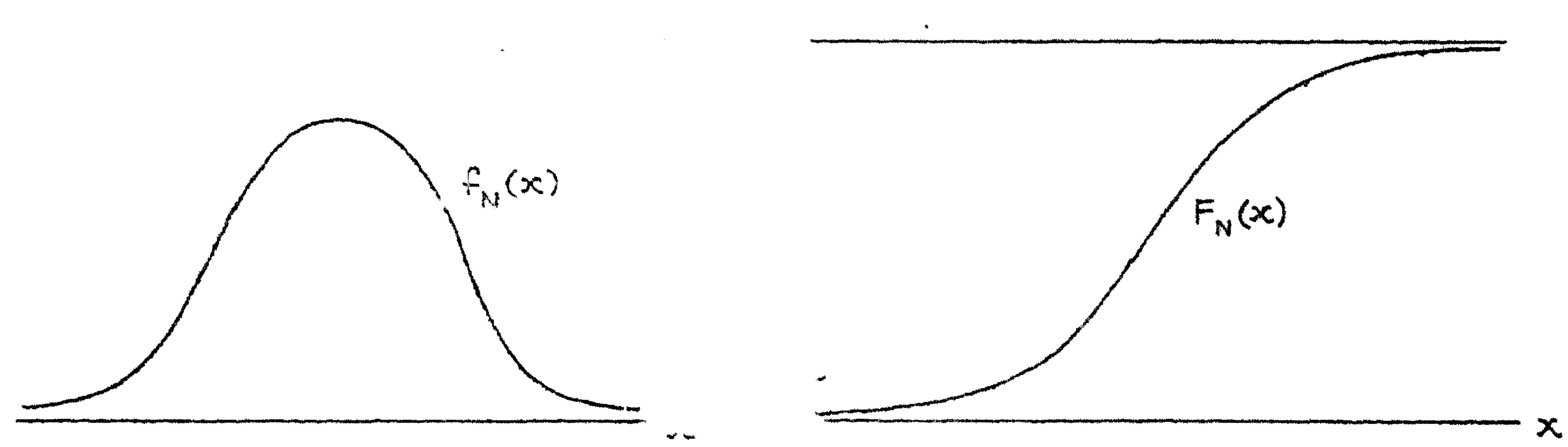


fig. 4: Schets van $f_N(x)$ en $F_N(x)$
(normale verdeling).

Zowel $f_N(x)$ als $F_N(x)$ zijn uit en te na getabelleerd. Men vindt deze tabellen in vrijwel ieder leerboek en in iedere statistische tabellenverzameling.

---

9) Deze beperking, die sterker is dan strikt nodig is, is voor de eenvoud ingevoerd. De continue verdelingen, die wij zullen ontmoeten, voldoen er aan.

Behalve discrete en continue verdelingen komen ook gemengde wh-verdelingen voor. Dit zijn wh-verdelingen, die in een continu en een discreet gedeelte kunnen worden gesplitst. Zij ontstaan veelal door gedeeltelijke groepering van waarden van een continu verdeelde grootheid. Voorbeeld: bij de bepaling van de levensduur van gloeilampen (een continu verdeelde grootheid) wordt vaak een bovengrens opgelegd aan de tijdsduur van de proef; van de lampen, die dan nog niet doorgebrand zijn, weet men dus alleen maar, dat hun levensduur groter dan een zekere waarde is, hetgeen als een discreet kenmerk kan worden opgevat. Ook op andere wijze kunnen gemengde verdelingen ontstaan. Er kunnen b.v. lampen zijn, die helemaal niet willen branden (levensduur 0), zodat er een positieve wh is voor het optreden van de waarde 0. Een ander voorbeeld is de uit te betalen brandschade bij een verzekering. Deze heeft een vrijwel continue wh-verdeling met aan de bovenkant een discrete grens: de maximale uitkering, die een positieve wh heeft.

Er zijn nog ingewikkelder types van wh-verdelingen te bedenken, die wij echter buiten beschouwing laten.

11. Momenten.

Het k^e moment van een discreet verdeelde stochastische grootheid $\underline{x}$, die de waarden x_i $(i=1,2,\ldots)$ aanneemt met whn p_i $(\Sigma\, p_i=1)$, wordt gedefinieerd als

(26) $$\mu_k = \mu_k\{\underline{x}\} \stackrel{\text{def}}{=} \sum_i p_i x_i^k .$$

Indien $\underline{x}$ continu verdeeld is, met verdelingsdichtheid $f(x)$, is de definitie

(27) $$\mu_k = \mu_k\{\underline{x}\} \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} x^k f(x)\,dx .$$

De som en de integraal kunnen divergent zijn; in dat geval bestaat het k^e moment niet. Het eerste moment heet het gemiddelde. Het nulde moment is steeds gelijk aan 1.

Voorbeelden:

Alternatieve stochastische grootheid (kans p op 1 en q op 0):

$$\mu_1 = 0.q + 1.p = p$$

$$\mu_2 = 0^2.q + 1^2.p = p.$$

Dobbelsteen (kans p_i op i; i=1,...,6):

$$\mu_1 = p_1 + 2p_2 + \ldots + 6p_6 \text{ ;voor } p_i = 1/6 \quad \mu_1 = 3\tfrac{1}{2}.$$

$$\mu_2 = p_1 + 2^2 p_2 + \quad + 6^2 p_6 \text{ :voor } p_i = 1/6 \quad \mu_2 = 9\tfrac{1}{6}.$$

Normale verdeling (zie (25)):

$$\mu_1 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx,$$

substitueer $y = \frac{x-\mu}{\sigma}$; $x = \sigma y + \mu$, $dx = \sigma dy$.

$$\mu_1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu) e^{-\frac{1}{2}y^2} dy =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y e^{-\frac{1}{2}y^2} dy + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} dy.$$

De eerste term is gelijk aan 0 wegens de symmetrie t.o.v. 0 van $e^{-\frac{1}{2}y^2}$. Verder geldt:

(28) $$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} dy = 1$$

(Bewijs: noem het linkerlid X, dan is

$$X^2 = \frac{1}{2\pi} \iint_{-\infty}^{+\infty} e^{-\frac{1}{2}(y^2+z^2)} dy dz =$$

$$= \frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_0^{\infty} \tau e^{-\frac{1}{2}\tau^2} d\tau = \int_0^{\infty} \tau e^{-\frac{1}{2}\tau^2} d\tau =$$

$$= \int_0^1 d e^{-\frac{1}{2}\tau^2} = 1.)$$

Dus $\mu_1 = \mu$.

Verder is:

$$\mu_2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu)^2 e^{-\frac{1}{2}y^2} dy =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{1}{2}y^2} dy + \mu^2 \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} dy =$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y \, d e^{-\frac{1}{2}y^2} + \mu^2 =$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \left\{ \left[y e^{-\frac{1}{2}y^2} \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} dy \right\} + \mu^2 =$$

$$= \sigma^2 + \mu^2.$$

De momenten van gemengde verdelingen worden geheel analoog gedefinieerd en berekend, na de wh-verdeling gesplitst te hebben in een continu en een stochastisch deel. Zij zijn dus gelijk aan de som van een reeks als (26) (maar nu met $\sum p_\nu \leq 1$) en een integraal als (27) (die nu voor k=0 $\leqq 1$ is). Wij zull deze gemengde verdelingen in het volgende gewoonlijk buiten beschouwing laten.

12. <u>Mathematische verwachting</u>.

De mathematische verwachting van een functie $\varphi(\underline{x})$ van een stochastische grootheid $\underline{x}$, wordt gedefinieerd door:

(29) $$\mathcal{E}\varphi(\underline{x}) \overset{\text{def}}{=} \sum_i \varphi(x_i).p_i \quad \text{resp.} \quad \int_{-\infty}^{+\infty} \varphi(x)f(x)dx.$$

Wij hebben dus (vgl. (26) en (27)):

(30) $$\mu_k = \mathcal{E}\underline{x}^k.$$

Een algemene eigenschap van de mathematische verwachting, die direct uit de definitie volgt, is:

(31) $$\mathcal{E}\sum_j \varphi_j(\underline{x}) = \sum_j \mathcal{E}\varphi_j(\underline{x}),$$

geldig indien één der leden van deze vergelijking eindig is. Verder geldt, als c een constante is:

(32) $\mathcal{E}c = c$ en (33) $\mathcal{E}c\underline{x} = c.\mathcal{E}\underline{x}$.

Een stochastische grootheid $\underline{x}$ gaat over in de <u>gereduceerde</u> $\underline{\tilde{x}}$, indien men hem met zijn verwachting (gemiddelde dus) vermindert:

(34) $$\underline{\tilde{x}} \overset{\text{def}}{=} \underline{x} - \mathcal{E}\underline{x}.$$

De <u>variantie</u> (of het <u>spreidingskwadraat</u>) Var $\underline{x}$ van een stochastische grootheid $\underline{x}$ wordt gedefinieerd door

(35) $$\text{Var}\,\underline{x} \overset{\text{def}}{=} \mathcal{E}\underline{\tilde{x}}^2.$$

Een andere, veelgebruikte notatie voor Var $\underline{x}$ is: $\sigma^2\{\underline{x}\}$; $\sigma\{\underline{x}\}$ wordt de <u>spreiding</u> (ook wel <u>standaarddeviatie</u> of <u>standaard-fout</u>) genoemd. Wij hebben volgens (31) en (34):

$$\mathcal{E}\underline{\tilde{x}}^2 = \mathcal{E}(\underline{x} - \mathcal{E}\underline{x})^2 = \mathcal{E}\{\underline{x}^2 - 2\underline{x}\,\mathcal{E}\underline{x} + (\mathcal{E}\underline{x})^2\} =$$

$$= \mathcal{E}\underline{x}^2 - 2\,\mathcal{E}\underline{x}.\mathcal{E}\underline{x} + (\mathcal{E}\underline{x})^2 = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2 = \mu_2 - \mu_1^2.$$

Dus

(36) $$\sigma^2\{\underline{x}\} = \mu_2\{\underline{x}\} - \mu_1^2\{\underline{x}\}.$$

Verder hebben wij de gemakkelijk te bewijzen eigenschappen

(36a) $\sigma^2\{\underline{x}+c\} = \sigma^2\{\underline{x}\}$ en $\sigma^2\{c\underline{x}\} = c^2\sigma^2\{\underline{x}\}$.

<u>Voorbeeld</u>.

Voor een alternatieve stochastische grootheid is

$$\sigma^2 = p - p^2 = pq \qquad (q = 1-p).$$

Voor de normale verdeling is:

$$\mathcal{E}^2\{\underline{x}\} = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

Vandaar het gebruik van de letters μ en σ in de formule voor de normale verdeling. Een normale verdeling met gemiddelde μ en spreiding σ geven wij aan met het symbool N(μ,σ).

De grootheid $\underline{x}$ wordt <u>gestandaardiseerd</u> door de gereduceerde grootheid $\underline{\tilde{x}}$ door zijn spreiding te delen.

Een <u>gereduceerde</u> stochastische grootheid heeft dus steeds de vorm

$$\frac{\underline{x} - \mathcal{E}\underline{x}}{\sigma\{\underline{x}\}} \tag{37}$$

en heeft als gemiddelde 0 en als spreiding 1.

13. Ongelijkheid van BIENAYMÉ-TCHEBYCHEFF.

Voor iedere stochastische grootheid $\underline{x}$ geldt:

$$P[|\underline{\tilde{x}}| > k\sigma] < \frac{1}{k^2}. \qquad (k \geqq 0;\ \sigma = \sigma\{\underline{x}\}). \tag{38}$$

<u>Bewijs</u>: Als $\underline{x}$ discontinu verdeeld is en de waarden x_i aanneemt met whn p_i, zodat $\underline{\tilde{x}}$ de waarden $\tilde{x}_i = x_i - \mathcal{E}\underline{x}$ met deze whn aanneemt, geldt:

$$\sigma^2 = \sum_i \tilde{x}_i^2 \cdot p_i = \sum_{\tilde{x}_i < -k\sigma} + \sum_{\tilde{x}_i > k\sigma} + \sum_{-k\sigma \leqq \tilde{x}_i \leqq k\sigma} >$$

$$> k^2\sigma^2 \Big\{ \sum_{\tilde{x}_i < -k\sigma} p_i + \sum_{\tilde{x}_i > k\sigma} p_i \Big\} = k^2\sigma^2 \cdot P[|\underline{\tilde{x}}| > k\sigma].$$

Hieruit volgt (38). Voor continu verdeelde grootheden verloopt het bewijs analoog.

De ongelijkheid is zeer onscherp en is voornamelijk van theoretisch belang. Wij behandelen in § 16 een zeer belangrijke toepassing.

<u>Voorbeelden</u>.

Voor een alternatieve stochastische grootheid $\underline{x}$, die de waarden 0 en 1 met whn $\frac{1}{2}$ aanneemt, is $\mathcal{E}\underline{x} = \frac{1}{2}$ en $\sigma\{\underline{x}\} = \frac{1}{2}$. Derhalve geldt voor $k < 1$, resp. $\geqq 1$: $P[|\underline{\tilde{x}}| > k\sigma] = 1$ resp. 0.
Hieruit blijkt, dat de ongelijkheid (38) niet kan worden verscherpt, zonder extra onderstellingen te maken.

Voor de normale verdeling geldt, zoals uit een tabel blijkt: $P[|\underline{\tilde{x}}| > 3\sigma] = 0{,}0027$. De ongelijkheid (38) geeft als resultaat slechts, dat deze wh $< \frac{1}{9}$ is.

14. Hulpstelling: Het aantal verschillende verdelingen van k objecten in twee groepen van h resp. k-h objecten.bedraagt

(39) $$\binom{k}{h} = \frac{k!}{h!(k-h)!}$$

Bewijs. Wij leggen de k objecten op een rij en nemen de eerste h als de ene groep. Er zijn k! rangschikkingen, maar daar de h! permutaties van de eerste h objecten in de rij en eveneens de (k-h)! van de overige aan de verdeling in twee groepen niets veranderen, moet het aantal rangschikkingen door h!(k-h)! gedeeld worden, om het aantal verdelingen in twee groepen te verkrijgen. H.u.v. (39).

Opmerking: is $h=k-h=\frac{1}{2}k$, dan geldt het lemma slechts, indien verwisseling van de twee groepen geacht wordt tot een nieuwe verdeling te leiden. M.a.w. men moet dan een eerste en een tweede groep onderscheiden.

15. De binomiale verdeling.

Wij beschouwen nu een reeks experimenten, zoals de in § 4 bij de experimentele wet der grote getallen besprokene. Bij ieder experiment treedt een bepaald kenmerk S al of niet op. Het wel optreden van S noemt men vaak, om wat fleurigheid in de terminologie te brengen, een succes, het niet optreden van een succes ($\bar{S}$) een mislukking. Wij onderstellen, dat de kans op een succes bij ieder der experimenten dezelfde is, en wel gelijk aan p, en dat de experimenten stochastisch onafhankelijk zijn, hetgeen dus betekent, dat de kans op een succes bij een bepaald experiment dezelfde is, wat ook de uitslag van de vorige experimenten geweest moge zijn (vgl.(14)). Deze onderstelling van onafhankelijkheid komt in de practijk overeen met een herhaling van het experiment in dezelfde, van tevoren vastgestelde, vorm, zonder beïnvloeding van deze vorm door de eerder behaalde resultaten. Het aantal successen in de reeks noemen wij $\underline{x}$. Indien de reeks uit slechts één experiment bestaat, bezit $\underline{x}$ een alternatieve verdeling. Bij een langere reeks behoort bij het $i^{\underline{e}}$ experiment een alternatieve stochastische grootheid $\underline{x}_i$, die het aantal successen (0 of 1) van dat experiment aangeeft, en wij hebben dan

(40) $$\underline{x} = \sum_i \underline{x}_i ,$$

dus ook

(41) $$\mathcal{E}\underline{x} = \sum_i \mathcal{E}\underline{x}_i = np ,$$

indien p de kans op een succes is (dus $E\underline{x}_i = 0.q+1.p=p$) en de reeks experimenten de lengte n heeft. De grootheden $\underline{x}_i$ zijn, volgens de gemaakte onderstelling, onderling onafhankelijk verdeeld. Wij leiden nu de wh-verdeling van $\underline{x}$ af en wel zullen wij de volgende stelling bewijzen:

Stelling 3. Bij een reeks van n onafhankelijke experimenten, die ieder een wh p op succes hebben, bezit het aantal successen $\underline{x}$ een wh-verdeling van de vorm

$$P[\underline{x}=x] = \binom{n}{x} p^x q^{n-x} \qquad (q=1-p). \tag{42}$$

Opmerkingen: formule (42) geldt voor iedere gehele waarde van x, indien wij $\binom{n}{x}=0$ stellen voor $x<0$ en $x>n$. De door (42) aangegeven wh-verdeling wordt de binomiale verdeling of de verdeling van Bernoulli genoemd.

Bewijs van de stelling.

De kans, dat er een bepaalde reeks uitkomsten van de vorm $S\bar{S}\bar{S}S\ldots S$ verkregen wordt, waaronder x successen, is gelijk aan $pqqp\ldots p=p^x q^{n-x}$. Dit volgt direct uit de vermenigvuldigingsregel (15) voor onafhankelijke gebeurtenissen. Er zijn, volgens de hulpstelling van § 14, $\binom{n}{x}$ verschillende dergelijke reeksen. Het optreden van twee verschillende reeksen tegelijk is onmogelijk, zodat wij formule (19) kunnen toepassen. Dan blijkt, dat de kans op het verkrijgen van één van deze reeksen gelijk is aan het rechterlid van (42), maar het is tevens de kans om x successen te verkrijgen, dus gelijk aan het linkerlid van (42).

Het gemiddelde van $\underline{x}$ hebben wij reeds berekend (zie (42)). Wij berekenen nu het tweede moment langs een omweg.

$$\mathcal{E}\underline{x}(\underline{x}-1) = \sum_{i=0}^{n} i(i-1)\binom{n}{i} p^i q^{n-i} = \sum_{i=2}^{n} \frac{n!}{(i-2)!\,(n-i)!}\, p^i q^{n-i} =$$

$$= n(n-1)p^2 \sum_{i=2}^{n} \binom{n-2}{i-2} p^{i-2} q^{(n-2)-(i-2)} =$$

$$= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j q^{n-2-j} = n(n-1)p^2.$$

Verder is

$$\mathcal{E}\underline{x}(\underline{x}-1) = \mathcal{E}(\underline{x}^2-\underline{x}) = \mathcal{E}\underline{x}^2 - \mathcal{E}\underline{x} = \mathcal{E}\underline{x}^2 - np,$$

dus

$$\mathcal{E}\underline{x}^2 = n(n-1)p^2+np$$

en

$$\mathcal{E}\underline{\tilde{x}}^2 = \sigma^2\{\underline{x}\} = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2 = n(n-1)p^2+np-n^2p^2 =$$

$$= np-np^2 = npq.$$

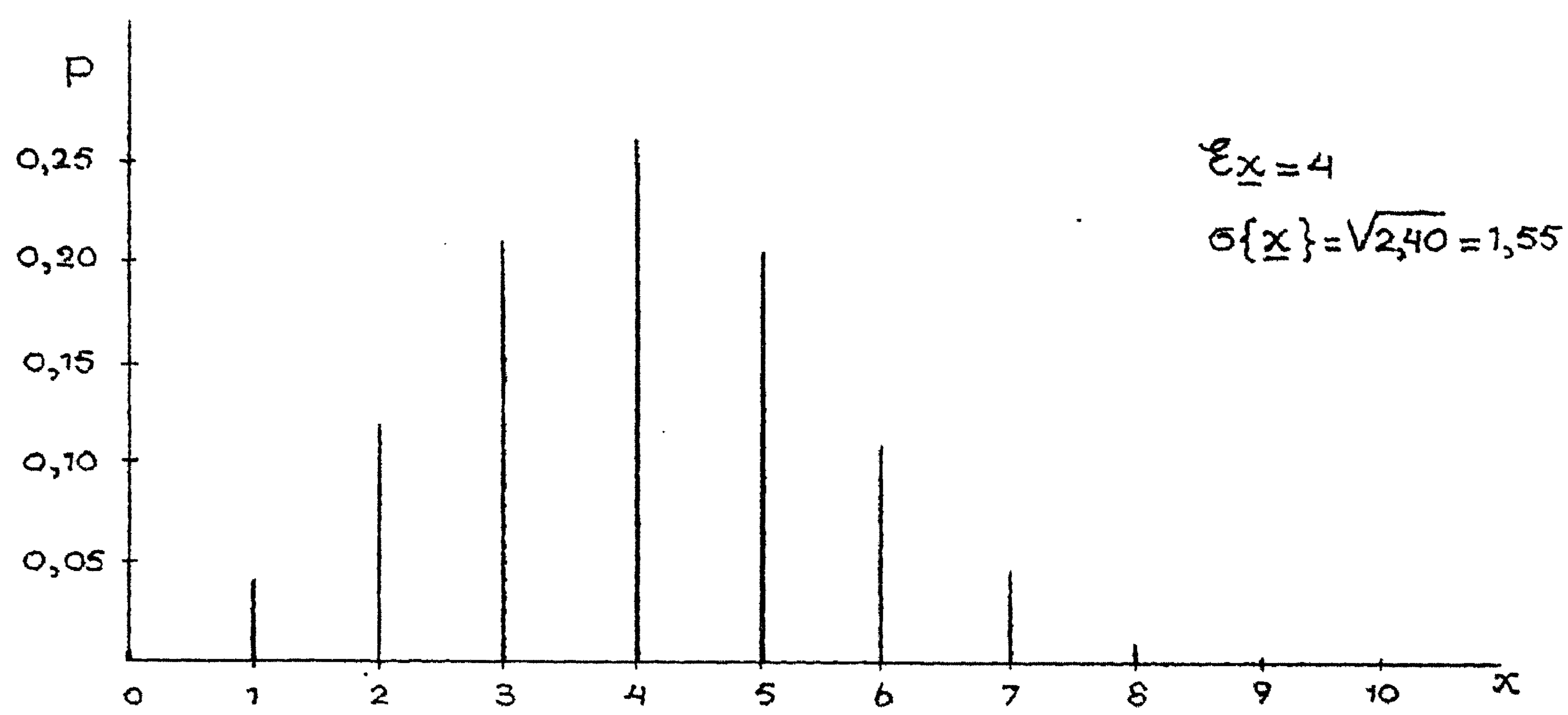
Wij hebben dus:

(41) $\mathcal{E}\underline{x} = np$ en $\mathcal{E}\underline{\tilde{x}}^2 = npq$. (43).

