

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Rapport S 199 (V14)

De sequente binomiale toets:

A. H. I. De statistische techniek

door

Ph. van Elteren

Mei 1956

De Statistische techniek van de sequente binomiale toets

door

Ph. van Elteren, medewerker van de Statistische Afdeling van het Mathematisch Centrum.

1. Inleiding

Wanneer men de gangbare klassieke statistische toetsingsmethoden wil toepassen, dient men tevoren het aantal te verrichten waarnemingen vast te stellen. Bij farmacologische experimenten in het laboratorium moet dus tevoren worden vastgelegd hoeveel proefdieren men zal gebruiken. Bij het onderzoek van een geneesmiddel in de kliniek dient reeds bij de aanvang te worden bepaald bij welk aantal patiënten het te bestuderen middel zal worden toegepast.

Bij vele methoden is het mogelijk op verantwoorde wijze het aantal waarnemingen vast te leggen, b.v. door te eisen, dat men een effect van een gegeven grootte met een gegeven waarschijnlijkheid kan aantonen. Indien het effect echter veel groter is dan men verwacht, zal men in de regel een overduidelijk resultaat vinden en het gevoel hebben, dat men niet zuinig genoeg te werk is gegaan. Dit is bijzonder onaangenaam bij experimenten in de kliniek, waar men om ethische redenen liefst zo min mogelijk patiënten in het onderzoek wil betrekken.

Ook is het mogelijk, dat men het effect juist niet kan aantonen op grond van de verrichte waarnemingen. De verleiding is dan groot, om nog enkele waarnemingen toe te voegen en dan de toets nog eens op het uitgebreide materiaal toe te passen. Dit is echter, zoals reeds gezegd, bij de gangbare toetsen niet toegestaan. De reden hiervan is de volgende: de genoemde toetsen zijn zodanig geconstrueerd, dat men een gegeven kleine kans α (de zg. onbetrouwbaarheidsdrempel) riskeert om, indien men het vastgestelde aantal waarnemingen verricht, te besluiten tot het bestaan van het gezochte effect, als dit er niet is. Gaat men echter door na dit vastgestelde aantal, dan vermeerdert men de onbetrouwbaarheidsdrempel α , met de kans dat men bij het vastgestelde aantal het gezochte effect niet, doch met de uitgebreide waarnemingsreeks wel kan aantonen. Men vergroot dus de kans op het aantonen van een effect dat er niet is, dus op een foutieve conclusie. Het bedrag waarmee deze kans toeneemt, is echter niet gemakkelijk uit te rekenen. De enige toelaatbare methode is om de proef te herhalen met een groter aantal waarnemingen.

Al deze moeilijkheden doen zich bij de toepassing van de sequente methoden niet voor. Het is bij deze methoden slechts vereist, dat men tevoren de onbetrouwbaarheidsdrempel vaststelt, alsmede de kans die men wil riskeren om een effect van gegeven grootte ten onrechte niet aan te tonen. Door deze gegevens wordt een procédé vastgelegd, waarbij men begint met een zeker minimum aantal waarnemingen. Indien het resultaat reeds voldoende duidelijk is, kan men een conclusie trekken betreffende het onderzochte effect; zo niet, dan moet men een nieuwe waarneming verrichten, waarna men of een conclusie kan trekken of weer een waarneming moet verrichten en zo voorts tot men tot een conclusie komt. Het procédé is steeds zo gedefinieerd, dat het zeker afloopt en wel met gemiddeld minder waarnemingen dan men nodig heeft bij de overeenkomstige klassieke toets.

2. Beschrijving van de toets

Wij zullen ons hier verder bezighouden met een bepaalde sequente toets, namelijk de sequente binomiale toets. Deze methode kan gebruikt worden voor het toetsen van hypothesen betreffende onbekende fracties of kansen. Men kan dit gemakkelijk uitleggen aan de hand van een vaas met een groot aantal ballen waarvan een onbekende fractie p rood gekeurd is en de rest wit. Indien men zonder in de vaas te kijken een bal trekt, heeft men een kans p dat het een rode bal zal zijn. Uit de fractie rode ballen in een reeks trekkingen wil men nu een conclusie trekken betreffende p .

Zoals bekend wordt deze vaas als model gebruikt voor vele soorten experimenten. Indien men b.v. geneesmiddelen beproeft bij ratten, komen alle beschikbare ratten van een bepaalde stam overeen met de ballen in de vaas; de ratten die gevoelig zijn voor het middel met de rode en de ratten die ongevoelig zijn met de witte ballen. Het toedienen van het middel bij een aantal ratten komt dan overeen met het nemen van een steekproef uit de vaas. Een ander voorbeeld heeft men in de kwaliteitscontrole. De vaas met ballen komt dan overeen met een partij exemplaren van een massa-artikel, de rode ballen zijn dan de defecte exemplaren. Voor dit toepassingsgebied zijn de sequente methoden aanvankelijk ontwikkeld; wij hebben hier echter meer farmacologische toepassingen op het oog. In de lezing van Dr Rümke zal een aantal voorbeelden op dit gebied besproken worden. Wij zullen echter verder met het model van de vaas met ballen werken.

Bij de sequente binomiale toets stelt men zich ten doel om de hypothese te toetsen dat de fractie rode ballen in de vaas niet groter is dan p_0 . Men wil dit doen aan de hand van een betrekkelijk kleine steekproef en kan dus uiteraard niet eisen dat men met zekerheid te weten komt of $p \leq p_0$ dan wel $p > p_0$ is. Daarom stelt men de gewone eis: als $p = p_0$ is, wil men een kans toelaten om de hypothese $p \leq p_0$ te verwerpen die niet groter is dan α , de zogenaamde onbetrouwbaarheidsdrempel. Zoals in de inleiding vermeld is, moeten wij ook nog vastleggen welke kans wij willen riskeren om een effect van gegeven grootte niet te vinden; dit betekent voor ons geval: wij moeten aangeven welke kans β wij ten hoogste willen riskeren om $p \leq p_0$ niet te verwerpen, als de fractie p een gegeven waarde $p_1 > p_0$ heeft. De grootheden p_0, p_1, α en β bepalen dan het procédé.

Om een voorbeeld de tegeven: indien wij kiezen $p_0 = 0,5$, $p_1 = 0,8$, $\alpha = 0,05$ $\beta = 0,01$, toetsen wij dus de hypothese, dat ten hoogste de helft van de ballen in de vaas rood is. Indien precies de helft rood is, hebben wij een kans $\frac{1}{20}$, om te besluiten dat meer dan de helft rood is; indien 80% van de ballen rood is, hebben wij een kans om te besluiten dat minder dan de helft rood is van $\frac{1}{100}$.

De toets kan op twee manieren worden toegepast:

a. De grafische methode. Hiertoe zetten wij de waarnemingsreeks uit in een grafiek. Men begint daarbij in de oorsprong. Voor iedere witte bal die wij aantreffen gaan wij één eenheid naar rechts, voor iedere rode één eenheid naar boven. De waarnemingsreeks wordt zodoende voor gesteld door een traplijn. Het punt dat na n trekkingen bereikt wordt, heeft dan als x -coördinaat het aantal witte en als y -coördinaat het aantal rode ballen dat getrokken is. Tevoren trekt men in de grafiek twee evenwijdige rechte lijnen L_0 en L_1 , waarvan de helling s en de stukken h_0 resp. h_1 , die zij afsnijden van de y -as bepaald worden door de grootheden p_0, p_1, α en β (formules (1)-(7); figuren 1 en 2).

Na iedere trekking gaat men nu na waar het eindpunt der traplijn ligt die de waarnemingsreeks voorstelt. Ligt dit punt tussen de lijnen L_0 en L_1 , dan trekt men een nieuwe bal. Ligt het boven de lijn L_1 , dan trekt men géén nieuwe bal meer en verwerpt men de hypothese $p \leq p_0$, en neemt men dus aan dat de fractie rode ballen groter is dan p_0 ; ligt het beneden de lijn L_0 , dan wordt het experiment eveneens beëindigd en verwerpt men de hypothese $p \geq p_1$. De kans dat L_1 overschreden en dus $p \leq p_0$ verworpen wordt, is, als $p = p_0$ is, gelijk aan α ; de kans dat L_0 overschreden en dus $p \geq p_1$ verworpen wordt is, als $p = p_1$ is, gelijk aan β .