<u>Voorbeeld</u>. Voor n=10 en p=0,4 is de binomiale verdeling in onderstaande tabel en grafiek gegeven.

<u>Binomiale verdeling voor n=10, p=0,4</u>.

x	P[$\underline{x}$=x]	x	P[$\underline{x}$=x]
0	0,006	5	0,201
1	0,040	6	0,111
2	0,121	7	0,042
3	0,215	8	0,011
4	0,251	9	0,002
		10	0,0001



16. <u>De wet der grote getallen</u> (theoretisch).

Volgens de ongelijkheid van BIENAYMÉ-TCHEBYCHEFF, (38) geldt voor de binomiale verdeling

$$P[|\underline{x} - np| > k\sqrt{npq}\,] < \frac{1}{k^2}.$$

Het frequentiequotiënt $\underline{x}/n$ van het aantal successen fq($\underline{x}$) noemende, verkrijgen wij hiervoor:

$$P\left[|fq(\underline{x}) - p| > k\sqrt{\frac{pq}{n}}\right] < \frac{1}{k^2}$$

of ook (met $h = k\sqrt{\frac{pq}{n}}$):

$$P[|fq(\underline{x}) - p| > h\,] < \frac{pq}{h^2 n}.$$

Dit betekent dus, dat bij iedere positieve h, de kans op een verschil groter dan h tussen het gevonden fq en de wh p voor $n \to \infty$ tot 0 nadert. Dit komt echter precies overeen met de experimentele wet der grote getallen, indien wij aannemen, dat een gebeurtenis, die een kleine wh bezit, gewoonlijk niet optreedt.

17. Toepassingsprincipe.

Hierop berust nu de toepassing van de whr in de statistiek. De statistiek ontwikkelt methoden, die tot niet geheel zekere conclusies leiden, maar waarbij de kans op een foute conclusie bekend is, b.v. gelijk aan α (voor α wordt vaak de waarde 1/20 genomen). Men weet dan, dat er slechts een zeer geringe kans is, om in een lange reeks van dergelijke conclusies meer dan ongeveer een fractie α aan foute conclusies te verkrijgen. Deze kleine kans wordt dan verwaarloosd, hetgeen, zoals experimenteel kan worden aangetoond (nl. met de experimentele wet der grote getallen of op grond van andere experimenten), voor toepassingen verantwoord is. Ook bij de andere in § 4 genoemde experimentele verschijnselen, waarbij bepaalde, niet onmogelijke, uitkomsten niet voorkomen, blijken, bij berekening, deze uitkomsten een zeer geringe wh te bezitten.

Dit toepassingsprincipe lijkt wellicht minder exact dan de methode van andere wetenschappen, zoals b.v. de natuurkunde, maar dit is slechts in schijn het geval. Ook de natuurkunde geeft geen absoluut zekere uitkomsten; anders zouden er geen discrepanties tussen verschillende metingen van eenzelfde grootheid en geen mislukte experimenten voorkomen. Een bromfiets zou nooit mogen weigeren. Eerder kan men zeggen, dat de statistiek exacter is, daar de aan het experiment inhaerente onzekerheid in de berekeningen betrokken wordt en tenslotte in een getal wordt uitgedrukt, dat, in een lange reeks experimenten, een nauwkeurige schatting van het fq der onjuiste conclusies of voorspellingen vertegenwoordigt.

Cursus:

Elementaire Mathematische Statistiek

S 95

Errata bij Hoofdstuk I

pag.	regel	staat	moet staan
5	4 v.b.	voor iedere A	voor A, maar ook voor B, C,... .
6	7 v.o.	van Mises	von Mises
8	13 v.b.	$\bigcup_{i=1}^{n} =$	$\bigcup_{i=1}^{n} A_i =$
9	16 v.b.	cummulatieve	cumulatieve
	fig. 3	p	q De dikke lijn langs de x-as moet bij O in een pijltje eindigen.
11	2 v.o.	$9\frac{1}{6}$	$\frac{91}{6}$
12	3 v.o.	stochastisch	discreet
14	9 v.b.	gereduceerde	gestandaardiseerde
15	5 v.o.	,	.
	1,2,3, 4 v.o.	"dus ook...lengte n heeft" weglaten.	
16	7 en 9 v.b. 13 en 15 v.o.	(42)	(41)
	12 v.o.	Voor "Het gemiddelde... (zie (42))" komt in de plaats:	

Het gemiddelde van $\underline{x}$ is nu, volgens (26) en (41):

$$\mathcal{E}\underline{x} = \sum_{i=0}^{n} i\binom{n}{i} p^i q^{n-i} = \sum_{i=1}^{n} \frac{n!}{(i-1)!\,(n-i)!} p^i q^{n-i} =$$

$$= np\sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} q^{n-i} = np\sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} = np,$$

daar $\sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} = (p+q)^{n-1} = 1$

is. Dus

(42) $\mathcal{E}\underline{x} = np.$

pag.	regel	staat	moet staan
17	8 v.o.	met $h = k\sqrt{\frac{pq}{n}}$	met $k = h\sqrt{\frac{n}{pq}}$

MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 95 (R3)

Cursus:

Elementaire Mathematische Statistiek

Hoofdstuk II

Binomiale verdeling.

door

Prof. Dr J.Hemelrijk .

1952.

II Binomiale verdeling.

1. Binomiale verdeling.

In I § 15 is de volgende stelling bewezen.

Stelling 1: Bij een reeks van n onafhankelijke experimenten, die ieder een wh p op succes hebben, bezit het aantal successen $\underline{x}$ een wh verdeling van de vorm

(1) $$P[\underline{x}=x]=\binom{n}{x}p^x q^{n-x} \qquad (q=1-p).$$

Deze verdeling wordt de binomiale verdeling genoemd; het gemiddelde en de spreiding van $\underline{x}$ zijn:

(2) $$\mathcal{E}\underline{x}=np \qquad \mathcal{E}\underline{\tilde{x}}^2=npq.$$

2. Asymptotische normaliteit.

Wij vermelden, zonder bewijs, de volgende stelling, die voor practische toepassingen van groot belang is.

Stelling 2: Noemen wij de gestandaardiseerde binomiale grootheid $\underline{y}$, dus

(3) $$\underline{y} \overset{\text{def}}{=} \frac{\underline{x}-np}{\sqrt{npq}},$$

dan is $\underline{y}$, bij gegeven waarde van p, asymptotisch voor $n\to\infty$ normaal verdeeld, d.w.z. voor iedere y geldt:

(4) $$\lim_{n\to\infty} P[\underline{y}\leq y]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}u^2}du.$$

Deze stelling wordt gewoonlijk de stelling van LAPLACE genoemd. Zij is echter reeds in 1738 door DE MOIVRE bewezen [1]). Een uitvoerig elementair bewijs vindt men in J.NEYMAN, First course in probability and statistics, N.Y. 1950, hoofdstuk 4. De stelling volgt verder direct uit de centrale limietstelling der whr, die wij hier echter niet bespreken.

3. Aangepaste normale verdeling.

Stelling 2 schept de mogelijkheid de binomiale verdeling, die voor kleine n zeer handelbaar is, maar voor grote n practisch niet meer op directe wijze berekenbaar, door een normale te benaderen. Een normale verdeling met hetzelfde gemiddelde en dezelfde spreiding als een gegeven binomiale verdeling wordt de aan deze binomiale verdeling aangepaste normale verdeling genoemd. Volgens stelling 2 moeten deze verdelingen, voor grote n, zeer veel op elkaar gelijken. Wij gaan de aanpassing uitvoeren voor het in I.15 [2]) behandelde voorbeeld n=10, p=0,4.

---

[1]) Zie b.v. H.M.WALKER, Studies in the history of statistical method, Baltimore 1931, p. 14.

[2]) I.15= hoofdstuk I, § 15.

Voor dit geval is $\mu=np=4$ en $\sigma=\sqrt{npq}=1{,}55$.

De aangepaste normale verdeling heeft dus als verdelingsdichtheid:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \qquad \text{met } \mu=4,\ \sigma=1{,}55. \tag{5}$$

Deze grootheid is, voor $\mu=0$ en $\sigma=1$, getabelleerd, b.v. in W.J.DIXON and F.J.MASSEY Jr, Introduction to statistical analysis, N.Y.-Toronto-London 1951, p.305 [3]. De berekening der aangepaste verdeling is uitgevoerd in tabel I, waarin

$$y=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}$$

is. De derde kolom van de tabel is opgezocht in de genoemde tabel van de N(0,1)-verdeling. De laatste kolom bevat de exacte verdeling (vgl. I.14).

Tabel I

Aanpassing van een normale verdeling aan een binomiale met $n=10$, $p=0{,}4$ ($\mu=4$, $\sigma=1{,}55$) (verdelingsdichtheid als schatting der whn)

x	$\lvert z\rvert=\left\lvert\frac{x-4}{1{,}55}\right\rvert$	y	$y/1{,}55$	p
4	0	0,399	0,257	0,251
3 en 5	0,645	0,324	0,209	0,215 en 0,201
2 en 6	1,29	0,174	0,112	0,121 en 0,111
1 en 7	1,935	0,061	0,039	0,040 en 0,042
0 en 8	2,58	0,014	0,009	0,006 en 0,011
9	3,225	0,002	0,001	0,002
10	3,87	0,0002	0,0001	0,0001

De exacte en de aangepaste verdeling zijn verder in fig. 1 geschetst. De exacte is discontinu en de aangepaste continu. De overeenstemming is reeds vrij goed, ondanks het feit, dat n klein is. Daar $p\neq\frac{1}{2}$ is, is de exacte verdeling scheef, terwijl de aangepaste symmetrisch is. De benadering is daarom verreweg het beste, indien $p=\frac{1}{2}$ is.

Nu gaat het erom de exacte verdeling te schatten met behulp van de benadering. Voor $n=10$ is de binomiale verdeling nog gemakkelijk exact te berekenen en tot en met $n=49$ is zij bovendien in 7 decimalen getabelleerd (voor $p=0{,}01$ (0,01) 0,50) in de "Tables of the binomial probability distribution", National Bureau of Standards, Washington 1950. Voor grotere n en ook, als men deze tabel niet bij de hand heeft, is het gebruik van de normale benadering aan te bevelen, als p niet te sterk van 0,5 afwijkt.

---

[3]) Dit boek wordt verder als "DIXON and MASSEY" aangeduid.

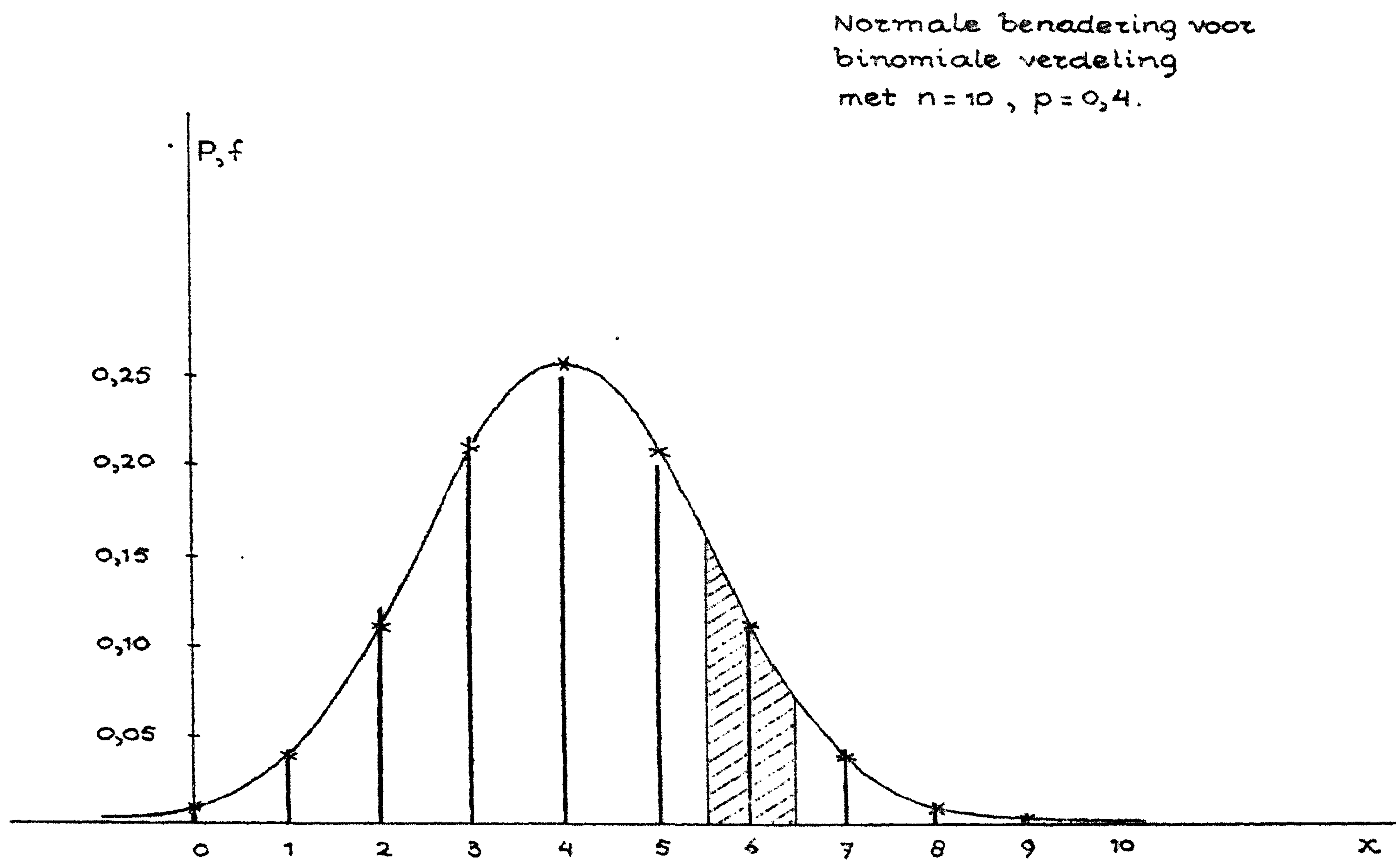


fig. 1. Aanpassing van een normale verdeling aan een binomiale met n=10 en p=0,4.

De wh P[x=x] wordt nu, zoals voor de hand ligt (zie fig. 1) benaderd door de oppervlakte onder de aangepaste normale verdeling boven het interval van $x-\frac{1}{2}$ tot $x+\frac{1}{2}$. Alleen voor de twee uiterste waarden (in ons voorbeeld 0 resp. 10) neemt men de oppervlakte vanaf $-\infty$ tot $\frac{1}{2}$ resp. van $9\frac{1}{2}$ tot $+\infty$, daar anders de som der benaderende whn kleiner dan 1 zou worden. Deze benadering door een oppervlak bezit, zoals wij verderop zullen zien, een groot voordeel boven de in tabel I, 4$^{\underline{e}}$ kolom, vermelde benadering met behulp van de ordinaten der aangepaste normale verdeling. Wij geven eerst in tabel II de benadering van de exacte whn door de genoemde oppervlakken weer. Deze kunnen worden opgezocht in een tabel van de cumulatieve verdelingsfunctie der normale verdeling (b.v. DIXON and MASSEY, pag. 306). Uit tabel II blijkt de wijze van berekenen voldoende duidelijk. Wij maken echter nog de volgende opmerkingen. De tabel is naar afdalende waarden van x gerangschikt, omdat dit de overgang van de 4$^{\underline{e}}$ op de 5$^{\underline{e}}$ kolom gemakkelijker maakt; de 5$^{\underline{e}}$ kolom bevat de verschillen van de 4$^{\underline{e}}$. De derde kolom is het gemakkelijkst te berekenen door gebruik te maken van het feit, dat de breedte van de intervallen, uitgedrukt in σ, 1/1,55=0,645 bedraagt. Ook de symmetrie van de aangepaste verdeling maakt in dit geval de berekening eenvoudig. Dit zou echter niet zo zijn, als het gemiddelde ($np = 4$) geen geheel getal was.

Tabel II

Aanpassing van een normale verdeling aan een binomiale met $n = 10$, $p = 0{,}4$ ($\mu = 4$, $\sigma = 1{,}55$)

(oppervlakken der aangepaste verdeling als schatting der whn)

x	ondergrens interval (g)	$g^* = \frac{g-4}{1{,}55}$	$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{g^*} e^{-\frac{1}{2}u^2}du$	$P[\underline{x}=x]$ benaderd	$P[\underline{x}=x]$ exact
10	9,5	3,548	0,9998	0,0002	0,0001
9	8,5	2,903	0,998	0,001	0,002
8	7,5	2,258	0,988	0,010	0,011
7	6,5	1,613	0,946	0,042	0,042
6	5,5	0,968	0,834	0,112	0,111
5	4,5	0,323	0,627	0,207	0,201
4	3,5	-0,323	0,373	0,254	0,251
3	2,5	-0,968	0,166	0,207	0,215
2	1,5	-1,613	0,054	0,112	0,121
1	0,5	-2,258	0,012	0,042	0,040
0	$-\infty$	$-\infty$	0	0,012	0,006

Bij vele toepassingen van deze benadering zijn niet de relatieve fouten van belang, maar de absolute en zolang deze niet meer dan enkele promillen bedragen zijn zij niet hinderlijk. De beide benaderingen (tabel I en II) zijn dus ongeveer even goed. Het voordeel van het gebruik van de oppervlakkenmethode (dus van de verdelingsfunctie van de aangepaste normale verdeling) is gelegen in het feit, dat het er gewoonlijk niet om gaat de whn $P[\underline{x}=x]$ ieder afzonderlijk te berekenen. Meestal moet men whn van de vorm

$$P[\underline{x} \leqq x_1]\,;\quad P[x_1 \leqq \underline{x} \leqq x_2]\,;\quad P[\underline{x} \geqq x_2]$$

berekenen. Bij de eerste benaderingsmethode zou men deze slechts verkrijgen als sommen van reeksen van whn van de vorm $P[\underline{x}=x]$, hetgeen veel te omslachtig zou zijn. Volgens de tweede benaderingsmethode hebben wij echter:

$$(6)\quad \begin{cases} P[\underline{x} \leqq x_1] \approx F_N\left(\frac{x_1 - np + \frac{1}{2}}{\sqrt{npq}}\right) \\[2ex] P[x_1 \leqq \underline{x} \leqq x_2] \approx F_N\left(\frac{x_2 - np + \frac{1}{2}}{\sqrt{npq}}\right) - F_N\left(\frac{x_1 - np - \frac{1}{2}}{\sqrt{npq}}\right) \\[2ex] P[\underline{x} \geqq x_2] \approx 1 - F_N\left(\frac{x_2 - np - \frac{1}{2}}{\sqrt{npq}}\right), \end{cases} \qquad 4)$$

---

4) $\approx$ betekent: is ongeveer gelijk aan.

waarin $F_N(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}u^2} du$ is.

Vergelijken wij dit met (4) van Stelling 2, dan zien wij, dat de eerste vergelijking van (6) met (4) overeenstemt op de de term $\frac{1}{2}$ in de teller na. Voor $n \to \infty$ telt deze term niet mee, voor eindige n echter wel. Men noemt deze term de continuïteitscorrectie, daar zij ingevoerd wordt wegens de continuïteit van de aangepaste normale verdeling, terwijl de binomiale verdeling discreet is. Bij het toepassen van de continuïteitscorrectie stelle men zich steeds fig. 1 goed voor ogen.

Voorbeeld. Een van de oudste voorbeelden op dit gebied is de kwestie der jongens- en meisjes-geboorten. Wij zullen de volgende twee vragen beantwoorden met behulp van de normale benadering.

1ᵉ Als de kans op een jongen gelijk is aan die op een meisje, hoe groot is dan de kans om bij 10.000 geboorten een aantal jongensgeboorten $\underline{x}$ tussen 4990 en 5010 te vinden?

2ᵉ In welk interval zal, in deze omstandigheden, het aantal jongensgeboorten behoudens een wh 0,05 liggen? (Deze tweede vraag is niet voor ondubbelzinnige beantwoording vatbaar;zie verderop.)