In vele gevallen wordt niet y tegen x uitgezet, doch tegen het totale aantal getrokken ballen $n = x + y$. In dat geval moet men bij iedere trekking één eenheid naar rechts gaan, en bovendien één eenheid naar boven, als een rode bal getrokken is. De grenzen worden dan eveneens twee evenwijdige lijnen (zie formules (8) en (9) en figuur 3). Uit deze formules kan men gemakkelijk voor iedere waarde van n de waarde a_n voor het aantal rode ballen berekenen, waarbij L_0 juist overschreden wordt, en evenzo de waarde b_n , waarbij L_1 juist overschreden wordt. Precies gedefinieerd is a_n het grootste gehele getal dat niet groter is dan:

$$\frac{h_0}{1+s} + \frac{s}{1+s} n$$

en b_n het kleinste gehele getal dat niet kleiner is dan:

$$\frac{h_1}{1+s} + \frac{s}{1+s} n$$

Hierop berust:

b. De tabellarische methode. Men berekent vóór het experiment de waarden van a_n en b_n voor $n = 1, 2, 3, \dots$ tot b.v. $n = 30$ (zie b.v. tabel 4 en tabel 7). Daarna schrijft men voor $n = 1, 2, 3, \dots$ het aantal getelde rode ballen d_n op; dit wordt dan een geleidelijk oplopende getallenrij. Zodra d_n gelijk is aan b_n , wordt de hypothese $p \leq p_0$ verworpen, en zodra d_n gelijk is aan a_n de hypothese $p \geq p_1$ (zie tabel 5).

Werkingskromme, kans op een foutieve conclusie

Wij hebben in het voorafgaande gezien, dat de sequente binomiale toets met zekerheid afloopt en dat dit kan gebeuren, door: A de lijn L_0 het eerste wordt overschreden en

B de lijn L_1 het eerste wordt overschreden.

Hieruit volgt dat de kans op mogelijkheid A en de kans op mogelijkheid B tezamen steeds 1 zijn. De kans op A hangt af van de fractie p der rode ballen in de vaas. Indien b.v. alle ballen rood zijn ($p=1$) dan zal men uitsluitend rode ballen trekken en zal de trekkingsreeks in figuur 1 voorgesteld worden door punten op de y -as. In dat geval zal men dus met zekerheid de lijn L_1 snijden, zodat de kans op A dan 0 is. Zijn daarentegen alle ballen wit ($p=0$) dan ziet men gemakkelijk in dat de kans op A gelijk is aan 1. Verder volgt gemakkelijk uit de definitie van de toets, dat voor $p=p_0$ de kans op A gelijk is aan $1-\alpha$, voor $p=p_1$ gelijk aan β .

Wij zullen hier de kans op A verder voorstellen door het symbool $L(p)$; de letter p is in het symbool opgenomen, omdat de kans op A van p afhangt. In tabel 5 vindt men de formules voor $L(p)$ bij $p=0$; p_0 ; $\frac{s}{1+s}$; p_1 en 1. Voor andere waarden van p is $L(p)$ moeilijker te berekenen, doch in de regel heeft men aan de genoemde waarden voldoende om het verband tussen $L(p)$ en p in een grafiek te schetsen. Men dient daarbij te bedenken, dat $L(p)$ afneemt, naarmate p groter wordt. Immers hoe groter p is, des te groter is de kans dat wij ons bewegen in de richting van de grens L_1 , en des te kleiner is dus de kans op de mogelijkheid A. Het verband tussen p en $L(p)$ hebben wij in figuur 5 weergegeven door een kromme voor de gevallen $p_0 = 0,5$, $p_1 = 0,8$, $\alpha = 0,05$, $\beta = 0,01$ én $\beta = 0,05$. Men noemt deze kromme in het Engels de "operating characteristic" van de toets, hetgeen wij hier zullen vertalen door werkingskromme (WK)

Indien wij $L(p)$ voor alle waarden van p kennen, kunnen wij ook afleiden hoe groot de kans is om een foutieve conclusie te maken. Er zijn hierbij 3 gevallen te onderscheiden:

1e. Als $p \leq p_0$ is, maken wij een fout, indien wij L_1 overschrijden; dan verwerpen wij immers $p \leq p_0$; de kans dat L_1 overschreden wordt is $1 -$ (de kans dat L_0 overschreden wordt), en dus, zoals wij zien in figuur 4 hoogstens gelijk aan α .

2e. Als $p_0 < p < p_1$ is, zijn beide mogelijke conclusies goed. De kans op een foutieve conclusie is dan dus gelijk aan 0.

3e. Als $p \geq p_1$ is, maken wij een fout, indien wij L_0 overschrijden; de kans dat dit gebeurt is $L(p)$ en hoogstens gelijk aan β .

In figuur 4 hebben wij door de onderbroken lijn het verloop van de kans $L(p)$ weergegeven en door de getrokken lijn het verloop van de kans op een foutieve conclusie.

Uit ons overzicht blijkt, dat de kans op een foutieve conclusie nergens groter is dan de grootste van de twee grootheden α en β .

4. Gemiddelde steekproefgrootte G.S.G.

Op grond van het voorafgaande zal men wellicht menen, dat het verstandig is p_0 en p_1 dicht bij elkaar te leggen en α en β klein te nemen. Men verkrijgt dan immers een scherpe scheiding: men verwerpt $p \leq p_0$ bijna zeker, als p maar een weinig groter is dan p_0 en omgekeerd: men verwerpt $p \geq p_1$ bijna zeker, als p maar een weinig kleiner is dan p_0 . Een dergelijke scherpe scheiding moet echter betaald worden met een gemiddeld groter aantal trekkingen. Het is daarom van belang om na te gaan hoeveel trekkingen

men gemiddeld nodig heeft om tot een beslissing te komen. Men vermijdt dan dat men zo hoge eisen stelt, dat het vereiste aantal trekkingen prohibitief groot wordt.

Wij willen er hierbij met nadruk op wijzen, dat men het aantal trekkingen niet tevoren kan voorspellen; men weet slechts dat dit aantal een kansverdeling heeft, die afhangt van de constanten p_0, p_1, α en β , alsmede van de fractie p van de rode ballen in de vaas. Deze verdeling is moeilijk te bepalen; gewoonlijk volstaat men met het gemiddelde hiervan op te geven voor een aantal waarden van p . Dit gemiddelde wordt in het Engels "average sample number" genoemd; in het Nederlands vertalen wij dit door gemiddelde steekproefgrootte (afgekort G.S.G.)

In de uiterste gevallen: $p = 0$ en $p = 1$ wordt dit zeer eenvoudig; in dat geval ligt nl. het aantal waarnemingen wél vooruit vast, omdat de steekproef verloopt langs de x -as resp. de y -as (zie figuur 1). Men vindt voor dit aantal dan $-\frac{h_0}{s}$ resp. h_1 , of juist: het kleinste gehele getal dat groter is dan $-\frac{h_0}{s}$ resp. h_1 . Voor de waarde van de G.S.G. bij de andere waarden van p verwijzen wij naar tabel 2 en de bijbehorende figuur 6. In de figuur is het verband tussen de G.S.G. en p weergegeven voor gegeven waarden van p_0, p_1, α en β . Wij zien dat de G.S.G. laag is voor waarden van p dichtbij 0 en 1, en dat zij haar grootste waarde ergens tussen p_0 en p_1 bereikt. Deze kromme komt hoger boven de x -as te liggen (de G.S.G. neemt toe), naarmate p_0 en p_1 dichterbij elkaar komen en α en β kleiner worden.