Oplossing:

1ᵉ Wij hebben dus een binomiale verdeling met $n=10.000$ en $p=q=\frac{1}{2}$. Dus $\mu=5.000$ en $\sigma=\sqrt{npq}=25$. Volgens formule (6) is dus:

$$P[4990 \leqq \underline{x} \leqq 5010] \approx F_N\left(\frac{5010-5000+\frac{1}{2}}{25}\right) - F_N\left(\frac{4990-5000-\frac{1}{2}}{25}\right) =$$

$$= F_N(0{,}42) - F_N(-0{,}42) = 1 - 2F_N(-0{,}42) = 0{,}33\,,$$

zoals bij opzoeken in een tabel van de verdelingsfunctie der $N(0,1)$-verdeling blijkt.

2ᵉ Er zijn zeer vele intervallen aan te geven, waarvoor het gevraagde geldt, daar in de regel bij een verschuiving van de ondergrens naar boven een verschuiving van de bovengrens naar boven gevonden kan worden, die de verschuiving van de ondergrens compenseert (en vice versa). Wij zullen er drie berekenen, die het meest voor de hand liggen en wel zullen wij x_1, x_1', x_2' en x_2 zo bepalen, dat voldaan is aan

$$P[\underline{x} \geqq x_1] = P[x_1' \leqq \underline{x} \leqq x_2'] = P[\underline{x} \leqq x_2] = 0{,}95\,,$$

terwijl $x_1' + x_2' = 10.000$ is, dus het interval (x_1', x_2') symmetrisch is t.o.v. het gemiddelde 5000.

Volgens formule (6) is nu (let op het teken van de continuïteitscorrectie):

$$P[\underline{x} \geqq x_1] \approx 1 - F_N\left(\frac{x_1 - 5000 - \frac{1}{2}}{25}\right)$$

en dit moet =0,95 zijn. Nu is, zoals bij opzoeken in een tabel van F_N blijkt, $F_N(-1{,}645) = 0{,}05$, dus

$$\frac{x_1 - 5000 - \frac{1}{2}}{25} = -1{,}645.$$

H.u.v. $x_1 = 4959{,}4$; daar $\underline{x}$ slechts gehele waarden aan kan nemen, vinden wij dus als oplossing:

$$P[\underline{x} \geqq 4960] \approx 0{,}95.$$

Uit symmetrie-overwegingen volgt dan tevens:

$$P[\underline{x} \leqq 5040] \approx 0{,}95.$$

Verder is:

$$P[x_1' \leqq \underline{x} \leqq x_2'] \approx F_N\left(\frac{x_2' - 5000 + \frac{1}{2}}{25}\right) - F_N\left(\frac{x_1' - 5000 - \frac{1}{2}}{25}\right) =$$

$$= 1 - 2F_N\left(\frac{x_1' - 5000 - \frac{1}{2}}{25}\right).$$

Nu is $F_N(-1{,}96) = 0{,}025$, dus

$$\frac{x_1' - 5000 - \frac{1}{2}}{25} = -1{,}96.$$

H.u.v. $x_1' = 4951{,}5$ en dus $x_2' = 5048{,}5$. Daar $\underline{x}$ alleen gehele waarden aan kan nemen, is

$$P[4952 \leqq \underline{x} \leqq 5048] \approx 0{,}95.$$

4. Exacte voorstelling door de onvolledige Bêta-functie.

Behalve uit de in § 3 genoemde tabellen der binomiale verdeling, kan men deze verdeling ook aflezen uit tabellen en nomogrammen van de onvolledige Bêta-functie.

Wij bewijzen daartoe de volgende stelling.

Lemma 1:

Voor $0 \leqq t < n$ en $p > 0, q > 0, p + q = 1$, geldt

$$(7) \qquad \sum_{i=0}^{t} \binom{n}{i} p^i q^{n-i} = 1 - \frac{\int_0^p x^t (1-x)^{n-t-1} dx}{\int_0^1 x^t (1-x)^{n-t-1} dx}.$$

(Opmerking: voor $t=n$ is het linkerlid gelijk aan 1; van de tweede term in het rechterlid zijn dan de integralen in teller en noemer divergent.)

Bewijs. Noem het linkerlid van (7) $S(t,n,p)$ (of bij afkorting: S), dan geldt:

$$\frac{dS}{dp}=\sum_{i=1}^{t}\frac{n!}{(i-1)!(n-i)!}p^{i-1}q^{n-i}-\sum_{j=0}^{t}\frac{n!}{j!(n-j-1)!}p^{j}q^{n-j-1}=$$

$$=\sum_{h=0}^{t-1}\frac{n!}{h!(n-h-1)!}p^{h}q^{n-h-1}-\sum_{j=0}^{t}\frac{n!}{j!(n-j-1)!}p^{j}q^{n-j-1}=$$

$$=-\frac{n!}{t!(n-t-1)!}p^{t}q^{n-t-1}.$$

Derhalve is

$$\int_0^p\frac{dS}{dp}\,dp=S(t,n,p)+C,$$

waarin C een integratieconstante is. Het linkerlid hiervan is gelijk aan

$$-\frac{n!}{t!(n-t-1)!}\int_0^p x^t(1-x)^{n-t-1}\,dx$$

(met voor de duidelijkheid van de notatie x als integratievariabele in plaats van p). Vullen wij hierin $p=0$ in, dan wordt dit linkerlid gelijk aan 0. Verder is in het rechterlid $S(t,n,0)=1$, daar voor $p=0$ de eerste term van (7) gelijk aan 1 wordt en de overige gelijk aan 0. Dus is

$$C=-1.$$

Wij hebben dus

$$S(t,n,p)=1-\frac{n!}{t!(n-t-1)!}\int_0^p x^t(1-x)^{n-t-1}\,dx.$$

Vullen wij hierin $p=1$ in, dan wordt het linkerlid, $S(t,n,1)$, gelijk aan 0 (daar nu $q=0$ is), zodat wij vinden

$$\frac{t!(n-t-1)!}{n!}=\int_0^1 x^t(1-x)^{n-t-1}\,dx.$$

Vullen wij dit in de vorige vergelijking in, dan verkrijgen wij (7). q.e.d.

Gewoonlijk wordt de volgende notatie gebruikt:

(8) $$B(r,s)=\int_0^1 x^{r-1}(1-x)^{s-1}\,dx$$

(9) $$B_p(r,s) = \int_0^p x^{r-1}(1-x)^{s-1}\,dx$$

(10) $$I_p(r,s) = \frac{B_p(r,s)}{B(r,s)}\ .$$

Deze laatste notatie is afkomstig van K.PEARSON, Tables of the incomplete Beta-function, London 1934, waarin deze functie $I_p(r,s)$, de onvolledige Bêta-functie, uitvoerig getabelleerd is.

In de notatie van K.PEARSON wordt (7):

(11) $$\sum_{i=0}^{t} \binom{n}{i} p^i q^{n-i} = 1 - I_p(t+1, n-t) = I_q(n-t, t+1).$$

De laatste overgang volgt uit de gemakkelijk te bewijzen eigenschap

(12) $$I_p(r,s) + I_q(s,r) = 1 \qquad (p+q=1).$$

Vervangen wij in (11) t door $u-1$ en trekken wij ieder der leden van (11) van 1 af, dan verkrijgen wij

(13) $$\sum_{i=u}^{n} \binom{n}{i} p^i q^{n-i} = I_p(u, n-u+1) = 1 - I_q(n-u+1, u).$$

Met behulp van de formules (11) en (13) kan men dus de verdelingsfunctie van de binomiale verdeling in de genoemde tafel van PEARSON opzoeken. De argumenten r en s doorlopen de waarden 1 tot en met 50, zodat wij tot en met $n=99$ kunnen komen. Deze tafel is dus uitgebreider dan de in § 3 genoemde tafel der binomiale verdeling van het "National Bureau of Standards". Voor $n \leqq 49$ is deze laatste echter veel gemakkelijker.

In dit verband verdient ook een zeer ingenieus nomogram van de onvolledige Bêta-functie vermeld te worden. Dit is te vinden in H.O.HARTLEY and E.R.FITCH, A chart for the incomplete Beta-function and the cumulative binomial distribution, Biometrika 38 (1951), pp. 423-426. Deze schrijvers hebben de functie $I_p(r,s)$ voor $r \leqq 200$ en $s \leqq 60$, die in Pearson's tafel honderden pagina's beslaat (voor r en $s \leqq 50$), op één pagina samen weten te persen.

De genoemde tabellen zijn vooral nuttig, indien n klein is en p sterk van $\frac{1}{2}$ verschilt. Voor grote n kan men gewoonlijk de normale benadering gebruiken, of als p zeer dicht bij 0 of 1 ligt een andere benadering (met behulp van een aangepaste Poissonverdeling), die wij hier niet behandelen. Voor kleine n en p dicht bij 0 of 1 zijn deze benaderingen echter vaak niet nauwkeurig genoeg.

Voorbeeld. In de kwaliteitsbeheersing heeft men vaak te maken met de keuring van grote partijen door het controleren van één of een reeks van kleine steekproeven daaruit. Ieder gecontroleerd exemplaar krijgt dan b.v. het kenmerk "deugdelijk" of "ondeugdelijk" (of "defect"), en het gaat erom zoveel mogelijk te weten te komen omtrent de fractie ondeugdelijke exemplaren in de gehele partij. Deze fractie is vaak klein en dan doen zich bij de uitwerking der statistische analyse vragen voor van het volgende type: indien het aantal gekeurde exemplaren gelijk aan 10 is en de fractie defecte exemplaren in de partij is 0,06 (6%), hoe groot is dan de kans, om onder de 10 gekeurde exemplaren 3 of meer ondeugdelijke te vinden. Indien de partij groot is en de 10 exemplaren er door loting [5]) uit gekozen worden komt deze vraag wh-theoretisch gezien neer op de vraag: hoe groot is de kans, dat de binomiaal verdeelde stochastische grootheid $\underline{x}$, met parameters $n=10$ en $p=0{,}06$, minstens de waarde 3 aanneemt. Wij moeten dan dus de grootheid

$$P[\underline{x} \geqq 3 \mid n=10,\, p=0{,}06] = \sum_{i=3}^{10} \binom{10}{i} (0{,}06)^{i} (0{,}94)^{10-i}$$

berekenen. Met behulp van de tafel van het "National Bureau of Standards" of van de in deze paragraaf beschreven tafel of nomogram van de Bêta-verdeling, vinden wij voor deze kans de waarde $0{,}019$ (of, desgewenst in meer decimalen, 0,0188378). Deze waarde is dus een afronding van de exacte waarde. Passen wij de normale benadering toe, dan hebben wij:

$$\mathcal{E}\underline{x} = 0{,}6\,, \quad \sigma\{\underline{x}\} = \sqrt{0{,}06 \times 0{,}94 \times 10} = 0{,}75\,,$$

dus $P[\underline{x} \geqq 3 \mid n=10, p=0{,}06] \approx \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{1}{2}u^2} du$ met $a = \frac{2{,}5-0{,}6}{0{,}80} = 2{,}53$. Derhalve vinden wij $P \approx 0{,}006$, een vrij slechte benadering.

----

[5]) Keuze door loting, ook "aselecte keuze" genoemd (zie D.VAN DANTZIG, De natuur als tegenspeler, Statistica 5 (1951), p. 153), is een der grondbeginselen van de toepassing van de mathematische statistiek. Men kan op die wijze bereiken, dat de waarnemingen binnen het mathematische model als onafhankelijke waarnemingen kunnen worden beschouwd. Wij gaan daar hier ter plaatse niet nader op in.

MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 95 (l3)

Cursus:

Elementaire Mathematische Statistiek

Hoofdstuk III

Principes der toetsingstheorie en de tekentoets.

door

Prof.Dr J. Hemelrijk.

1952
3e druk 1955

III Principes der toetsingstheorie en de tekentoets.

1. Hypothesen.

De toetsingstheorie, ontwikkeld door J.NEYMAN en E.S.PEARSON, houdt zich bezig met het toetsen van onderstellingen (die hypothesen worden genoemd) over niet geheel bekende wh-verdelingen, op grond van waarnemingsresultaten. Wij zullen de principes van deze theorie aan de hand van een zeer eenvoudig voorbeeld uiteenzetten.

Laat van een stochastische grootheid $\underline{x}$ gegeven zijn, dat deze een binomiale verdeling bezit:

(1) $P[\underline{x}=x] = \binom{n}{x} p^x q^{n-x} \qquad (q=1-p),$

waarin n bekend is, maar p onbekend. Wij zeggen dan, dat de verzameling van toegelaten hypothesen bestaat uit alle binomiale wh-verdelingen met de gegeven waarde van n. Deze verzameling kan dus worden voorgesteld door

(2) $\Omega : \quad 0 \leqq p \leqq 1$ (n gegeven),

de z.g. parameterruimte. Een onderstelling omtrent p van het type: $p=p_0$ wordt nu een enkelvoudige hypothese genoemd. In het algemeen is een enkelvoudige hypothese een hypothese, die de betrokken wh-verdeling volledig specificeert. Een onderstelling omtrent p van de vorm $p \leqq p_0$ (of ook $p \geqq p_1$, of $p_0 \leqq p \leqq p_1$ e.d.), waardoor de wh-verdeling niet volledig wordt gespecificeerd, wordt een samengestelde hypothese genoemd. Een dergelijke hypothese kan altijd worden voorgesteld als een deelverzameling ω van de parameterruimte Ω, b.v.

(3) $\omega : \quad p_0 \leqq p \leqq p_1 \qquad (\omega \subset \Omega),$

terwijl een enkelvoudige hypothese door één punt van Ω voorgesteld wordt.

2. Kritieke zône.

Onderstel nu b.v., dat $n=100$ is en dat wij $p=0,9$ wensen te toetsen op grond van een waarneming x_0 van $\underline{x}$. Dit betekent dus, dat onder 100 onafhankelijke proeven, ieder met een onderstelde kans 0,9 op succes, x_0 successen gevonden worden. Wanneer kan men nu zeggen, dat de gevonden waarde x_0 behoorlijk in overeenstemming is met de hypothese $p=0,9$? Intuitief is het duidelijk, dat dit zo is, als x_0 niet te ver van 90 (dit is nl. $\mathcal{E}\underline{x}$) afwijkt. Als men echter $x_0=0$ vindt, dus geen enkel succes, dan is dit resultaat wel bijzonder slecht in overeenstemming met de hypothese

$p=0{,}9$ en men zal dan geneigd zijn deze hypothese te verwerpen.

De toetsingstheorie is een precizering van dit soort redeneringen. In het bovenstaande geval b.v. zal men de hypothese $p=0{,}9$ niet alleen verwerpen, als $x_0=0$ is, maar ook als $x_0=1,2,\ldots$ tot een zekere grens, is. Met andere woorden, men kiest een deelverzameling van de verzameling van alle mogelijke resultaten en verwerpt de getoetste hypothese, indien het gevonden resultaat tot die deelverzameling behoort. Die deelverzameling wordt de kritieke zone van de toets genoemd. Valt het gevonden resultaat niet in de kritieke zone, dan wordt de hypothese niet verworpen.

De verzameling der mogelijke resultaten noemt men vaak de steekproefruimte (waarnemingen-ruimte of resultaten-ruimte zou misschien beter zijn) en de hypothese, die men toetst wordt de getoetste hypothese genoemd en gewoonlijk met H_0 aangegeven. De steekproefruimte bestaat in het door ons beschouwde geval uit getallen $0,1,\ldots,100$, nl. de waarden, die het aantal successen aan kan nemen. Een toets wordt geheel bepaald door de bijbehorende kritieke zone. De meest doelmatige keuze van een kritieke zone en de eigenschappen van de bij verschillende kritieke zones behorende toetsen zijn de onderwerpen van onderzoek van de toetsingstheorie.

3. Kritieke zone van de tekentoets.

Indien Ω, zoals in § 1, bestaat uit alle mogelijke binomiale verdelingen met gegeven n en als men de enkelvoudige hypothese $p=\frac{1}{2}$ wenst te toetsen;

(4) $\qquad H_0 : \; p=\frac{1}{2}$,

dan gebruikt men daarvoor gewoonlijk een kritieke zone van de vorm

(5) $\qquad Z : \; |x-\frac{1}{2}n| \geqq a.$

Dit betekent dus, dat men H_0 verwerpt, indien het gevonden resultaat x_0 (het aantal successen bij n proeven) meer dan een bepaald bedrag van $\frac{1}{2}n$ afwijkt. In fig. 1 is een dergelijke Z geschetst. De binomiale verdeling met $p=\frac{1}{2}$ is daarbij voor de eenvoud vervangen door de aangepaste normale verdeling.

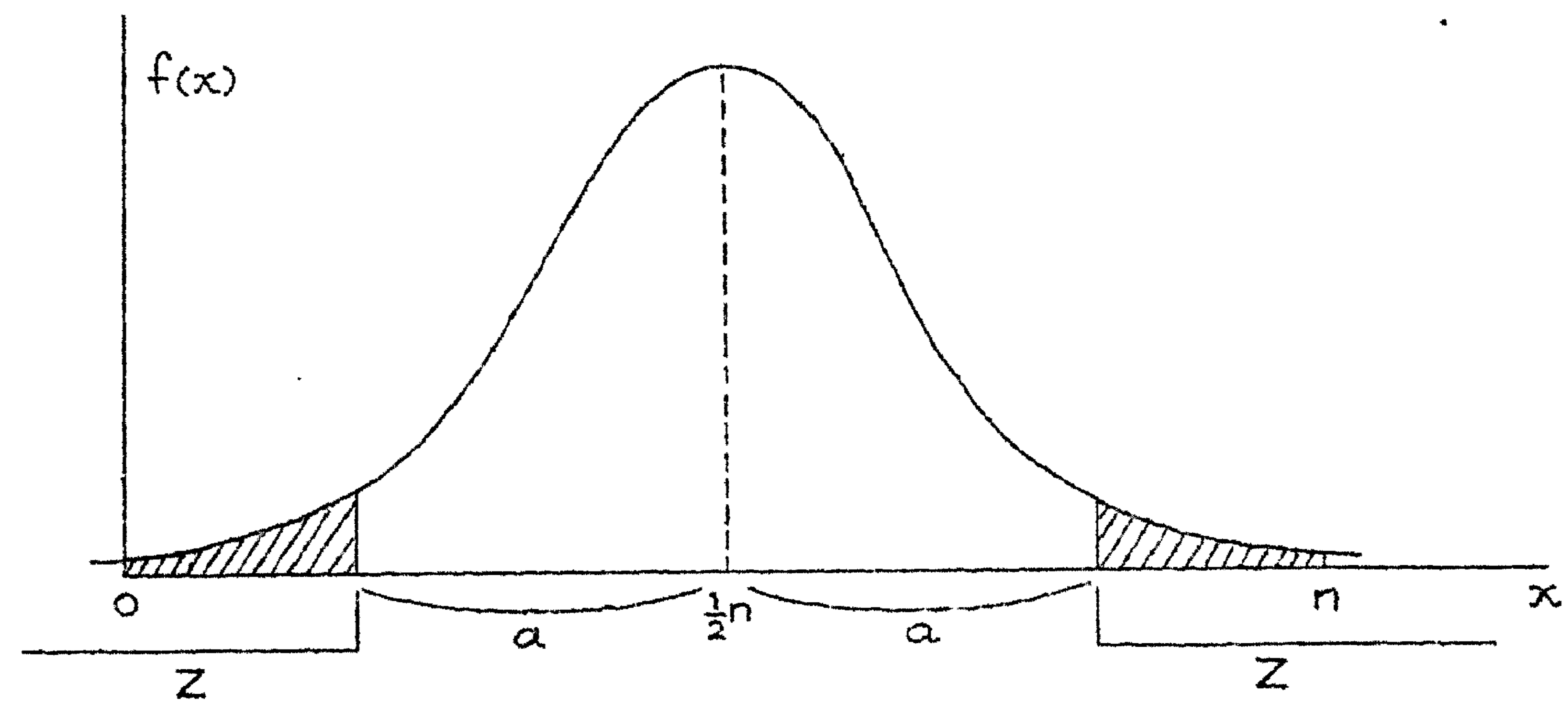


fig. 1. Z bij de tekentoets.

Wij verwerpen dus de hypothese H_o $(p=\frac{1}{2})$, als het bij de n proeven verkregen aantal successen kleiner dan of gelijk aan $\frac{1}{2}n-a$ is of groter dan of gelijk aan $\frac{1}{2}n+a$. In het eerste geval is het aantal successen te klein en trekken wij de conclusie, dat $p<\frac{1}{2}$ is, in het tweede geval te groot, hetgeen tot de conclusie $p>\frac{1}{2}$ leidt.

De voor de toets gebruikte stochastische grootheid $\underline{x}$, in dit geval het aantal successen, noemen wij de <u>toetsingsgrootheid</u>. Er kunnen ook twee of meer toetsingsgrootheden tegelijk worden gebruikt; wij beperken ons voorlopig tot één.

Wij zullen in het volgende speciaal de tekentoets beschouwen. Deze naam "tekentoets" wordt verderop duidelijk gemaakt. Behalve het door (5) gekarakteriseerde type van kritieke zones, die <u>tweezijdige kritieke zones</u> worden genoemd, worden, onder later te bespreken omstandigheden, ook <u>éénzijdige kritieke zones</u> gebruikt, van het type

(6) $\quad Z_r: \quad x-\frac{1}{2}n \geqq b$ (rechter- éénzijdige kr. zone)

en

(7) $\quad Z_l: \quad x-\frac{1}{2}n \leqq b$ (linker- éénzijdige kr. zone)

4. <u>Fout van de eerste soort</u>.