Men kan de G.S.G. verder gebruiken om de doeltreffendheid van de sequente binomiale toets te vergelijken met die van andere toetsen voor hetzelfde doel opgesteld. Men moet dan, om een eerlijke vergelijking mogelijk te maken, een toets kiezen, waarvan de W.K. zoveel mogelijk met die van de sequente toets overeenstemt, dan zal immers ook de kans op een foutieve conclusie gelijk zijn. Het blijkt dan ^{dat} de G.S.G. van de bovenbeschreven toets steeds kleiner is dan het vaste aantal waarnemingen benodigd voor de overeenkomstige klassieke toets.

Hoewel de G.S.G. van de sequente binomiale toets dus relatief klein is, loopt men bij deze toets toch het risico dat het aantal waarnemingen zeer groot wordt, voor een beslissing kan worden genomen. Om praktische redenen kan het gewenst zijn een maximum te stellen voor het aantal waarnemingen dat men wil verrichten. Indien men dit maximum n_m bereikt vóórdat L_0 of L_1 overschreden is, neemt men toch een beslissing b.v. als:

$y > \frac{h_0 + h_1}{2} + sx$ verwerpt men $p \leq p_0$ en als

$y < \frac{h_0 + h_1}{2} + sx$ verwerpt men $p \geq p_1$.

Men noemt dit procédé afknotting van het sequente proces. Op deze wijze verhoogt men de kans op een foutieve conclusie, doch vermindert men uiteraard de G.S.G.

Indien n_m niet te klein gekozen wordt, d.w.z. op minstens 2 à 3 maal de maximale waarde van de G.S.G. wordt gesteld, is de invloed van de afknotting op de kans op een foutieve conclusie en de G.S.G. gering.

5. Toets met drie beslissingsmogelijkheden

De boven beschreven toets leidt tot de conclusies: verwerp $p \leq p_0$ of verwerp $p \geq p_1$. Deze mogelijkheden zitten ook in de overeenkomstige klassieke toets, hoewel men daarbij gewoonlijk slechts let of één van beide mogelijke conclusies. Er zijn echter ook gevallen, waarbij men wil weten of p in de buurt van een bepaalde waarde ligt, dan wel veel groter of veel kleiner is. De klassieke toets voor dit geval is een tweezijdige binomiale toets. De overeenkomstige sequente toets is de sequente binomiale toets met 3 beslissingsmogelijkheden, welke ontwikkeld is door J. DE BOER naar het voorbeeld van andere dergelijke toetsen.

Het principe van deze toets is zeer eenvoudig, zij is samengesteld uit twee sequente binomiale toetsen T en T' met bepalende grootheden: p_0, p_1, α en β resp. p'_0, p'_1, α' en β' . Daarbij kiezen men $p_0 < p_1 < p'_0 < p'_1$. Deze toets wordt grafisch uitgevoerd door de grenslijnen L_0, L_1, L'_0 en L'_1 van de beide toetsen te bepalen, zoals beschreven is in par. 2. Daarna gaat men de waarnemingen in de grafiek uitzetten (zie figuur 7). Zodra men één der grenslijnen van één van beide toetsen overschrijdt, trekt men de bij die toets behorende conclusie. Daarna gaat men verder, tot men ook één der grenslijnen van de andere toets overschrijdt, waarna de waarnemingen worden beëindigd en de bij de tweede toets behorende conclusie wordt toegevoegd. Het is nu mogelijk, dat er twee tegenstrijdige conclusies uit de bus komen: b.v. indien men eerst L_0 overschrijdt en dan L'_1 . Men verwerpt dan eerst $p \geq p_1$ volgens T en dan $p \leq p'_0$ volgens T' , hetgeen tegenstrijdig is, omdat wij ondersteld hebben, dat $p'_0 > p_1$ is. Deze moeilijkheid doet zich niet voor, indien L_0 de x -as verder van O snijdt dan L'_0 en L'_1 de y -as verder van O dan L_1 . In dat geval is bij overschrijding van L_0 steeds L'_0 al overschreden en dus T' reeds afgelopen; de conclusies zijn achtereenvolgens:

verwerp $p \geq p_1'$ en verwerp $p \geq p_1$, welke niet tegenstrijdig zijn. Aan de gestelde voorwaarde is o.a. voldaan indien men de bepaalde grootheden zo kiest dat geldt:

$$p_0 + p_1' = 1 \quad ; \quad p_1 + p_0' = 1 \quad ; \quad \alpha = \beta' = 2\beta = 2\alpha' = \gamma$$

Dit geval zullen wij meer in het bijzonder beschouwen. Er doen zich nu de volgende mogelijkheden voor:

- 1e. Men overschrijdt eerst L_0' en dan L_0 . Men verwerpt dan uiteindelijk $p \geq p_1$,
- 2e. Men overschrijdt eerst L_0' en vervolgens L_1 of eerst L_1 en dan L_0' . Nu verwerpt men $p \leq p_0$ en $p \geq p_1'$.
- 3e. Men overschrijdt eerst L_1' en dan L_1 . Dan verwerpt men $p \leq p_0'$

De 2e mogelijkheid doet zich voor als men uiteindelijk belandt in het binnengebied van de hoek $L_0 A L_1'$, het is echter ook mogelijk dat men eerst vóór het punt A de lijn L_0' tweemaal passeert en dan L_1 (of tweemaal L_1 en dan L_0').

Door gebruik te maken van de werkingskrommen van de toetsen T en T' kan men voor iedere waarde van p de kans op een foutieve conclusie bepalen. Dit zou ons hier echter te ver voeren. Men kan aldus aantonen, dat deze kans nergens groter is dan γ . De G.S.G. voor de toets met drie beslissingsmogelijkheden vertoont twee maxima resp. in het interval $p_0 < p < p_1$ en in het interval $p_0' < p < p_1'$. Verder is zij steeds groter dan de G.S.G. van T en van T' . Ook op de G.S.G. kunnen wij niet nader ingaan.

De toets met 3 beslissingsmogelijkheden kan ook tabellarisch worden toegepast. Men berekent daartoe eerst de kritieke waarden a_n, b_n, a_n' en b_n' van de samenstellende toetsen (zie tabel 8). Vervolgens schrijft men weer voor iedere n het aantal getelde rode ballen d_n op (zie tabel 9). Dit doet men bij voorkeur in de kolom tussen b_n en a_n' . Indien d_n eerst gelijk wordt aan a_n' (de kolom rechts van d_n) is de toets T' afgelopen met de verwerping van $p \geq p_1'$; men gaat nu naar de ruimte tussen de kolommen a_n en b_n (dus aan de linkerkant) over; indien nu d_n voor een gelijke of hogere waarde van n gelijk wordt aan a_n verwerpt men $p \geq p_1$; bereikt men daarentegen één van de waarden b_n , dan verwerpt men $p \leq p_0$ (naast $p \geq p_1'$). De lezer wordt verzocht zelf na te gaan welke mogelijkheden er zijn, indien d_n eerste gelijk wordt aan b_n .

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Rapport S 199 (V14)

De sequente binomiale toets

II. Enkele toepassingen op farmacologisch terrein

door

Dr Chr.L. Rünke, verbonden aan het
Laboratorium voor Farmacologie en Neu-
robiochemie der Vrije Universiteit.

Mei 1956

Wanneer men een sequente binomiale toets wil toepassen, dienen voor de bepalende grootheden β_0, β_1, α en β zekere waarden te worden gekozen. Het spreekt vanzelf, dat deze waarden in het algemeen van geval tot geval verschillen. In het volgende willen wij bij wijze van voorbeeld uiteenzetten, welk soort overwegingen tot een bepaalde keuze kunnen leiden. Bovendien is ons gebleken, dat de toets met de te beschrijven waarden voor de bepalende grootheden zeer bruikbaar is bij vele verschillende soorten vraagstukken, waarmee wij ons in ons laboratorium bezig plegen te houden.

In de praktijk kan men de sequente binomiale toets dikwijls toepassen, waar men tot dusverre de tekentoets gebruikte. De in het voorafgaande beschreven binomiale toets met twee beslissingsmogelijkheden moet dan met de éézijdige tekentoets worden vergeleken.