Indien wij het in § 3 beschreven procédé toepassen, moeten wij ons goed realiseren, dat verwerping van H_o niet behoeft te betekenen, dat p werkelijk $\neq \frac{1}{2}$ is. Immers ook als $p=\frac{1}{2}$ is, kan het aantal successen wel $\leqq \frac{1}{2}n-a$ of $\geqq \frac{1}{2}n+a$ zijn, zodat ook als H_o juist is, de toets tot verwerping kan leiden. In dat geval bereiken wij dus een onjuiste conclusie en wel wordt dan de hypothese H_o verworpen, hoewel zij juist is. Men noemt dit een <u>fout van de eerste soort</u> en, onderstellende dat H_o juist is, kan men de kans op het maken van deze fout van de eerste soort berekenen. Immers als H_0 juist is, is de wh-verdeling van het aantal successen ($\underline{x}$) bekend en wel is

(8) $$P[\underline{x}=x \mid H_0] = 2^{-n}\binom{n}{x},$$

waarin het symbool H_0 achter de streep aangeeft, dat H_0 ondersteld wordt juist te zijn. De kans op een fout van de eerste soort is nu bij tweezijdige toetsing:

(9) $$\beta(H_0) = 2^{-n}\sum_{i=0}^{\frac{1}{2}n-a}\binom{n}{i} + 2^{-n}\sum_{i=\frac{1}{2}n+a}^{n}\binom{n}{i} = 2^{-n+1}\sum_{i=0}^{\frac{1}{2}n-a}\binom{n}{i},$$

en bij éénzijdige toetsing:

(10) $$\beta_r(H_0) = \beta_l(H_0) = 2^{-n}\sum_{i=0}^{\frac{1}{2}n-b}\binom{n}{i},$$

als wij a en b zo kiezen, dat $\frac{1}{2}n-a$ en $\frac{1}{2}n-b$ geheel zijn.

Niet alleen bij de tekentoets geldt, dat men deze kans berekenen kan. In het algemeen geldt dit voor het toetsen van een enkelvoudige hypothese, daar een dergelijke hypothese (vgl. § 1) de wh-verdeling van de beschouwde grootheid (of grootheden) en dus ook van de toetsingsgrootheid, volledig bepaalt. Men kiest nu de kritieke zone zo, dat

(11) $$\beta(H_0) \leqq \alpha$$

is, waarin α een constante is, die de <u>onbetrouwbaarheidsdrempel</u> (Engels: "level of significance") wordt genoemd. De grootheid $\beta(H_0)$ wordt de (werkelijke) <u>onbetrouwbaarheid</u> genoemd. Indien mogelijk, maken wij deze gelijk aan α; bij discreet verdeelde toetsingsgrootheden, zoals bij de tekentoets, is dit in de regel niet mogelijk. Wij maken dan $\beta(H_0)$, om later te bespreken redenen, zo groot mogelijk, voor zover (11) dit toelaat.

Volgens de theoretische wet der grote getallen (zie I.10) kunnen wij nu het volgende zeggen: indien wij een lange reeks van stochastisch onafhankelijke toetsingen van <u>juiste</u> hypothesen uitvoeren, steeds met dezelfde onbetrouwbaarheid, dan zal het fq der verwerpingen, behoudens een zeer geringe wh, liggen tussen twee dicht om $\beta(H_0)$ gelegen grenzen. Volgens het in I.17 besproken "uitschakelings"- of "toepassingsprincipe" (of ook volgens de experimentele wet der grote getallen) beheersen wij dus het aantal fouten van de eerste soort. Het fq daarvan op de collectie van alle mogelijke toetsingen van juiste hypothesen is $\beta(H_0) \leqq \alpha$. Indien de getoetste hypothese onjuist is, kan men geen fout van de eerste soort maken. Op de collectie van alle mogelijke toetsingen is het fq van fouten van de eerste soort dus a fortiori $\leq \alpha$.

Voor α neemt men veelal de waarde 0,05 ; dit berust op traditie. Indien men het ten onrechte verwerpen van een juiste hypothese "één op de twintig maal" te ernstig acht, kan men een kleinere waarde kiezen. Soms neemt men ook een grotere, b.v. 0,1. Bij bepaalde toepassingen behoeft daar geen bezwaar tegen te bestaan. Voorbeeld: indien verwerping van een getoetste hypothese H_0 bij een industrieel procédé beschouwd wordt als een aanwijzing voor een ongunstige wending in dit procédé en aanleiding geeft tot voorzorgsmaatregelen, kan het zeer wel acceptabel zijn, dat deze voorzorgsmaatregel ongeveer één op de tien maal, dat zij overbodig is, toch wordt genomen. Vooral indien het nalaten van deze voorzorgsmaatregel, indien deze wel nodig is, ernstige gevolgen zou hebben.

Samenvatting: Het verwerpen van een juiste hypothese wordt een fout van de eerste soort genoemd. De kritieke zone van een toets wordt zo gekozen, dat de kans op het maken van deze fout, indien de getoetste hypothese juist is, hoogstens gelijk aan een gegeven getal α (de onbetrouwbaarheidsdrempel) is. De werkelijke waarde van deze kans, die bij discrete toetsingsgrootheden meestal niet precies gelijk aan α kan worden gemaakt, wordt met α_0 aangegeven en de onbetrouwbaarheid genoemd.

5. Bepaling van de kritieke zones bij de tekentoets.

Deze bepaling komt neer op de berekening van de grootste a resp. b in (9) en (10), waarvoor nog aan (11) voldaan is. Voor kleine n kan men dit gemakkelijk doen met behulp van de in hoofdstuk II genoemde tabellen der binomiale verdeling. Voor $p=\frac{1}{2}$ is er bovendien een tabel van de door (8) gegeven wh-verdeling tot $n=200$ toe: A. VAN WIJNGAARDEN, Table of the cumulative symmetric binomial distribution, Proc.Kon.Nederl.Akad.v.Wet. 53 (1950), pp. 857-868, Indag.Math. 12 (1950), pp. 301-312.

Voor $n \leq 100$ is een tabel van de waarden $\frac{1}{2}n-a$ berekend door W.J.DIXON and A.M.MOOD, The statistical sign test, Jrn.Am.Stat. Ass. 41 (1946), pp. 557-566, die als tabel I (bijlage) in dit hoofdstuk overgenomen is. De waarde $\frac{1}{2}n-a$ is daarin de kritieke waarde genoemd. Dezelfde waarden zijn kritieke waarden voor de linker- éénzijdige tekentoets, met de halve onbetrouwbaarheidsdrempel. Dit volgt direct uit (9) en (10). De kritieke waarden voor de rechter- éénzijdige tekentoets, met eveneens de halve onbetrouwbaarheidsdrempel, zijn gelijk aan:

$\frac{1}{2}n+a = n-(\frac{1}{2}n-a)$.

Indien er geen geschikte tabel bij de hand is en men wel de beschikking over een tabel van de normale verdeling heeft, kan men gebruik maken van de normale benadering van de door (8) gegeven verdeling. Daar deze verdeling $\frac{1}{2}n$ als gemiddelde en $\frac{1}{2}\sqrt{n}$ als spreiding heeft, is de aangepaste normale verdelingsdichtheid

(12) $$f(x) = \frac{2}{\sqrt{2\pi n}} e^{-2\frac{(x-\frac{1}{2}n)^2}{n}},$$

of ook, de grootheid

(13) $$\underline{y} \overset{\text{def}}{=} \frac{\underline{x}-\frac{1}{2}n}{\frac{1}{2}\sqrt{n}}$$

is bij benadering $N(0,1)$ verdeeld. Bepalen wij dus (uit een tabel van de normale verdeling) bij gegeven α de waarde $y_{\frac{1}{2}\alpha}$ van $\underline{y}$ zo, dat

(. $$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y_{\frac{1}{2}\alpha}} e^{-\frac{1}{2}u^2}du = \frac{1}{2}\alpha$$

is, dan kunnen wij uit (13) de bijbehorende waarde $x_{\frac{1}{2}\alpha}$ van $\underline{x}$ oplossen. Om de continuïteitscorrectie (zie II.3, blz.23) toe te passen, verminderen wij $x_{\frac{1}{2}\alpha}$ met $\frac{1}{2}$ en als wij nu het grootste gehele getal, dat hoogstens gelijk is aan $x_{\frac{1}{2}\alpha}-\frac{1}{2}$, bepalen, hebben wij een benadering van $\frac{1}{2}n-a$; dit wordt uitgedrukt door de volgende notatie:

(14) $$\tfrac{1}{2}n-a \approx [x_{\frac{1}{2}\alpha}-\tfrac{1}{2}].$$

Voeren wij deze berekening uit voor $n \leqq 100$ en $\alpha = 0{,}05$, dan blijkt, dat (14) voor al deze waarden van n, behalve voor $n=17$ en $n=94$, een gelijkheid is. Voor de 2 genoemde waarden bevat de benaderde kritieke zone in het tweezijdige geval aan beide zijden één waarde van $\underline{x}$ te weinig. De benadering is dus zeer goed.

Indien men er niet op uit is de kritieke zone zelf te berekenen, maar alleen wenst te weten, of een gevonden waarnemingsresultaat x_0 in de kritieke zone ligt, kan men met een eenvoudigere berekening volstaan. Immers, indien men het tweezijdige geval beschouwt (voor de éénzijdige toetsen gaat het geheel analoog), dan ligt x_0 in Z, als

$$|\tfrac{1}{2}n-x_0| \geqq a$$

is, dus als

(15) $$P[|\tfrac{1}{2}n-\underline{x}| \geqq |\tfrac{1}{2}n-x_0| \mid H_0] \leqq P[|\tfrac{1}{2}n-\underline{x}| \geqq a] \leqq \alpha$$

is. Het eerste lid van deze ongelijkheid wordt de <u>tweezijdige</u>

overschrijdingskans van x_0 genoemd.(de definitie van de éénzijdige overschrijdingskansen is analoog). Daar echter $\underline{y}$ (zie (15)) bij benadering $N(0,1)$ verdeeld is, is de tweezijdige overschrijdingskans van x_0, berekend met continuïteitscorrectie, ongeveer gelijk aan

$$(16) \qquad k = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{-y^*} e^{-\frac{1}{2}u^2} du ,$$

met

$$(17) \qquad y^* = \frac{|x_0 - \frac{1}{2}n| - \frac{1}{2}}{\frac{1}{2}\sqrt{n}}$$

en wij behoeven dus slechts y^* te berekenen en k op te zoeken en na te gaan of k kleiner of groter dan α is.

Voorbeeld 1. "Lady tasting tea".

In Engeland, waar men de thee met veel melk drinkt, treft men mensen aan, die beweren te kunnen proeven, of de melk of de thee het eerst in het kopje is gedaan. J.NEYMAN, First course in probability and statistics, N.Y. 1950, gebruikt dit voorbeeld onder de bovenstaande titel, om er verschillende aspecten van de toetsingstheorie aan te demonstreren. Er is een uitvoerige polemiek geweest tussen hem en R.A.FISHER, die in zijn boek "The design of experiments", Edinburgh, 1925, een methode aangaf, om deze bewering statistisch te toetsen. Wij zullen op deze polemiek niet ingaan en het voorbeeld slechts gebruiken, om te laten zien, hoe men zich in dit geval van de tekentoets kan bedienen. Als proefopzet kan men b.v. 14 kopjes nemen, en bij ieder kopje met een zuivere munt werpen, om uit te maken, of men er de thee of de melk het eerst in zal doen. Vervolgens laat men de thee proevende dame van deze 14 kopjes zeggen, wat er volgens haar het eerst ingedaan is. Indien haar bewering, dat zij dit kan proeven, onjuist is en zij er maar naar raadt, is in deze omstandigheden de kans, om juist te raden, precies gelijk aan $\frac{1}{2}$. Het aantal successen $\underline{x}$ (waarbij zij het goede antwoord geeft) wordt nu als toetsingsgrootheid gebruikt, om de hypothese H_0 te toetsen, inhoudende, dat zij het niet kan proeven, dus dat de kans op het geven van een juist antwoord steeds gelijk aan $\frac{1}{2}$ is. Voor de toetsing wordt in dit geval een rechter- éénzijdige kritieke zone gebruikt. De redenen hiervoor, en in het algemeen richtlijnen voor de keuze van een bepaalde kritieke zone uit alle mogelijke, worden verderop behandeld. In dit geval kan men echter wel direct, op intuïtieve gronden, zien, dat de rechter- éénzijdige kritieke zone de aangewezene is. Immers alleen als de dame vaker goed raadt, dan men op grond van H_0 zou verwachten, zal men geneigd zijn H_0 te verwerpen ten gunste van de hypothese, dat haar bewering juist is [1]). Dit verklaart de éénzijdigheid van de toets. Bovendien neemt natuurlijk de bewijskracht van de gevonden waarde x_0 van $\underline{x}$ toe, naarmate hij groter is. Vandaar, dat men werkt met

(6) $\quad Z_\tau: \ x - \frac{1}{2}n \geqq b.$

Nemen wij nu $\alpha = 0{,}05$, dan zien wij uit tabel I, dat $\frac{1}{2}n+b$ voor $n=14$ de waarde $14-3=11$ aanneemt (wegens de éénzijdigheid moeten wij de kolom met onbetrouwbaarheid 0,10 gebruiken). Wij zullen dus de dame gelijk geven, als zij minstens 11 van de 14 maal het goede antwoord geeft. De kans, dat zij dit bereikt, indien zij in werkelijkheid niet kan proeven, of de melk of de thee het eerst in het kopje is gedaan, is dan $\leqq 0{,}05$ (en wel, om precies te zijn, gelijk aan $0{,}029$; voor $\frac{1}{2}n+b=10$ wordt dit $0{,}09$).

Wij maken bij dit voorbeeld terloops enkele opmerkingen over de gebruikte proefopzet. De kwesties, waar het hier om gaat komen later uitvoeriger ter sprake.

De proefopzet berust op het gebruik van een "zuivere munt" voor het bepalen of de melk of de thee het eerst in het kopje zal worden gedaan. Dit betekent, dat men over een lotingsmechanisme moet beschikken met $p=\frac{1}{2}$, of althans met een zo dicht bij $\frac{1}{2}$ gelegen waarde, dat het verschil verwaarloosd kan worden. Hoe kan men nagaan, of een bepaald lotingsmechanisme daaraan voldoet? Wel, door de tekentoets toe te passen op één of meer reeksen van uitkomsten, die met dit mechanisme verkregen zijn. In het bijzonder zou men in dit geval een groot aantal reeksen van 14 lotingen kunnen beschouwen, om na te gaan, of de tekentoets in dat geval niet te vaak tot verwerping leidt.

Maar waarvoor is het gebruik van dit lotingsmechanisme eigenlijk nodig? Kunnen wij niet eenvoudig in 7 kopjes het eerst de melk en in 7 andere het eerst de thee doen en de kopjes dan, door elkaar gezet, alle tegelijk aan de dame voorzetten? Wij zouden dit kunnen doen, maar dan behoeft de kans op een juist antwoord

---

[1]) Tenminste indien wij geen rekening houden met de mogelijkheid, dat zij wel een verschil proeft, maar juist de melk als eerst ingeschonken aanwijst, als het de thee is en andersom. Indien wij die mogelijkheid ook wensen te beschouwen, zouden wij de toets tweezijdig moeten uitvoeren. Wij laten deze complicatie voor de eenvoud buiten beschouwing.

ook onder H_0, dus als zij het niet kan proeven, niet meer gelijk aan $\frac{1}{2}$ te zijn. Indien de dame, om welke reden dan ook, geneigd is, om vaker te zeggen, dat de thee er eerder in gedaan is dan de melk, dan is de kans op het geven van een juist antwoord onder H_0 bij de kopjes, waar de thee het eerst in gedaan is groter dan $\frac{1}{2}$ en bij de andere kleiner. De tekentoets kan dan dus niet meer worden toegepast. Deze moeilijkheid bestaat niet bij het gebruik van het lotingsmechanisme met $p=\frac{1}{2}$.

Voorbeeld 2. Jongens- en meisjesgeboorten.

In 1930 werden in Amsterdam 6848 jongens en 6374 meisjes geboren. Noemen wij het aantal jongensgeboorten onder 6848+6374= =13.222 geboorten $\underline{x}$, dan bezit $\underline{x}$, onder de hypothese H_0, dat de kans op een jongen gelijk is aan die op een meisje, een binomiale verdeling met gemiddelde $\frac{13222}{2} = 6611$ en spreiding $\frac{1}{2}\sqrt{13222} \approx 57{,}5$. Derhalve is (zie (17)):

$$y^* = \frac{6848 - 6611 - \frac{1}{2}}{57{,}5} = \frac{236{,}5}{57{,}5} \approx 4.$$

Derhalve vinden wij voor k (zie (16)) bij tweezijdige toetsing de waarde 0,00006. Dit is zeer veel kleiner dan alle gebruikelijke onbetrouwbaarheidsdrempels, zodat wij met slechts een zeer gering voorbehoud tot verwerping van de getoetste hypothese over kunnen gaan en besluiten, dat de kans op een jongensgeboorte groter is dan die op een meisjesgeboorte.

Deze twee voorbeelden zijn van een nogal speciaal karakter. Wij zullen verderop zien, hoe een grote groep van technische en wetenschappelijke vragen met behulp van de tekentoets kan worden onderzocht.

6. Fouten van de tweede soort; onderscheidingsvermogen.

Behalve de mogelijkheid, dat een juiste hypothese H_0 verworpen wordt, moeten wij ook rekening houden met de mogelijkheid van de complementaire fout, dat een onjuiste H_0 niet verworpen wordt. Dit wordt een fout van de tweede soort genoemd. In de regel beheerst men de kans op deze fout minder goed dan die op een fout van de eerste soort. Bij de keuze van de toetsingsmethode en vooral van de te toetsen hypothese kan men daar vaak rekening mee houden, door H_0 zo te kiezen, dat een fout van de tweede soort minder ernstig is dan één van de eerste soort. J.NEYMAN gaat zelfs zover de ernstigste van de twee mogelijke foutensoorten altijd de fout van de eerste soort te noemen en de te toetsen hypothese naar aanleiding daarvan op te stellen. Daar echter in vele gevallen niet voor elke gewenste hypothese een toets ontwikkeld is, kan men daar in de practijk niet streng de hand aan houden. Wij zullen daarom die terminologie niet overnemen.

Evenals het maken van een fout van de eerste soort onmogelijk is, indien H_o onjuist is, is het maken van een fout van de tweede soort onmogelijk, indien H_o juist is. Indien H_o echter onjuist is, hangt de kans op het maken van een fout van de tweede soort ervan af, welke hypothese nu wel juist is. Geven wij deze juiste hypothese aan met H, dan wordt

$$(18) \qquad \beta(H) \overset{\text{def}}{=} P[\underline{x} \in Z \mid H] \qquad (H \in \Omega) \qquad \text{(zie § 1)}$$

het <u>onderscheidingsvermogen</u> van de toets, of van de kritieke zone Z genoemd. Dit onderscheidingsvermogen is dus de kans op het verwerpen van H_o, indien H juist is. Voor $H=H_o$ is β dus de onbetrouwbaarheid, en voor $H \neq H_o$ is $1-\beta$ de kans op een fout van de tweede soort. Indien gewenst wordt het symbool, dat de kritieke zone aangeeft als index aan de β gehangen: $\beta_Z(H)$, ter onderscheiding van verschillende toetsen.

De eigenschappen van een toets worden met grote mate van volledigheid beschreven door het onderscheidingsvermogen. Wij zullen later aangeven, hoe men, in het geval van tweezijdige toetsing, een nog iets verder gaande beschrijving van een toets kan geven.

7. <u>Het onderscheidingsvermogen van de tekentoets</u>.

Bij de tekentoets wordt Ω gegeven door (2) en een hypothese H komt dus overeen met een waarde van p tussen 0 en 1. Het onderscheidingsvermogen is dus een functie van p. De wh-verdeling van de toetsingsgrootheid $\underline{x}$ wordt dan gegeven door (1) en het bij de door (5), (6) en (7) gegeven kritieke zones Z, Z_r en Z_l behorende onderscheidingsvermogen is dus (vgl. ook (9) en (10)):

$$(19) \qquad \begin{cases} \beta_Z(p) = \sum\limits_{i=0}^{\frac{1}{2}n-a} \binom{n}{i} p^i q^{n-i} + \sum\limits_{i=\frac{1}{2}n+a}^{n} \binom{n}{i} p^i q^{n-i} \\[2ex] \beta_{Z_r}(p) = \sum\limits_{i=\frac{1}{2}n+b}^{n} \binom{n}{i} p^i q^{n-i} \\[2ex] \beta_{Z_l}(p) = \sum\limits_{i=1}^{\frac{1}{2}n-b} \binom{n}{i} p^i q^{n-i}. \end{cases}$$

Deze drie functies kunnen dus, met behulp van de vroeger genoemde tabellen en benaderingen zonder veel moeite worden bepaald. Nemen wij b.v., evenals in voorbeeld 1 van § 5, $n=14$ en $\alpha=0{,}05$, dan vinden wij uit de tabel voor de binomiale verdeling van het National Bureau of Standards als kritieke zones:

Z : $x \leqq 2$ of $\geqq 12$ $\beta_Z(\frac{1}{2}) = 0{,}013$ ($\alpha = 0{,}05$).

Z_r: $x \geqq 11$ $\beta_{Z_r}(\frac{1}{2}) = 0{,}029$ ($\alpha = 0{,}05$).