Laten wij als voorbeeld veronderstellen, dat wij willen onderzoeken of onder invloed van een bepaald geneesmiddel, alloxaan, het bloedsuikergehalte in het eerste halve uur na inspuiting daalt. Volgens de klassieke werkwijze moeten wij voor het begin van de proef vaststellen bij hoeveel ratten wij dit gaan onderzoeken, wanneer wij later de tekentoets willen toepassen. Meestal waren dat bij ons werk ongeveer dertig ratten. Wanneer wij nog meer dieren nodig zouden hebben om het effect aan te tonen, zal de invloed van alloxaan op het bloedsuikergehalte wel niet erg groot zijn, en dan interesseert het effect ons niet bijzonder. Als onbetrouwbaarheidsdrempel voor de uit te voeren tekentoets kiezen wij 0,05.

Wij toetsen met de tekentoets de hypothese, dat de kansen op een stijging en een daling gelijk zijn, nl. ieder 0,5, en wij hopen deze veronderstelling te kunnen verwerpen, omdat wij verwachten, dat de kans op een daling groter is dan die op een stijging. Hoe groot de kans op een daling is weten wij niet; als wij dat wél wisten, viel er niets te onderzoeken. Intuïtief voelt men echter, dat hoe groter de kans op een daling is - dus hoe meer deze van 0,5 verschilt - hoe minder waarnemingen nodig zijn. Tabel 3 laat dit in getallen zien. Als de tekens in werkelijkheid in de verhouding 45-55 verdeeld zijn, moeten steekproeven van 1080 waarnemingen genomen worden om, behoudens een kans 0,05 een (bij een onbetrouwbaarheidsdrempel 0,05) significante afwijking van de verhouding 50-50 te vinden. Het getal 1080 is zo bepaald, dat men, indien de kans p_1 op een plus-teken 0,45 is, in ongeveer 95% van de gevallen een aantal

plustekens vindt, dat kleiner is dan de kritieke waarde $\chi = 512$ (indien $\mu_1 = 0,55$ zal 95% van de steekproeven een aantal plustekens groter dan $n - \chi = 568$ vertonen). Dat wil dus zeggen, dat men mag verwachten, dat wanneer een groot aantal steekproeven van de omvang 1080 worden getrokken uit een populatie, waarin de verdeling 45-55 is, ongeveer 95% van de steekproeven een significant verschil vertoont met de verdeling 50-50.

Onze aandacht richten wij in het bijzonder op de regel waaruit blijkt, dat bij $\mu_1 = 0,20$ of $0,80$ bij een onbetrouwbaarheid $0,05$ de steekproef uit tenminste 28 waarnemingen moet bestaan. De steekproefomvang moet dus ongeveer 28 zijn om een verschil tussen $\mu_0 = 0,5$ en $\mu_1 = 0,8$ aan te tonen! Omgekeerd geldt echter ook, dat wij bij toepassing van een tekentoets op de gegevens uit een steekproef bij 30 ratten, een toets verrichten, die gevoelig is voor waarden van μ_1 , die ongeveer $0,8$ of hoger zijn. Bij de keuze van de waarde van μ_1 voor de sequente binomiale toets kunnen wij hiervan gebruik maken. De keuze $\mu_0 = 0,5$ en $\mu_1 = 0,8$ sluit goed aan bij hetgeen wij tot nu toe deden. Dit betekent, dat wij het als een ernstige fout beschouwen onze hypothese, dat het bloedsuikergehalte in het algemeen daalt te verwerpen als de kans op een daling $\geq 0,8$ is. Bovendien vinden wij het een ernstige fout de veronderstelling, dat het bloedsuikergehalte gelijk blijft of stijgt te verwerpen, als $\mu \leq 0,5$ is. Als μ tussen $0,5$ en $0,8$ in ligt kan het ons niet veel schelen, welke beslissing valt.

Vooraf het feit, dat het eerst een ernstige fout wordt genoemd een overwegende daling te verwerpen wanneer μ de vrij hoge waarde van $0,8$ heeft, lijkt een tamelijk grove bewering. In de praktijk hebben wij echter vroeger door het nemen van een steekproef van dertig ratten niet veel anders gedaan dan wat in deze uitspraak wordt geconstateerd!