Z_l: $x \leqq 3$ $\beta_{Z_l}(\frac{1}{2}) = 0{,}029$ ($\alpha = 0{,}05$).

Met behulp van (19) vinden wij verder uit deze zelfde tabel de in tabel II vermelde waarden.

Tabel II

Onderscheidingsvermogen van de tekentoets bij $n=14$ en $\alpha=0{,}05$.

p	$\beta_Z(p)$	$\beta_{Z_r}(p)$	$\beta_{Z_l}(p)$
0,0	1	0	1
0,1	0,842	0,000	0,956
0,2	0,448	0,000	0,698
0,3	0,161	0,000	0,355
0,4	0,039	0,004	0,124
0,5	0,013	0,029	0,029
0,6	0,039	0,124	0,004
0,7	0,161	0,355	0,000
0,8	0,448	0,698	0,000
0,9	0,842	0,956	0,000
1	1	1	0

De drie functies zijn in figuur 2 geschetst.

De vorm van ieder der krommen kan men op grond van eenvoudige overwegingen gemakkelijk begrijpen. Beschouw b.v. $\beta_{Z_r}(p)$. Z_r bevat alleen grote waarden van x en voor $p=0$ zal er zeker geen enkel succes optreden, dus is β dan gelijk aan 0. Met toenemende p neemt uiteraard de kans op een groot aantal successen toe, dus ook de kans, om een resultaat in de kritieke zone te vinden, dus β neemt toe. Voor $p=1$ komen wij zeker in de kritieke zone terecht, dus is β dan gelijk aan 1. Soortgelijke overwegingen gelden voor de twee andere krommen. $\beta_{Z_l}(p)$ is de om de verticale lijn $p=\frac{1}{2}$ gespiegelde kromme van $\beta_{Z_r}(p)$.

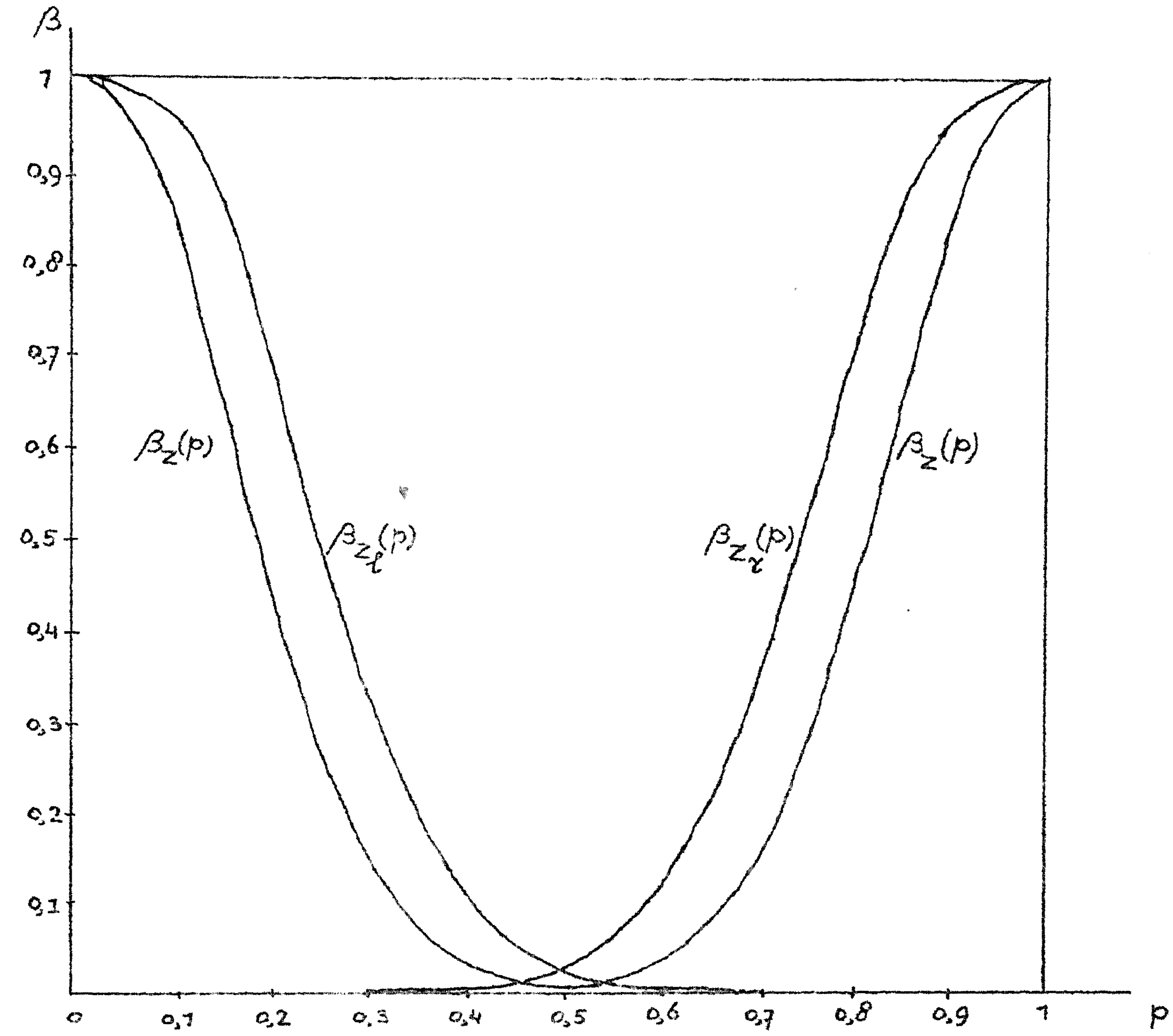


Fig. 2: Het onderscheidingsvermogen van de tekentoets voor $n=14$; $\alpha=0{,}05$.

Opmerking: Nemen wij voor de éénzijdige toetsen $\alpha'=\frac{1}{2}\alpha$, dus in ons geval een onbetrouwbaarheidsdrempel 0,025, dan is $Z_r \cup Z_\ell = Z$ en dus $\beta_Z(p) = \beta_{Z_r}(p) + \beta_{Z_\ell}(p)$.

8. Eén- en tweezijdige toetsing.

Het kiezen van een toets voor een te toetsen hypothese H_0 komt neer op het kiezen van een kritieke zone. Bij het door ons beschouwde geval $H_0: p=\frac{1}{2}$, hebben wij tot nu toe drie mogelijkheden beschouwd; Z, Z_r en Z_ℓ. Er zijn natuurlijk veel meer mogelijkheden. Men zou b.v. ook kritieke zones van de vorm

$$c \leqq x \leqq d$$

kunnen nemen. Dit is, intuïtief beschouwd, weinig voor de hand liggend en wij zullen later ook aantonen, dat het weinig doeltreffend zou zijn. Voorlopig bepalen wij ons echter tot het geven van richtlijnen voor een keuze uit de 3 kritieke zones, waarvan in fig. 2 (voor een speciaal geval: $n=14$; $\alpha=0{,}05$) het onderscheidingsvermogen getekend is.

Deze keuze geschiedt naar aanleiding van het onderscheidingsvermogen en als richtsnoer kan men de volgende nemen. Indien men er vooral of uitsluitend op gesteld is de hypothese H_0 te verwerpen, indien $p>\frac{1}{2}$ is, maar niet, indien $p<\frac{1}{2}$ is, gebruikt men Z_r ; is de situatie juist andersom, dan Z_ℓ. Wenst men geen onderscheid te maken tussen $p>\frac{1}{2}$ en $p<\frac{1}{2}$, dan Z . Bij de jongens- en meisjesgeboorten b.v. en in het algemeen zeer vaak bij het toetsen van wetenschappelijke hypothesen, zal dit laatste het geval zijn, zodat men dan tweezijdig toetst. Bij de "lady tasting tea" kunnen zich beide mogelijkheden voordoen. Indien men alleen wenst na te gaan, of $p=\frac{1}{2}$ is, kan men tweezijdig toetsen. Ook als zij veel te vaak juist het verkeerde antwoord geeft, wijst dit er nl. op, dat zij een verschil proeft, maar dit juist verkeerd verklaart. Indien zij echter niet alleen beweert een verschil te kunnen proeven, maar bovendien, dat zij dan ook de juiste oorzaak weet aan te geven, dan zal men H_o (inhoudende, dat haar bewering onjuist is) slechts willen verwerpen ten gunste van $p>\frac{1}{2}$ en dus Z_r als kritieke zone nemen. Z_ℓ komt bij dit voorbeeld niet in aanmerking.

De keuze van een kritieke zone berust, algemeen geformuleerd, op de keuze van een te toetsen hypothese en een speciale klasse A van alternatieve hypothesen; indien een hypothese van A juist is, zou men H_o graag willen verwerpen. Men wenst dus het onderscheidingsvermogen van de toets speciaal voor de tot A behorende hypothesen zo groot mogelijk te maken. Het is duidelijk, dat dit bij de keuze uit de 3 zones Z, Z_r en Z_ℓ bij de bovenstaande voorbeelden is gedaan. Men kan nog andere eisen aan kritieke zones stellen. In dit verband voeren wij in de volgende paragraaf enkele begrippen hieromtrent in.

9. Begrippen, die verband houden met de keuze van een kritieke zone.

Een toets, en ook de bijbehorende kritieke zone Z, wordt zuiver (Engels: unbiased) genoemd als toets van H_o tegen de klasse van alternatieve hypothesen A, indien voor iedere hypothese H van A geldt:

(20) $$\beta_Z(H) \geqq \beta_Z(H_o), \quad (H \in A)$$

waarin $\beta_Z(H_o)$ de onbetrouwbaarheid van de toets is (die dus $\leqq \alpha$, de onbetrouwbaarheidsdrempel, is). Dit betekent dus, dat bij een zuivere toets van H_o tegen de klasse A, de kans op verwerping van H_o, indien H_o juist is, hoogstens even groot is, als indien H_o onjuist is en één der hypothesen van A juist is. Dit is uiteraard een eigenschap, die men van een goede toets voor H_o tegen A mag vergen; zelfs zal men gewoonlijk wensen, dat (20) met het $<$-teken geldt.

Voorbeelden.

Uit figuur 2 zien wij, dat Z een zuivere kritieke zone is voor toetsing van H_o tegen de klasse van alternatieven

(21) $$A_Z : \; p \neq \tfrac{1}{2} \quad (0 \leqq p \leqq 1).$$

Dit geldt echter niet voor Z_ℓ en Z_r, daar de door (21) ...

met Z_ℓ resp. Z_r in plaats van Z, niet geldt.

Z_ℓ resp. Z_r zijn echter zuiver als toets van H_o tegen

(22) $A_{Z_\ell}: 0 \leq p < \frac{1}{2}$

resp.

(23) $A_{Z_r}: \frac{1}{2} < p \leq 1$.

De in de vorige paragraaf besproken keuze tussen Z, Z_ℓ en Z_r kunnen wij dus ook beschrijven als de keuze van een zuivere toets voor H_o tegen één der drie klassen A_Z, A_{Z_ℓ} resp. A_{Z_r} van alternatieve hypothesen. Hiermee wordt in andere woorden bijna hetzelfde gezegd als in de vorige paragraaf.

Daar komt echter nog iets bij. Immers Z is óók zuiver als toets van H_o tegen Z_ℓ en tegen Z_r. Indien wij dus H_o wensen te toetsen tegen de door (23) gegeven klasse van alternatieve hypothesen, dan voldoen zowel Z als Z_r aan de eis van zuiverheid. Toch gebruiken wij in dat geval bij voorkeur Z_r en niet Z, omdat voor iedere hypothese van deze klasse Z_r een groter onderscheidingsvermogen heeft dan Z. De kans op een fout van de tweede soort is dan dus bij Z_r kleiner dan bij Z, terwijl de kans op een fout van de eerste soort bij beide $< \alpha$ is.

Wij noemen de kritieke zone Z_1 <u>meer onderscheidend</u> dan Z_2 met betrekking tot een hypothese H, indien

(24) $\beta_{Z_1}(H) > \beta_{Z_2}(H)$

is, terwijl voor de getoetste hypothese H_o geldt:

(25) $\beta_{Z_1}(H_o) \leq \alpha$ en $\beta_{Z_2}(H_o) \leq \alpha$.

Indien het er speciaal om gaat H_o te toetsen tegen H, wordt, indien aan (24) en (25) voldaan is, aan Z_1 de voorkeur gegeven boven Z_2.

Z_1 heet <u>uniform meer onderscheidend</u> dan Z_2 met betrekking tot de klasse A van hypothesen, indien voor iedere H van A aan (24) voldaan is, terwijl ook (25) geldt. Zo is b.v. bij bovenstaand voorbeeld (zie fig. 2) Z_r uniform meer onderscheidend dan Z met betrekking tot de klasse A_{Z_r}, die door (23) wordt gegeven.

Een kritieke zone Z heet de <u>meest onderscheidende</u> kritieke zone met betrekking tot een hypothese H, indien Z_1 aan (25) voldoet en indien voor <u>iedere</u> Z_2, die ook aan (25) voldoet, ook (24) geldt. Indien dit voor een gehele klasse A van hypothesen geldt, heet Z_1 <u>uniform meest onderscheidend</u> voor de betreffende klasse.

Kritieke waarden van de tekentoets (tweezijdig)

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

Opmerking. In § 4 is opgemerkt, dat wij de onbetrouwbaarheid $\beta(H_0)$ van een toets vaak niet precies gelijk aan α kunnen maken. In dat geval maken wij $\beta(H_0)$ zo groot mogelijk (maar niet groter dan α). De reden hiervan is, dat daardoor het onderscheidingsvermogen zo groot mogelijk wordt. Dit is het gemakkelijkst aan een voorbeeld te zien. Stel, dat wij een rechter-eenzijdige kritieke zone willen gebruiken, met $\alpha = 0{,}05$ en $n = 14$, dan is de grootst mogelijke Z_r: $x \geqq 11$. Wij zouden ook $x \geqq 12$ kunnen nemen, die één punt minder bevat. Daardoor neemt de kans, om in deze kritieke zone te komen af, zowel indien H_0 juist is, als indien H_0 onjuist is. Dit betekent dus, dat zowel de onbetrouwbaarheid van de toets als zijn onderscheidingsvermogen afnemen. Indien nu α gegeven is, betekent dit, dat de onderzoeker deze waarde als kans op een fout van de eerste soort wil toelaten. Wenst men dus het onderscheidingsvermogen zo groot mogelijk te maken - en daarmee dus de kans op een fout van de tweede soort te minimaliseren - dan moet men dus de kritieke zone, en daarmee $\beta(H_0)$, zo groot mogelijk maken, voor zover de gekozen waarde van α dit toelaat.

9. Toetsen van samengestelde hypothesen.

In verband met de eis van zuiverheid van een toets, kunnen wij in het algemeen de parameterruimte Ω, bij gegeven α en kritieke zone, in twee klassen verdelen, die wij Ω_0 en Ω_1 noemen en waarvoor geldt:

(26) $\beta(H) \leqq \alpha$ voor alle $H \in \Omega_0$

en

(27) $\beta(H) > \alpha$ voor alle $H \in \Omega_1$.

Wij zeggen dan, dat de toets een zuivere toets is voor Ω_0 tegen Ω_1, met α als onbetrouwbaarheid. In figuur 3 is een enigszins vreemdsoortig geval geschetst, terwijl de bijbehorende klassen Ω_0 en Ω_1 in de figuur zijn aangegeven.

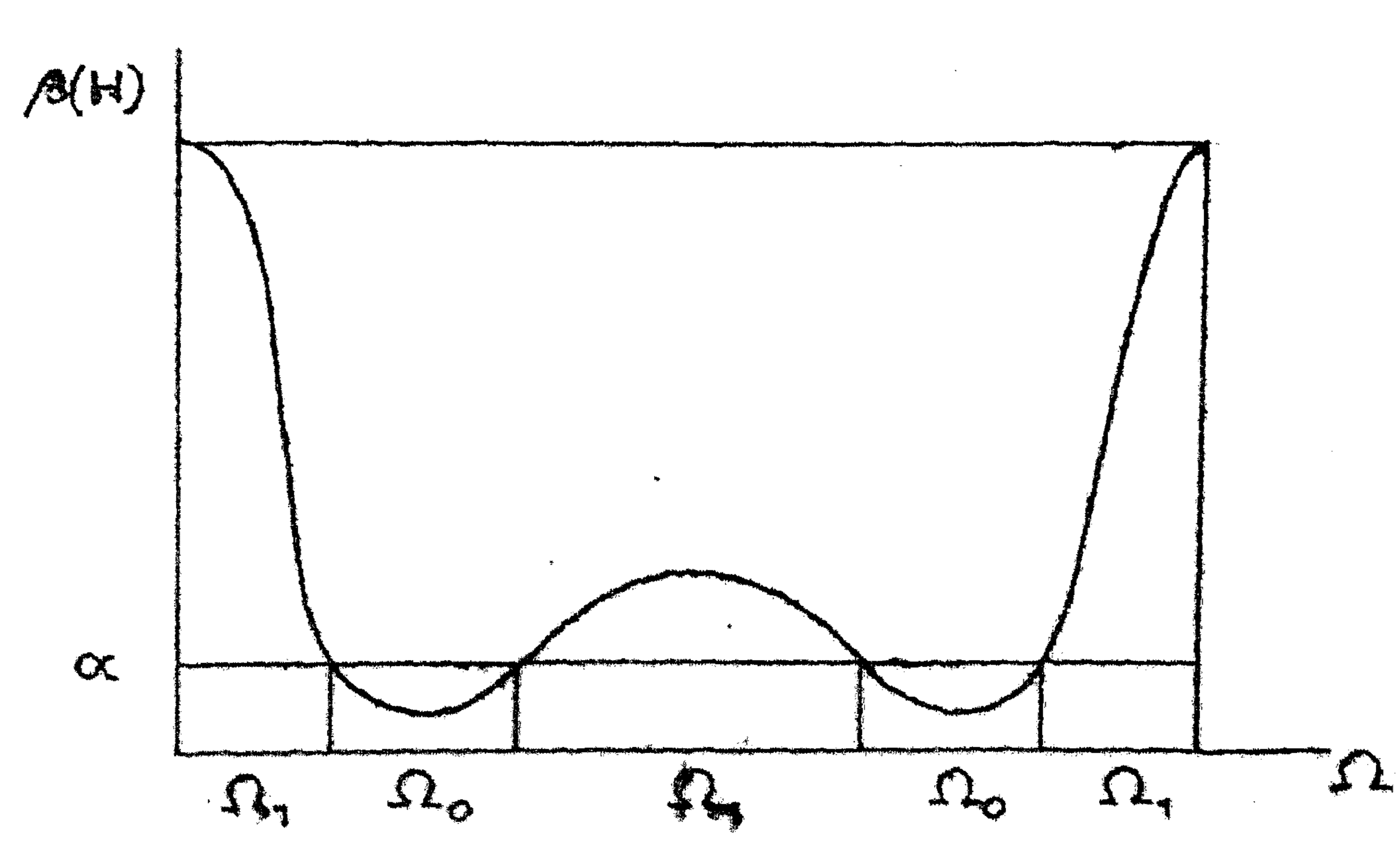


Fig. 3. Verdeling van Ω in twee

Beschouwen wij het speciale geval, dat in figuur 2 is geschetst, dan zien wij, dat Z, met $\alpha = 0{,}013$, een zuivere toets is voor $H_0: p=\frac{1}{2}$ tegen de klasse $\Omega_1: p \neq \frac{1}{2}$; Z_r is zuiver, met $\alpha = 0{,}029$, voor $\Omega_0: p \leqq \frac{1}{2}$ tegen $\Omega_1: p > \frac{1}{2}$ en Z_ℓ voor $\Omega_0: p \geqq \frac{1}{2}$ tegen $\Omega_1: p < \frac{1}{2}$. Houden wij vast aan $\alpha = 0{,}05$, dan zijn deze klassen enigszins anders: bij Z wordt Ω_0 een klein interval om $p=\frac{1}{2}$, terwijl Ω_0 zich bij Z_r iets verder naar rechts en bij Z_ℓ iets verder naar links dan de waarde $\frac{1}{2}$ uitstrekt.

Men kan zich afvragen, waarvoor deze nieuwe terminologie dient na de in de vorige paragraaf ingevoerde. De zin is daarin gelegen, dat de verdeling van Ω in twee klassen bij vele practische problemen een afspiegeling vormt van een dilemma, dat men met behulp van de toets wenst op te lossen. Indien b.v. een tabaksfabrikant een nieuwe methode voor het prepareren van zijn tabak wil vergelijken met de oude, kan hij een aantal rokers hun oordeel laten uitspreken over proefjes van volgens beide methoden bereide tabak. Noemen wij de kans, dat een roker de nieuwe tabak boven de oude zal prefereren p, dan hangt de beslissing van de fabrikant (al of niet invoering van de nieuwe methode) van de waarde van p af en men kan zich voorstellen, dat deze fabrikant van tevoren vaststelt: indien $p \leqq \frac{1}{2}$ is, laat ik alles bij het oude, maar is $p > \frac{1}{2}$ dan wil ik de nieuwe methode invoeren, omdat dit vermoedelijk de omzet zal vergroten. In dat geval ligt het dus voor de hand, om voor $\Omega_0: p \leqq \frac{1}{2}$ en voor $\Omega_1: p > \frac{1}{2}$ te nemen en te zoeken naar een zuivere toets voor Ω_0 tegen Ω_1.

Indien het invoeren van de nieuwe methode grote kosten vergt, zal men geneigd zijn de grens tussen Ω_0 en Ω_1 verder naar rechts te leggen, daar dit een grotere garantie geeft voor een vergroting van de omzet, indien p tot Ω_1 behoort. Ook de keuze van α zal van de kosten van de verandering in het procédé afhankelijk worden gesteld.

Zo zijn er vele situtaies denkbaar, waarbij de klassen Ω_0 en Ω_1 gekozen kunnen worden op grond van het doel van het onderzoek. Indien Ω_0 en Ω_1 op een dergelijke wijze gekozen zijn, zal men gaan zoeken naar een toets voor Ω_0 tegen Ω_1. In sommige gevallen zal men deze gemakkelijk, in andere moeilijk kunnen vinden. In ieder geval is een nieuwe definitie van de onbetrouwbaarheid nodig. Immers, voor de verschillende hypothesen van Ω_0 zal deze in de regel ook verschillend zijn. Bij deze toetsing van een samengestelde hypothese Ω_0 voeren wij daarom de volgende definitie in.

Indien men de samengestelde hypothese Ω_o toetst met behulp van een kritieke zone Z, dan is de onbetrouwbaarheid van Z met betrekking tot Ω_o gelijk aan

(28) $$\beta_Z(\Omega_o) \overset{def}{=} \max_{H\in\Omega_o} \beta_Z(H).$$

Z heet zuiver als toets van Ω_o tegen Ω_1, indien voor iedere $H\in\Omega_o$ en $H'\in\Omega_1$ geldt:

(29) $$\beta_Z(H) < \beta_Z(H').$$

11. Asymptotische eigenschappen van een toets.

De uitdrukking "asymptotische eigenschappen" slaat op eigenschappen, die in de limiet gelden voor $n\to\infty$, als n het aantal waarnemingen is, waarop de toets berust.

Wij noemen een toets (of ook de bijbehorende kritieke zone) asymptotisch onderscheidend (Engels: consistent) met betrekking tot de hypothese H, indien

(30) $$\lim_{n\to\infty} \beta(H) = 1$$

is. De kans op verwerping van de getoetste hypothese kan dan dus willekeurig dicht bij 1 gebracht worden door n voldoende groot te maken, indien H de juiste hypothese is.

Voor de tekentoets geldt de volgende stelling.

Stelling 1: De tweezijdige tekentoets is asymptotisch onderscheidend voor iedere hypothese met $p\neq\frac{1}{2}$; de linker- resp. rechter-éénzijdige tekentoets voor iedere hypothese met $p<\frac{1}{2}$ resp. $p>\frac{1}{2}$.

Bewijs.

Bij het bewijs maken wij gebruik van de ongelijkheid van Bienaymé-Tchebycheff (zie § I,13). Wij beschouwen eerst de tweezijdige toets, met onbetrouwbaarheidsdrempel α.

Onder $H_o(p=\frac{1}{2})$ wordt de wh-verdeling van het aantal successen $\underline{x}$ bij n waarnemingen gegeven door

(8) $$P[\underline{x}=x|H_o] = 2^{-n}\binom{n}{x}$$

met

$$\mathcal{E}\{\underline{x}|H_o\}=\tfrac{1}{2}n \quad \text{en} \quad \sigma\{\underline{x}|H_o\}=\tfrac{1}{2}\sqrt{n}.$$

Volgens de ongelijkheid van Bienaymé-Tchebycheff is dus

$$P[|\underline{x}-\tfrac{1}{2}n|>\tfrac{1}{2}k\sqrt{n}|H_o]<\frac{1}{k^2}.$$

Neem hierin $\frac{1}{k^2}=\alpha$, dus $k=\frac{1}{\sqrt{\alpha}}$, dan komt er

(31) $$P[|\underline{x}-\tfrac{1}{2}n|>\tfrac{1}{2}\sqrt{\tfrac{n}{\alpha}}|H_o]<\alpha.$$

De tweezijdige kritieke zone is

(5) $$Z:\ |x-\tfrac{1}{2}n| \geqq a,$$

waarin a zo gekozen moet worden, dat de onbetrouwbaarheid $\leqq\alpha$ is, dus dat

$$P[|\underline{x}-\tfrac{1}{2}n|\geqq a|H_0]\leqq\alpha$$

is. Indien wij dus de zone

(32) $$Z^*:\ |x-\tfrac{1}{2}n|>\tfrac{1}{2}\sqrt{\tfrac{n}{\alpha}}$$

beschouwen, zien wij, dat

(33) $$Z^*\subset Z \qquad (\subset = \text{"is bevat in"}).$$

geldt, en wel voor iedere n.

Derhalve geldt voor iedere n en p

$$\beta_{Z^*}(p)\leqq\beta_Z(p)$$

en indien wij dus voor een willekeurige waarde van p, die $\neq\frac{1}{2}$ is, bewijzen, dat

$$\lim_{n\to\infty}\beta_{Z^*}(p)=1$$

is, dan geldt dit a fortiori voor $\beta_Z(p)$, zodat de stelling dan voor het tweezijdige geval bewezen is.

Voor een kans $p(\neq\frac{1}{2})$ op een succes wordt de verdeling van het aantal successen bij n waarnemingen gegeven door

(1) $$P[\underline{x}=x|p]=\binom{n}{x}p^xq^{n-x} \quad (p+q=1)$$

met

$$\mathcal{E}\{\underline{x}|p\}=np \quad \text{en} \quad \sigma\{\underline{x}|p\}=\sqrt{npq}.$$

Beschouw het geval, dat $p>\frac{1}{2}$ is (voor $p<\frac{1}{2}$ verloopt het bewijs analoog). Dan geldt volgens de ongelijkheid van Bienaymé-Tchebycheff:

$$P[|\underline{x}-np|>k\sqrt{npq}|p]<\tfrac{1}{k^2}.$$

De ongelijkheid $|\underline{x}-np|>k\sqrt{npq}$ is vervuld, indien $\underline{x}-np<-k\sqrt{npq}$ is, dus geldt a fortiori

$$P[\underline{x}-np<-k\sqrt{npq}|p]<\tfrac{1}{k^2},$$

of ook

(34) $$P[\underline{x}-\tfrac{1}{2}n<np-\tfrac{1}{2}n-k\sqrt{npq}\,|p]<\tfrac{1}{k^2}.$$

Wij hebben nog de beschikking over de waarde van k, die wij naar believen kunnen kiezen, met de beperking $k>0$ (vgl. § I,13).

Wij kiezen nu k zo, dat het rechterlid van de ongelijkheid tussen de vierkante haken van (34) groter is dan het rechterlid van (32). Indien dan aan de ongelijkheid van (34) niet voldaan is, bevinden wij ons in Z^*, dus volgens (33) ook in Z. Daartoe lossen wij k op uit de vergelijking

$$np - \tfrac{1}{2}n - k\sqrt{npq} = \tfrac{1}{2}\sqrt{\tfrac{n}{\alpha}} + 1 .$$

Dit geeft

$$k = \frac{n(p-\frac{1}{2}) - \frac{1}{2}\sqrt{\frac{n}{\alpha}} - 1}{\sqrt{npq}}$$

en, voor voldoende grote n, is deze waarde positief, daar $p>\frac{1}{2}$.
Vullen wij deze waarde nu in in (34), dan komt er dus

$$P[\underline{x} - \tfrac{1}{2}n < \tfrac{1}{2}\sqrt{\tfrac{n}{\alpha}} + 1 \mid p] < \frac{npq}{\{n(p-\frac{1}{2}) - \frac{1}{2}\sqrt{\frac{n}{\alpha}} - 1\}^2}$$

Voor $n \to \infty$ heeft het rechterlid o als limiet, daar in de noemer n^2 voorkomt en in de teller slechts n. Het linkerlid heeft dus ook o tot limiet, dus gèldt

$$\lim_{n\to\infty} P[\underline{x} - \tfrac{1}{2}n > \tfrac{1}{2}\sqrt{\tfrac{n}{\alpha}} \mid p] = 1 ,$$

of ook

$$\lim_{n\to\infty} P[\underline{x} \in Z \mid p] = 1 ,$$

waarmee de stelling voor het tweezijdige geval bewezen is.

Voor het éénzijdige geval volgt de stelling nu direct uit het voorafgaande. Beschouw b.v. Z_r. Bij gegeven α bevat Z_r de rechterhelft van Z geheel en in het bovenstaande is bewezen, dat de kans op een x in deze rechterhelft, die dus ook in Z_r ligt, tot 1 nadert. Hetzelfde geldt mutatis mutandis voor Z_ℓ.

Men kan het bewijs aanschouwelijk voorstellen. Dit is in fig. 4 gedaan.

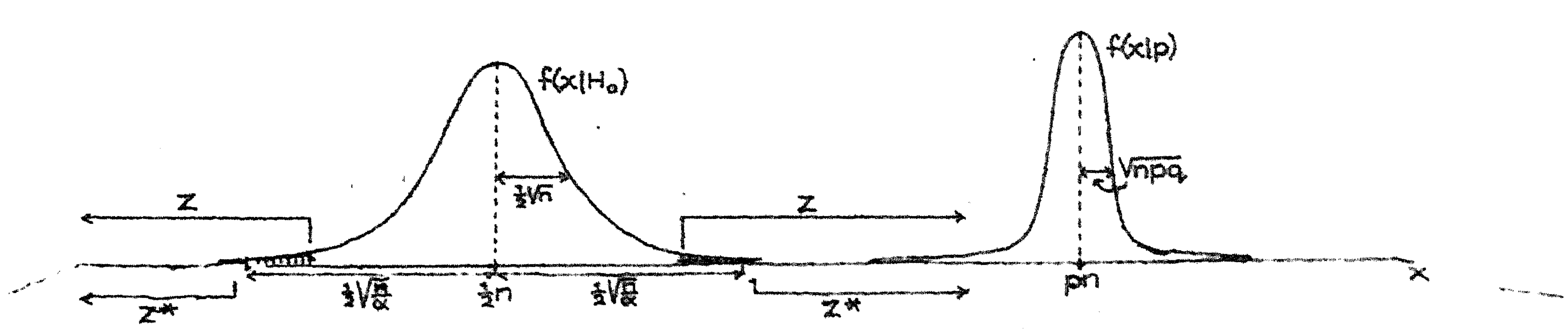


Fig. 4. Wh-verdelingen van x onder H_0 en H $(p>\frac{1}{2})$ voor grote n.

Voor toenemende n schuiven de punten $\frac{1}{2}n$ en pn steeds verder uit elkaar, terwijl de spreidingen $\frac{1}{2}\sqrt{n}$ en $\sqrt{npq}$ in absolute zin toenemen, maar in verhouding tot het verschil $pn - \frac{1}{2}n$ onbegrensd afnemen. De kans, onder hypothese H, om in Z te komen nadert derhalve voor $n \to \infty$ tot 1.

Uit de stelling blijkt, dat het onderscheidingsvermogen, dat in fig. 2 voor $n=14$ gegeven is, voor grote n de in fig. 5 geschetste vorm aanneemt, terwijl voor toenemende n de drie krommen steeds spitser, resp. steiler, worden.

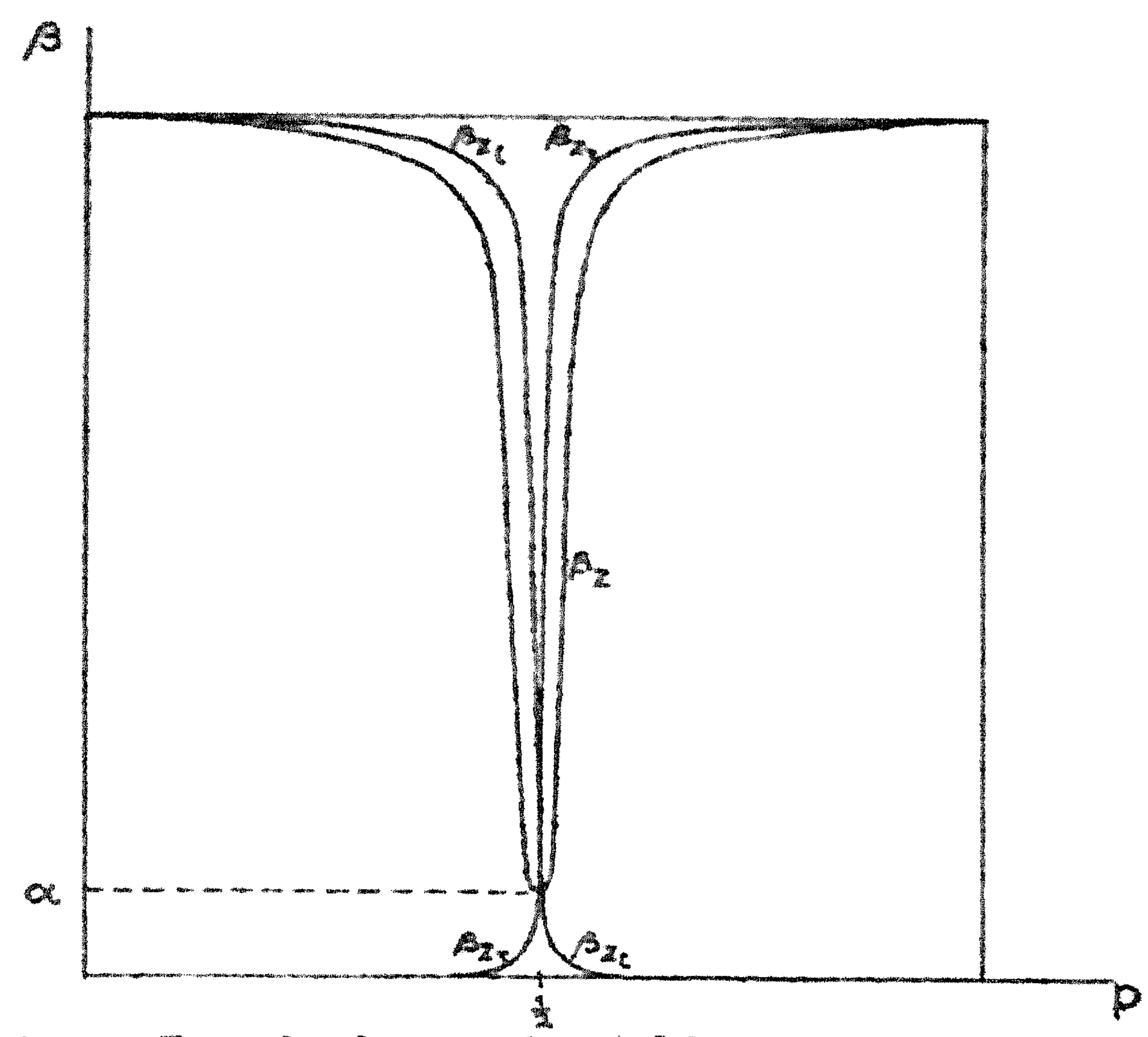


Fig. 5. Onderscheidingsvermogen van de tekentoets voor grote n.

Ook naar aanleiding van de asymptotische eigenschappen van een toets kunnen wij deelklassen van de klasse Ω van toegelaten hypothesen beschouwen.

Wij zeggen, dat een toets <u>asymptotisch volledig onderscheidt</u> tussen de klassen Ω_1 en Ω_2, indien

(35) $\lim_{n\to\infty} \beta(H) = 0$ voor $H \in \Omega_1$

en

(36) $\lim_{n\to\infty} \beta(H) = 1$ voor $H \in \Omega_2$.

De rechteréénzijdige tekentoets is dus asymptotisch volledig onderscheidend voor de klassen $\Omega_1 : p<\frac{1}{2}$ en $\Omega_2 : p>\frac{1}{2}$, terwijl deze klassen voor de linkeréénzijdige toets verwisseld worden. Voor de tweezijdige toets is $\Omega_2 : p\neq\frac{1}{2}$, terwijl Ω_1 ontbreekt. In alle drie gevallen is er een derde klasse, Ω_o, bestaande uit de hypothese $H_o(p=\frac{1}{2})$, waarvoor

(37) $\lim_{n\to\infty} \beta(H_o) = \alpha$.

is. Wij noemen de toets <u>asymptotisch onderscheidend</u> voor Ω_o tegen Ω_2, daarbij de toevoeging "volledig" weglatend, omdat het rechterlid van (37) >o is. De klasse Ω_o behoeft in principe niet uit één enkelvoudige hypothese te bestaan al is dit bij de tekentoets wel het geval.

12. Benadering van het onderscheidingsvermogen met behulp van de normale verdeling.

In de vorige paragraaf zijn enkele asymptotische eigenschappen van de tekentoets bewezen op grond van het gemiddelde en de spreiding van de binomiale verdeling alleen. In feite kwam de berekening neer op het aangeven van een ondergrens voor het onderscheidingsvermogen, die bij vaste $p \neq \frac{1}{2}$ voor $n \to \infty$ tot 1 nadert. Maken wij bovendien gebruik van de in § II.2 vermelde asymptotische normaliteit van de binomiale verdeling, dan kunnen wij meer doen, dan alleen maar een ondergrens aangeven, wij kunnen dan nl. het onderscheidingsvermogen bij benadering berekenen met behulp van een tabel van de normale verdeling.

Wij beschouwen daartoe eerst de door (6) (zie § 3) gegeven kritieke zone Z_r:

(6) $$Z_r: \ x - \tfrac{1}{2}n \geq b.$$

Het onderscheidingsvermogen is de kans, dat het aantal successen bij n proeven in Z_r zal vallen, als functie van de kans p op succes. Het wordt gegeven door (19) (zie § 7):

(19) $$\beta_r(p) = \sum_{i=\frac{1}{2}n+b}^{n} \binom{n}{i} p^i q^{n-i} \qquad (q = 1-p).$$

Volgens de reeds eerder gebruikte procedure gebruiken wij de gestandaardiseerde grootheid, die wij nu, omdat verschillende waarden van p beschouwd worden, met $\underline{y}_p$ aangeven:

(38) $$\underline{y}_p = \frac{\underline{x} - pn}{\sqrt{npq}}.$$

Deze grootheid is voor grote n bij benadering $N(0,1)$ verdeeld. De ongelijkheid (6) kunnen wij nu ook schrijven als

(39) $$Z_r: \ \underline{y}_p \geq \frac{b + (\frac{1}{2} - p)n}{\sqrt{npq}}$$

en de kans hierop kan (bij benadering) gevonden worden in een tabel van de $N(0,1)$-verdeling. Willen wij daarbij nog een continuïteitscorrectie toepassen, dan verminderen wij b met $\frac{1}{2}$.

Voor de linker-éénzijdige toets met Z_l als kritieke zone is het onderscheidingsvermogen, zoals in § 7 is opgemerkt, het gespiegelde van dat van Z_r. Voor de tweezijdige toetsing, met als kritieke zone (zie (5), § 3):

(5) $$Z: \ |x - \tfrac{1}{2}n| \geq a,$$

verloopt de berekening analoog, met a in plaats van b, terwijl dan achteraf nog met 2 vermenigvuldigd moet worden.

Voorbeeld. De in §10 op blz. 43 genoemde tabaksfabrikant, die overweegt een nieuwe methode in te voeren voor het prepareren van zijn tabak, acht het lonend hiertoe over te gaan, indien minstens 60% van de rokers van het door hem bestreken rayon de volgens de nieuwe methode geprepareerde tabak beter vindt dan de oude. Hij is er natuurlijk niet op gesteld de nieuwe methode in te voeren, indien deze slechter is dan de oude en stelt dus als eerste eis, dat de kans op deze foute beslissing (die in dit geval een fout van de eerste soort is) niet groter dan 0,05 mag zijn. Anderzijds is hij er echter zeer op gesteld, de nieuwe methode wèl in te voeren, als 60% of meer van zijn rokers de nieuwe tabak beter vinden dan de oude. Hij wil daarom het experiment zo opzetten, dat hij, als dit inderdaad zo is, minstens een kans 0,9 heeft, het ook te ontdekken. Hij kan dit bereiken door het aantal proefpersonen, die hij de beide soorten tabak laat vergelijken, voldoende groot te kiezen. Hij wenst dit aantal echter niet groter te nemen dan nodig is, om aan de beide bovenstaande eisen te voldoen, daar dit overbodige moeite en kosten met zich mee zou brengen. De vraag is nu dus, hoeveel proefpersonen hij moet kiezen.

Geven wij de kans, dat een willekeurige roker de nieuwe tabak beter vindt dan de oude, weer aan met p, dan kunnen wij de eisen van de tabaksfabrikant als volgt mathematisch preciseren.

1e: p kan in principe iedere waarde tussen 0 en 1 aannemen, dus de verzameling van toegelaten hypothesen is (vergelijk §1)

(2) $$\Omega:\ 0 \leqq p \leqq 1.$$

2e: De uitslag van de toets moet bij voorkeur negatief zijn, indien $p \leqq \frac{1}{2}$ is. Wij nemen daarom, als te toetsen hypothese de samengestelde hypothese

(40) $$\Omega_0:\ 0 \leqq p \leqq \tfrac{1}{2}.$$

Indien aan (40) voldaan is wordt slechts een kans van hoogstens 0,05 toegestaan, om toch in de kritieke zone terecht te komen. Immers indien dit geschiedt wordt Ω_0 verworpen en wordt de nieuwe methode ingevoerd. Wij kunnen dit aangeven met de vergelijking

(41) $$\beta(\Omega_0) \overset{\text{def}}{=} \max_{p \in \Omega_0} \beta(p) \leqq 0{,}05.$$

3e: Indien $p \geqq 0{,}6$ is, moet de kans, om in de kritieke zone te komen, minstens gelijk aan 0,9 zijn. Noemen wij dus

(42) $$\Omega^*:\ 0{,}6 \leqq p \leqq 1,$$

dan wordt als tweede eis gesteld:

(43) $$\beta^*(\Omega^*) \overset{\text{def}}{=} \underset{p\in\Omega^*}{\text{Min}} \beta(p) \geqq 0{,}9.$$

Wij kunnen deze eisen ook grafisch voorstellen. Dit is in fig. 6 gedaan.

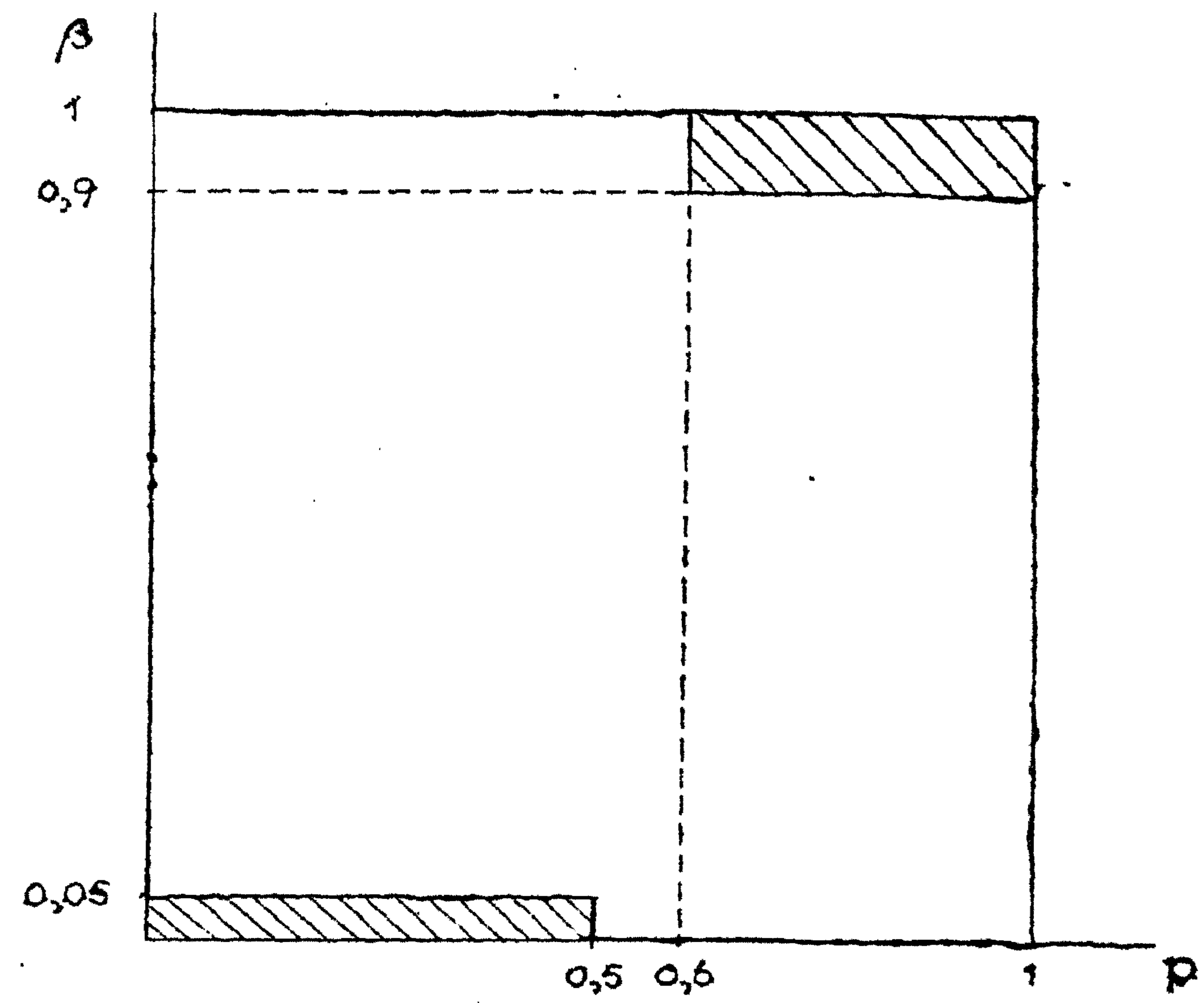


Fig. 6. Eisen van de tabaksfabrikant, grafisch voorgesteld.

De eerste eis, in formule voorgesteld door (41), houdt in, dat het onderscheidingsvermogen $\beta(p)$ voor $p \leqq 0{,}5$ in het links-onder gelegen gearceerde rechthoekje moet verlopen, de tweede eis, voorgesteld door (43), dat dit voor $p \geqq 0{,}6$ in het rechts-boven gelegen gearceerde rechthoekje moet geschieden. Wat er in het tussen 0,5 en 0,6 gelegen gebied gebeurt is niet zo erg belangrijk. De fabrikant heeft hierover geen eisen gesteld. Men noemt een dergelijk gebied vaak <u>indifferentiezone</u>. Beide mogelijke beslissingen worden, indien p in dit gebied ligt, acceptabel geacht.

Beschouwen wij fig. 2 op blz. 39, dan zien wij, dat alleen een rechteréénzijdige kritieke zone aan de beide eisen zal kunnen voldoen. Dit is dus een zone van het type (6).

Uit fig. 2 zien wij tevens, dat het maximum van $\beta(p)$ voor $p\in\Omega_0$ bereikt wordt voor $p=\frac{1}{2}$, zodat wij de door (41) aangegeven eis ook kunnen schrijven in de vorm

(41a) $$\beta_z(\tfrac{1}{2}) \leqq 0{,}05,$$

waarin de index z aangeeft, dat wij een rechteréénzijdige kritieke zone zullen gebruiken. Met behulp van de notatie (38) kunnen wij hier ook voor schrijven (in (39) wordt p=1 ingevuld):

$$\frac{b}{\frac{1}{2}\sqrt{n}} \geqq 1{,}645$$

of:

(41c) $$b \geqq 0{,}83\sqrt{n}.$$

Vervolgens gaan wij eis (43) op analoge wijze bewerken. Uit fig. 2 zien wij, dat het minimum van $\beta(p)$ voor $p \in \Omega^*$ bij een rech-~~ter~~éénzijdige kritieke zone bereikt wordt voor $p=0{,}6$ (het linker eindpunt). Met behulp van (38) kunnen wij dus (43) schrijven als

(43b) $$P\left[\underline{y}_{0,6} \geqq \frac{b-0{,}1n}{\sqrt{0{,}24n}}\right] \geqq 0{,}9$$

(nu is $p=0{,}6$ ingevuld). Uit de $N(0,1)$-tabel vinden wij nu

$$\frac{b-0{,}1n}{\sqrt{0{,}24n}} \leqq -1{,}28$$

of, daar $\sqrt{0{,}24} = 0{,}49$ is

(43c) $$b \leqq 0{,}1n - 0{,}63\sqrt{n}.$$

Wij moeten dus n zo groot nemen, dat zowel aan (41c) als aan (43c) voldaan kan worden door een geschikte keuze van b, dus van z_τ. Dit is mogelijk, indien n voldoet aan

$$0{,}83\sqrt{n} \leqq 0{,}1n - 0{,}63\sqrt{n},$$

dus aan

$$1{,}46\sqrt{n} \leqq 0{,}1n$$

of

$$14{,}6 \leqq \sqrt{n}$$

of wel

(44) $$n \geqq (14{,}6)^2 = 215{,}2.$$

De fabrikant zal dus minstens (ongeveer) 216 rokers om hun mening moeten vragen om met behulp van de tekentoets te voldoen aan de door hem gestelde eisen.

Opmerkingen.

1. Bij bovenstaande berekeningen is de continuïteitscorrectie niet aangebracht. De betrekkelijk grote uitkomst voor n rechtvaardigt deze verwaarlozing.

2. Men kan bij vraagstukken als het bovenstaande soms ook gebruik maken van de vroeger genoemde tabellen van de binomiale verdeling (nl. als de gezochte n niet te groot uitvalt). Een tabelletje van de kleinste n, nodig om een bepaald onderscheidingsvermogen te verkrijgen bij verschillende waarden van p en van de onbetrouwbaarheidsdrempel α vindt men o.a. in W.J.DIXON and F.J.MASSEY, Introduction to Statistical Analysis, N.Y. 1951, p. 250. Bij gebruik van deze tabel houde men in het oog, dat de onbetrouwbaarheidsdrempel α hier voor de tweezijdige toets is opgegeven.

13. Het λ-principe voor een enkelvoudige hypothese [1].

In § 8 is aangekondigd, dat wij nog zouden aantonen, dat de kritieke zones bij de tekentoets het best "in de staarten" genomen konden worden. Wij maken daartoe gebruik van een zeer algemene stelling van J.NEYMAN, die in vele gevallen de mogelijkheid biedt "beste" kritieke zones uit te zoeken. Voor de eenvoud geven wij deze stelling niet in zijn meest algemene vorm. Men raadplege daarvoor b.v. J.NEYMAN, First course in Probability and Statistics, N.Y. 1950, p. 304 e.v.

Wij beschouwen eerst het geval van een enkelvoudige hypothese H_0 (in ons geval $p=\frac{1}{2}$), die getoetst wordt met behulp van een discreet verdeelde toetsingsgrootheid $\underline{x}$ (in ons geval: het aantal successen). Zij verder H_1 een enkelvoudige alternatieve hypothese (in ons geval b.v. $p=0{,}6$). Wij definiëren nu

$$\lambda(x\,;H_0,H_1) \overset{\text{def}}{=} \frac{P[\underline{x}=x\,|\,H_0]}{P[\underline{x}=x\,|\,H_1]}. \tag{45}$$

De stelling van Neyman houdt nu in, dat van alle mogelijke kritieke zones die zone het grootste onderscheidingsvermogen voor H_1 heeft, die zoveel mogelijk waarden x met zo klein mogelijke λ bevat. Preciezer gezegd:

Stelling 1: Indien men een kritieke zone voor toetsing van H_0 wijzigt door één of meer waarden x ervan te vervangen door waarden met $\lambda(x;H_0,H_1)$ groter dan de grootste van die van de vervangen waarden, zonder dat door deze wijziging de onbetrouwbaarheid toeneemt, dan vermindert het onderscheidingsvermogen ten opzichte van H_1.

Deze stelling is intuitief gemakkelijk te begrijpen. Immers waarden x met een kleine λ zijn zeer gunstig voor het onderscheiden van H_0 en H_1: indien H_0 juist is zijn zij in verhouding weinig waarschijnlijk, zodat er nogal wat tot de kritieke zone gerekend kunnen worden, terwijl zij, als H_1 juist is, veel waarschijnlijker zijn, hetgeen het onderscheidingsvermogen groot maakt.

Bewijs van de stelling. Noemen wij de oorspronkelijke kritieke zone Z, dan kunnen wij Z in twee delen Z_1 en Z_2 splitsen, waarbij Z_2 die waarden voorstellen, die door waarden met grotere λ vervangen zullen worden. Geven wij dit "nieuwe" stuk aan met Z_3, dan

[1]) De grootheid λ wordt in de Angelsaksische litteratuur vaak als de "likelihood ratio" aangeduid.

is dus de oude kritieke zone

$$Z = Z_1 \cup Z_2 \qquad (\cup \text{ betekent: verenigd met})$$

en de nieuwe

$$Z' = Z_1 \cup Z_3 \; .$$

Daar nu de onbetrouwbaarheid door de wijziging niet toegenomen mag zijn (zie de stelling), moet gelden:

$$\beta_{Z_2}(H_0) \geqq \beta_{Z_3}(H_0),$$

hetgeen wij ook kunnen schrijven als

(46) $$\sum_{x \in Z_2} P[\underline{x}=x \mid H_0] \geqq \sum_{x \in Z_3} P[\underline{x}=x \mid H_0].$$

Verder is:

(47) $$\beta_{Z_2}(H_1) = \sum_{x \in Z_2} P[\underline{x}=x \mid H_1] = \sum_{x \in Z_2} \frac{P[\underline{x}=x \mid H_0]}{\lambda(x; H_0, H_1)}$$

volgens (45). Analoog geldt voor Z_3:

(48) $$\beta_{Z_3}(H_1) = \sum_{x \in Z_3} \frac{P[\underline{x}=x \mid H_0]}{\lambda(x; H_0, H_1)}.$$

Volgens de onderstelling zijn nu echter alle factoren λ in (48) groter dan die in (47), zodat, tezamen met (46) direct volgt:

$$\beta_{Z_3}(H_1) < \beta_{Z_2}(H_1).$$

De kans, onder H_1, dat $\underline{x}$ in Z_3 zal vallen is dus kleiner dan dat $\underline{x}$ in Z_2 terecht zal komen. Hetzelfde geldt dan echter voor Z' in vergelijking met Z, daar deze twee zones slechts in Z_2 en Z_3 verschillen.

<u>Opmerking</u>: Uit de stelling, in deze speciale vorm, volgt direct, dat wij in het geval van een discrete toetsingsgrootheid $\underline{x}$ de kritieke zone op moeten bouwen uit waarden van $\underline{x}$, die een zo klein mogelijke λ bezitten. Wij beginnen dus met de waarde (of waarden) met de kleinste λ en nemen er vervolgens waarden met zo langzaam mogelijk opklimmende λ's bij, tot de gekozen onbetrouwbaarheidsdrempel ons verhindert verder te gaan. Dit alles geldt slechts, indien wij één speciale alternatieve hypothese H_1 in het oog vatten. Hiermee valt echter, zoals wij zullen zien, soms al een aardig resultaat te bereiken. Een algemenere vorm van het λ-principe zullen wij verderop bespreken.

14. <u>Toepassing op de tekentoets</u>.

Bij de tekentoets is $H_0: p=\frac{1}{2}$ en $\underline{x}$: het aantal successen. Een alternatieve hypothese H_1 wordt aangegeven door de waarde p van de kans op succes. Wij hebben dan (zie (45)):

(49) $$\lambda(x;H_0,H_1)=\frac{\binom{n}{x}2^{-x}}{\binom{n}{x}p^x(1-p)^{n-x}} :: \frac{1}{p^x(1-p)^{n-x}},$$

waarin :: betekent: "op een constante factor na gelijk aan."

Wij zien dus, dat λ, bij gegeven n, voor iedere $p>\frac{1}{2}$ (dus $1-p<\frac{1}{2}$) monotoon daalt bij stijgende x en voor iedere $p<\frac{1}{2}$ monotoon stijgt met x.

Dit betekent echter, dat wij bij toetsing van H_0 tegen iedere H_1 met $p>\frac{1}{2}$ een rechteréénzijdige kritieke zone moeten nemen, en bij toetsing tegen iedere $p<\frac{1}{2}$ een linkeréénzijdige. De éénzijdige toetsingen zijn dus reeds op grond van dit eenvoudige λ-principe te rechtvaardigen. Immers zij zijn, voor alternatieve $p>\frac{1}{2}$ resp. $p<\frac{1}{2}$ de uniform meest onderscheidende toetsen, zoals uit de stelling volgt.

De tweezijdige toets kunnen wij op deze wijze echter nog niet afleiden.

15. Het λ-principe voor samengestelde hypothesen.

Ook hier volstaan wij met het geven van een speciale vorm van het λ-principe, waarbij wij de keuze van de toetsingsgrootheid reeds bepaald veronderstellen.

Voor de generalisatie van het λ-principe bij het toetsen van samengestelde hypothesen tegen samengestelde alternatieven zullen wij bovendien geen stellingen geven. Dit zou te ver voeren. Wij geven slechts de definitie van λ en vermelden, dat A.WALD, "Tests of statistical hypotheses concerning several parameters when the number of observations is large", Am..Math.Soc.Trans., Vol. 54 (1943), p. 426 e.v., bewezen heeft, dat toetsen, die op dit principe gebaseerd zijn, bepaalde optimale eigenschappen bezitten.

Indien men de samengestelde hypothese Ω_0 wenst te toetsen en Ω de verzameling van alle toegelaten hypothesen voorstelt (Ω_0 inbegrepen), dan kan men gebruik maken van

(50) $$\lambda(x;\Omega_0,\Omega)\overset{\text{def}}{=}\frac{\underset{H\in\Omega_0}{\text{Max}}\ P[\underline{x}=x|H]}{\underset{H\in\Omega}{\text{Max}}\ P[\underline{x}=x|H]}.$$

Hierin stelt $\underline{x}$ de toetsingsgrootheid voor, waarvan weer ondersteld is, dat hij discreet verdeeld is. Is $\underline{x}$ continu verdeeld, dan wordt $P[\underline{x}=x|H]$ vervangen door de verdelingsdichtheid van $\underline{x}$ onder hypothese H.

Indien wij in (50) voor Ω_0 een enkelvoudige hypothese H_0 substitueren, neemt (50) de volgende vorm aan:

(50a) $$\lambda(x;H_0,\Omega)=\frac{P[\underline{x}=x|H_0]}{\underset{H\in\Omega}{\text{Max}}\ P[\underline{x}=x|H]}.$$

In dit geval is het ook wel aannemelijk, dat het vormen van kritieke zones uit waarden x met zo klein mogelijke λ's gunstig is. Immers dit betekent voor ieder van die waarden, dat, althans onder sommige van de alternatieve hypothesen, deze waarden met veel grotere waarschijnlijkheid aangenomen worden dan onder H_o. Dit heeft echter een gunstige invloed op het onderscheidingsvermogen. De vorm (50) zelf is vanuit dit gezichtspunt ook wel plausibel. Indien λ in (50) klein is, betekent dit nl., dat, althans onder sommige hypothesen van Ω, de kans op de beschouwde waarde van $\underline{x}$ veel groter is dan de grootste kans op deze waarde van $\underline{x}$ onder Ω_o.

Bij deze intuitieve rechtvaardiging zullen wij het moeten laten. In de volgende paragraaf geven wij enkele toepassingen van het principe.

NB: Wij merken op, dat voor deze vorm van λ geldt:

$$0 \leqq \lambda \leqq 1.$$

16. Toepassing op de tekentoets.

Wij beschouwen nu het geval van de tweezijdige tekentoets, d.w.z. van de hypothese $H_o: p=\frac{1}{2}$ tegen de samengestelde hypothese $p \neq \frac{1}{2}$. Dus met $\Omega: 0 \leqq p \leqq 1$ als verzameling van toegelaten hypothesen. Dan is

$$\lambda(x; H_o, \Omega) = \frac{\binom{n}{x} 2^{-n}}{\underset{0 \leqq p \leqq 1}{\mathrm{Max}} \binom{n}{x} p^x q^{n-x}} \qquad (q=1-p),$$

dus

$$\lambda = 2^{-n} \Big\{ \underset{0 \leqq p \leqq 1}{\mathrm{Max}}\, p^x q^{n-x} \Big\}^{-1}.$$

Voor $x=0$ wordt het maximum tussen de accolades bereikt voor $q=1$ en voor $x=n$ voor $p=1$. Wij hebben dus

$$\lambda(0; H_o, \Omega) = \lambda(n; H_o, \Omega) = 2^{-n}.$$

Voor andere waarden van x sporen wij het maximum tussen de accoladen op door de afgeleide van $p^x q^{n-x}$ te beschouwen. Deze is

$$p^{x-1} q^{n-x-1} (x-np)$$

en deze is dus gelijk aan 0 voor $p=\frac{x}{n}$. De tweede afgeleide is in dit punt $(p=\frac{x}{n})$ gelijk aan:

$$-n\left(\frac{x}{n}\right)^{x-1}\left(1-\frac{x}{n}\right)^{n-x-1} < 0,$$

zodat wij inderdaad met een maximum te maken hebben.

Wij vinden dan dus voor λ:

$$\lambda = 2^{-n}\left\{\left(\frac{x}{n}\right)^x\left(1-\frac{x}{n}\right)^{n-x}\right\}^{-1} = \frac{2^{-n}n^n}{x^x(n-x)^{n-x}} .$$

De teller is constant. Het gedrag van de noemer als functie van x kan men het eenvoudigst bepalen door de afgeleide van de logarithme te beschouwen. Deze is:

$$\frac{d}{dx}\left\{x\ln x + (n-x)\ln(n-x)\right\} = \ln x + 1 - \ln(n-x) - 1 = \ln\frac{x}{n-x} .$$

Deze is dus positief voor $x > n-x$ (dus voor $x > \frac{1}{2}n$)
en negatief voor $x < n-x$ (dus voor $x < \frac{1}{2}n$).
Hieruit blijkt, dat λ afneemt, als x zich van de waarde $\frac{1}{2}n$ af beweegt. Bovendien is voor $x=1$ en $x=n-1$

$$\lambda = \frac{2^{-n}n^n}{(n-1)^{n-1}} = 2^{-n}\, n\left(\frac{n}{n-1}\right)^{n-1} > 2^{-n},$$

dus groter dan de bij $x=0$ en $x=n$ behorende waarde van λ. Daar bovendien de waarde van λ voor twee waarden van x, die symmetrisch t.o.v. $\frac{1}{2}n$ liggen (dus waarvoor $x=n-x$ is) dezelfde is, leidt het bijeenzoeken van die waarden van x, waarvoor λ zo klein mogelijk is, tot een symmetrische tweezijdige kritieke zone Z van het door ons steeds gebruikte type (5).

Opmerking. Op volkomen analoge wijze kan men de toetsing van de samengestelde hypothese $\Omega_0 : 0 \leq p \leq \frac{1}{2}$ tegen $\Omega_1 : \frac{1}{2} < p \leq 1$ (dus weer met $\Omega : 0 \leq p \leq 1$) behandelen. Er komt dan

$$\lambda(x;\Omega_0,\Omega) = \frac{\underset{0\leq p\leq\frac{1}{2}}{\mathrm{Max}} \binom{n}{x} p^x q^{n-x}}{\underset{0\leq p\leq 1}{\mathrm{Max}} \binom{n}{x} p^x q^{n-x}} \quad (q=1-p).$$

Voor $x \leq \frac{1}{2}n$ vindt men nu steeds $\lambda = 1$, daar teller en noemer gelijk worden (het maximum wordt nl. weer aangenomen voor $p=\frac{x}{n}$), terwijl voor $x > \frac{1}{2}n$ het maximum in de teller bereikt wordt voor $p=\frac{1}{2}$ en voor de noemer nog steeds voor $p=\frac{x}{n}$. Bij verder doorrekenen blijkt λ nu voor stijgende x af te nemen, zodat toepassing van het λ-principe nu tot een rechteréénzijdige kritieke zone leidt. Dit is hetzelfde resultaat, als in § 14 op andere wijze werd bereikt. Hetzelfde vonden wij ook in § 12, bij het voorbeeld van de tabaksfabrikant, op grond van redelijke eisen van deze fabrikant.

17. Toepassingen gebaseerd op verschillen.

Laat $\underline{u}$ en $\underline{v}$ twee stochastische grootheden zijn, die dezelfde wh-verdeling bezitten. Onderstel verder, dat $\underline{u}$ en $\underline{v}$ stochastisch onafhankelijk zijn, d.w.z. dat de waarde, die de ene aanneemt de door de andere aangenomen waarde niet beïnvloedt. In formule (vgl. ook § I.8):

(51) $$P[\underline{u} \leq u \wedge \underline{v} \leq v] = P[\underline{u} \leq u]\, P[\underline{v} \leq v]$$

voor iedere u en v.

Beschouw nu de stochastische grootheid

(52) $$\underline{z} = \underline{u} - \underline{v},$$

dan bezit deze dezelfde wh-verdeling als de grootheid $-\underline{z} = \underline{v} - \underline{u}$. Immers verwisseling van $\underline{u}$ en $\underline{v}$ in de gegevens verandert hieraan niets. Als echter een stochastische grootheid $\underline{z}$ dezelfde verdeling bezit als zijn tegengestelde $-\underline{z}$, moet de kans, dat $\underline{z}$ positief is, gelijk zijn aan de kans, dat $-\underline{z}$ positief, dus $\underline{z}$ negatief is. Dus geldt dan

(53) $$P[\underline{z} > 0] = P[\underline{z} < 0].$$

Is $P[\underline{z} = 0] = 0$, dan volgt uit (53):

(53a) $$P[\underline{z} > 0] = P[\underline{z} < 0] = \tfrac{1}{2}.$$

Is $P[\underline{z} = 0] > 0$, dan beschouwen wij

$$P[\underline{z} > 0 \mid \underline{z} \neq 0] = \frac{P[\underline{z} > 0 \wedge \underline{z} \neq 0]}{P[\underline{z} \neq 0]} = \frac{P[\underline{z} > 0]}{P[\underline{z} > 0 \vee \underline{z} < 0]} =$$

$$= \frac{P[\underline{z} > 0]}{P[\underline{z} > 0] + P[\underline{z} < 0]} = \frac{1}{2}$$

(de laatste overgang volgt uit (53)).

Wij hebben dan dus, daar dezelfde berekening voor $\underline{z} < 0$ opgaat:

(53b) $$P[\underline{z} > 0 \mid \underline{z} \neq 0] = P[\underline{z} < 0 \mid \underline{z} \neq 0] = \tfrac{1}{2}.$$

Wij kunnen dit samenvatten in de volgende stelling.

<u>Stelling 2</u>: <u>Indien $\underline{u}$ en $\underline{v}$ stochastisch onafhankelijk zijn en dezelfde wh-verdeling bezitten, dan is de voorwaardelijke wh, dat $\underline{u}-\underline{v}$ positief is, onder de voorwaarde, dat $\underline{u}-\underline{v}$ niet gelijk aan 0 is, gelijk aan $\frac{1}{2}$.</u>

<u>Opmerking</u>: De grootheid $\underline{u}-\underline{v}$ bezit zelfs een verdeling, die symmetrisch t.o.v. 0 is, hetgeen meer is dan in de stelling staat; wij gebruiken hier echter alleen het in de stelling vermelde resultaat.

Voorbeeld [1]). De bepaling van isolerende eigenschappen van verschillende oliesoorten kan geschieden door een glazen buisje met die te vullen en vervolgens op de uiteinden een spanningsverschil aan te brengen, dat men laat stijgen, tot de isolatie verbroken wordt. Men kan deze bepaling van de doorslagspanning met dezelfde olievulling herhalen zo vaak men wil. Wij zullen speciaal de eerste twee bepalingen beschouwen. Geven wij de doorslagspanning bij de eerste bepaling aan met $\underline{u}$ en bij de tweede met $\underline{v}$, dan zal men weliswaar niet verbaasd zijn, dat de uitkomsten van $\underline{u}$ en $\underline{v}$ niet dezelfde zijn, maar het ligt voor de hand te onderstellen, dat $\underline{u}$ en $\underline{v}$ wel dezelfde wh-verdeling bezitten, d.w.z., dat de volgorde der beide bepalingen geen invloed op de uitkomsten heeft.

Om te onderzoeken of dit juist is, werden 14 oliemonsters gebruikt. Van ieder daarvan werden twee vullingen gebruikt en bij iedere vulling werden twee opeenvolgende bepalingen van de doorslagspanning verricht. De bij deze paren van bepalingen gevonden verschillen tussen de tweede en de eerste bepaling zijn in tabel III vermeld.

Tabel III

Verschillen (in K.V.) van waargenomen doorslagspanningen.

olie-monster no.	bij eerste vulling z_1	bij tweede vulling z_2	$z_1 - z_2$
1	3	-1	4
2	2	1	1
3	4	2	2
4	4	5	-1
5	1	-2	3
6	5	1	4
7	6	3	3
8	0	0	0
9	-1	7	-8
10	1	5	-4
11	7	1	6
12	-1	-1	0
13	1	0	1
14	1	8	-7
n	13	12	12
x	2	3	4

z_1 resp. z_2 = uitkomst van de tweede bepaling verminderd met die van de eerste, bij eerste resp. tweede vulling.
n = aantal verschillen, die ongelijk aan 0 zijn.
x = aantal negatieve verschillen.

---

1) Ontleend aan W.J.YOUDEN 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.

Wij zullen nu de hypothese H_0 toetsen, dat de volgorde der bepaling geen invloed op de uitkomst heeft, dus dat zowel z_1 als z_2 een even grote wh bezit om positief als negatief te zijn. Wij merken daarbij op, dat de wh-verdelingen van u en v, dus van doorslagspanningen bij de eerste en de tweede bepaling, voor de verschillende oliemonsters, of zelfs voor de vullingen, niet dezelfde behoeven te zijn. De verdelingen van z_1 en z_2 mogen verschillen, ook die van z_1 en z_2 afzonderlijk voor de verschillende monsters. Volgens stelling 2 kunnen wij de tekentoets op de resultaten toepassen, daar deze stelling, als H_0 juist is, voor ieder der proeven afzonderlijk opgaat. Laten wij dus de verschillen, die gelijk aan 0 zijn, bij de toetsing weg [1]) en gebruiken wij het aantal negatieve verschillen (dat wij x noemen) als toetsingsgrootheid, dan kunnen wij H_0 met de tekentoets toetsen.

Wij kunnen de in de tweede en derde kolommen vermelde uitkomsten (van z_1 en z_2) als 28 onafhankelijke proeven beschouwen. Hieronder zijn er 25 met een uitkomst $\neq 0$, en daarvan zijn 5 verschillen negatief. Uit tabel I van dit hoofdstuk blijkt, dat voor $n=25$ het optreden van slechts 5 negatieve verschillen een tweezijdige overschrijdingskans bezit, die kleiner dan 0,01 is. Op grond hiervan besluiten wij tot verwerping van H_0. De uitkomsten der tweede bepaling liggen kennelijk hoger dan die van de eerste.

Beschouwen wij de tweede en de derde kolom apart, op grond van de overweging, dat dit effect misschien speciaal bij één van beide vullingen optreedt, dan vinden wij, weer uit tabel I, bij de tweede kolom ($n=13$, $x=2$) een overschrijdingskans tussen 0,01 en 0,05 en bij de derde ($n=12$, $x=3$) één tussen 0,1 en 0,25. Het verschil tussen de eerste en de tweede bepaling is dus bij de eerste vulling op zichzelf nog wel aantoonbaar, maar bij de tweede niet meer. Daarmee is echter nog geen verschil aangetoond tussen de eerste en de tweede vulling, want het feit, dat bij

---

1) Voor de behandeling der verschillen, die gelijk aan 0 zijn, worden soms ook andere voorstellen gedaan, zoals: de helft als positief en de andere helft als negatief tellen. Dit komt het onderscheidingsvermogen echter niet ten goede. Weglaten is beter. Zie hiervoor: J.HEMELRIJK, A theorem on the sign test when ties are present, Proc.Kon.Ned.Ak.v.Wet. A 55, Indag.Math. 14 (1952), pp. 322-326.

de tweede vulling een grotere overschrijdingskans gevonden wordt is niet verbazingwekkend; deze overschrijdingskans moet tenslotte bij één van de twee vullingen wel groter zijn dan bij de andere, men kan er niet op rekenen, dat beide precies dezelfde uitkomst zullen geven. Het is zeer wel mogelijk -een vergelijking van de twee kolommen verschillen doet wel vermoeden, dat dit inderdaad zo is; er is weinig verschil in deze twee reeksen- dat bij een groter aantal proeven ook de derde kolom een kleine overschrijdingskans gegeven zou hebben (volgens § 11 neemt het onderscheidinsgvermogen met het aantal proeven toe).

Wij gaan dus ook nog na, of de verzamelde gegevens reden geven, om een systematisch verschil te vermoeden tussen de grootheden $\underline{z}_1$ en $\underline{z}_2$, die bij de eerste resp. tweede vullingen behoren. Daartoe beschouwen wij de in de laatste kolom vermelde verschillen $z_1 - z_2$ en passen daarop opnieuw de tekentoets toe. Wij vinden dan $n=12$ en $x=4$, waarbij volgens tabel I een overschrijdingskans $>0{,}25$ behoort, zodat er geen reden tot verwerping van de getoetste hypothese is, die in dit geval inhoudt, dat $\underline{z}_1$ en $\underline{z}_2$ voor ieder van de monsters afzonderlijk dezelfde wh-verdeling bezitten. De boven reeds toegepaste combinatie van de tweede en derde kolom tot één reeks van 28 proeven wordt door dit resultaat gerechtvaardigd.

Onze conclusie luidt dus: de tweede bepaling geeft een systematisch hogere uitkomst dan de eerste; tussen de beide vullingen met één monster is in dit opzicht geen verschil aangetoond. De verklaring van het geconstateerde verschil is de taak van de physicus, niet van de statisticus.

Opmerking. Men kan op dit punt de tegenwerping maken, dat de uitkomsten van de tweede en van de derde kolom zo weinig verschillen ($n=12$ bij beide en $x=3$ resp. 4), dat de conclusie hierdoor een gezocht karakter heeft. Als men op grond van de uitkomsten in de derde kolom niet tot verwerping van de getoetste hypothese overgaat, kan men dit bij de tweede evenmin doen. Het boven gebruikte argument, dat er slechts 14 proeven genomen zijn, geldt niet alleen voor de derde, maar ook voor de vierde kolom. Deze tegenwerping is gerechtvaardigd, als wij ons uitsluitend van de tekentoets bedienen. Beschouwen wij echter niet alleen het teken van de verschillen z_1, z_2 en $z_1 - z_2$, dan zien wij, dat bij z_1 en z_2 de negatieve uitkomsten niet alleen geringer in aantal zijn dan de positieve, maar dat zij bovendien veel dichter bij 0 liggen. Bij de negatieve verschillen onder $z_1 - z_2$ is dit echter niet het geval. Overwegen wij nu, dat onder de getoetste hypothesen niet alleen de kans op een negatief verschil gelijk is aan die op een positief verschil, maar dat de verschillen symmetrisch t.o.v. 0 verdeeld zijn, dan zien wij, dat in de laatste kolom wèl veel minder argumenten tegen deze symmetrie aanwezig zijn dan in de tweede en ook in de derde. Dit ondersteunt onze conclusie.

Hiermee komt tevens een zwakheid van de tekentoets naar voren, althans bij toepassing op verschillen van waarnemingen. Zouden wij, behalve met de tekens van de verschillen ook nog met hun absolute grootte rekening kunnen houden, dan zouden wij ongetwijfeld het onderscheidingsvermogen van de toets vergroten. En dit is inderdaad mogelijk, zelfs op verschillende manieren. Wij geven hiertoe enkele litteratuurverwijzingen.

1. Indien de verschillen normaal verdeeld zijn, met gelijke spreidingen, kan men de toets van Student toepassen voor de hypothese, dat de mathematische verwachting (het gemiddelde) van de wh-verdeling der verschillen gelijk aan 0 is. Zie b.v. W.J.DIXON and F.J.MASSEY, "Introduction to statistical analysis," N.Y. 1951, pp. 97-100 (men stelle $\mu_0=0$). Wij merken op, dat het onderhavige materiaal geen garantie biedt voor het vervuld zijn van de nodige voorwaarden. Vooral de gelijkheid der spreidingen is een bedenkelijke voorwaarde, daar verschillende oliemonsters zijn gebruikt.

2. Zonder verdere onderstellingen te maken, dan bij de tekentoets nodig is (dus uitsluitend onderlinge onafhankelijkheid der verschillende proeven; een voorwaarde, die men zonder bezwaar kan stellen), kan men zich bedienen van symmetrietoetsen: J.HEMELRIJK,"Symmetrietoetsen," Diss.,Den Haag 1950. Verder zijn toetsen van een soortgelijk karakter ontwikkeld door J.E.WALSH. Zie b.v. DIXON and MASSEY, pp. 252-254.

18. Opmerkingen.

In § 6 is aangekondigd, dat men een tweezijdige toets nog uitvoeriger kan beschrijven, dan door het onderscheidingsvermogen alleen. Daarmee is het volgende bedoeld. Indien wij de hypothese $p=\frac{1}{2}$ tweezijdig toetsen, beschouwen wij zowel de alternatieve hypothese met $p<\frac{1}{2}$ als die met $p>\frac{1}{2}$ als toegelaten. Bij verwerping van de getoetste hypothese besluiten wij ofwel tot $p<\frac{1}{2}$ ofwel tot $p>\frac{1}{2}$. Behalve de reeds genoemde mogelijke foute conclusies, kunnen wij dus ook nog een tot nu toe verwaarloosde mogelijkheid beschouwen: p kan in werkelijkheid $<\frac{1}{2}$ zijn, terwijl wij besluiten tot $p>\frac{1}{2}$. Wij verkrijgen daarom een nog betere beschrijving van de toets dan door fig. 2, indien wij voor iedere waarde van p de kans op de drie mogelijke uitkomsten aangeven. Deze mogelijke uitkomsten zijn:

1. $p=\frac{1}{2}$ wordt niet verworpen,
2. wij besluiten tot $p<\frac{1}{2}$,
3. wij besluiten tot $p>\frac{1}{2}$.

De som van de kansen op 1., 2. en 3. is, voor iedere waarde van p, gelijk aan 1. Wij besluiten tot 1., als de gevonden waarde x van $\underline{x}$ niet in de kritieke zone ligt; tot 2., indien x in het linkerdeel en tot 3., indien x in het rechterdeel van deze zone valt. Op grond hiervan kunnen wij de kansen op 1., 2. en 3. berekenen op dezelfde wijze als vroeger met het onderscheidingsvermogen gedaan is. Voor $n=14$ en $\alpha=0{,}05$ (het geval van § 7; dus de Z van blz. 38 bovenaan), verkrijgen wij dan fig. 7. De kansen op de 3 beslissingen hebben wij daarin aangegeven als $\beta_1(p)$, $\beta_2(p)$ resp. $\beta_3(p)$. Men ziet, dat de kans, om tot $p<\frac{1}{2}$ te besluiten, als $p>\frac{1}{2}$ is zeer gering is, vooral als p b.v. $>0{,}6$ is. Het omgekeerde geldt natuurlijk evenzeer.

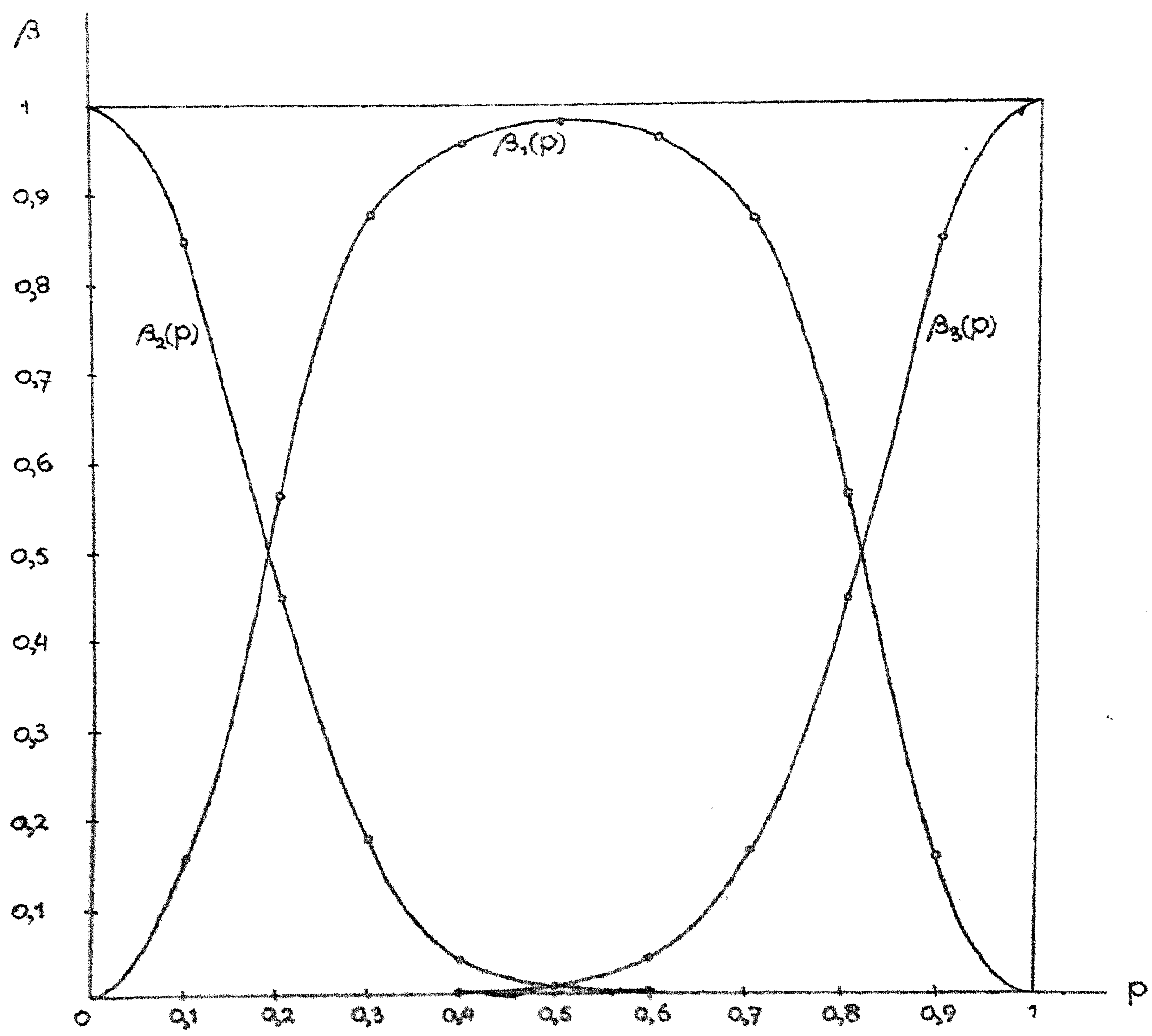


Fig. 7. De volledige karakteristiek van de tweezijdige tekentoets voor $n=14$, $\alpha=0{,}05$ ($Z: x\leq 2$ of $x \geq 12$).

$\beta_1(p)$ = kans, om $p=\frac{1}{2}$ niet te verwerpen,

$\beta_2(p)$ = kans, om tot $p<\frac{1}{2}$ te besluiten,

$\beta_3(p)$ = kans, om tot $p>\frac{1}{2}$ te besluiten.

Tenslotte is het wellicht niet overbodig, op te merken, dat de in dit hoofdstuk besproken tekentoets slechts één toets uit velen is. Wij hebben deze toets gekozen ter illustratie van de begrippen van de toetsingstheorie, omdat het de eenvoudigste toets is, die beschikbaar is. Voor een verzameling van andere toetsen verwijzen wij naar het boek van DIXON and MASSEY.