Nu moet worden vastgesteld welk risico wij bereid zijn te dragen, dat wij foutieve conclusies trekken. Hierbij geldt, evenals bij het kiezen van het verschil tussen de waarden van μ_0 en μ_1 , dat een grote nauwkeurigheid betaald moet worden met een grote omvang van de steekproef: Hoe kleiner risico voor fouten men wenst, hoe groter steekproef noodzakelijk is. Bij de klassieke proef plegen wij ons te houden aan het getal $0,05$ als onbetrouwbaarheidsdrempel. Het is logisch deze waarde ook te kiezen voor de maximale kans dat ten onrechte besloten wordt dat de waarschijnlijkheid van een daling $\geq 0,80$ is, terwijl zij in feite $\leq 0,5$ is. Zoals bij de klassieke toets soms een andere

waarde dan 0,05 gekozen wordt, kan dat hier eveneens geschieden. Men kan dat mede laten afhangen van de gevolgen die het trekken van een dergelijke foutieve conclusie mogelijk met zich meebrengt.

Over de maximaal toelaatbare kans de veronderstelling van een daling te verwerpen, terwijl hij in feite wel groter dan 0,8 is, spreekt men bij de klassieke methoden slechts weinig. Door de omvang van de steekproef wordt deze kans - veelal aangeduid met de letter β - ondubbelzinnig bepaald. Bij de sequente toets moeten wij dus in getal vastleggen, hoe groot de bovengrens van de kans mag zijn, dat wij beslissen, dat er geen daling is, hoewel hij er in feite wel is. In het algemeen kan men voor β dezelfde waarde kiezen als voor α , dus ook 0,05. Zonder enig bezwaar mag echter een andere waarde voor β gekozen worden. Zo kan bijvoorbeeld voor verschillende experimenten in een laboratorium een iets lagere waarde van β - bijvoorbeeld 0,01 - in overweging komen. Indien men nl. de hypothese $\mu \leq 1,5$ moet verwerpen, betekent dit in de meeste gevallen, dat het gezochte effect gevonden is. In het algemeen zal men dan de proef nog eens herhalen, voordat men er inderdaad consequenties aan verbindt. Bij deze wijze van werken wordt in de praktijk de waarde van de α , waarbij men werkt, kleiner dan 0,05. Als echter verworpen wordt, dat $\mu \geq 0,8$ is, betekent dit meestal, dat het gezochte effect niet kon worden aangetoond. Men hermaakt de proef dan veelal niet. Het risico om dit ten onrechte te doen kan daarom beter wat kleiner worden gesteld, bijvoorbeeld 0,01.

Zo komen wij op grond van deze overwegingen tot de sequente binomiale toets met de kenmerkende grootheden: $\mu_0 = 0,5$, $\mu_1 = 0,8$, $\alpha = 0,05$, $\beta = 0,01$, waarvan de grafische methode door Drs van Elteren reeds beschreven is (zie figuur 2). Tabellarisch kan deze toets volgens tabel 4 worden uitgevoerd.

Bij de grafische wijze van werken wordt na iedere waarneming een stukje aan de traplijn toegevoegd, totdat een van de beide evenwijdige grenslijnen gesneden wordt. Wanneer dat de bovenste is, wordt $\mu \leq 0,5$ verworpen bij een onbetrouwbaarheidsdrempel van 0,05, en wanneer de onderste lijn gesneden wordt, verwerpen wij, dat $\mu \geq 0,8$ is bij een onbetrouwbaarheidsdrempel van 0,01.

Wij bezien nu de werkingskromme van deze toets nog eens, en de curve van de gemiddelde steekproefgrootte (G.S.G.) bij verschillende waarden van μ . (figuur 5 en 6). Zowel voor $\beta = 0,05$ en 0,01 zijn deze getekend. Het is begrijpelijk, dat naar links de G.S.G. groter is wanneer $\beta = 0,01$ is.

Uit de G.S.G. curve kunnen wij dus voor iedere waarde van μ schatten, hoeveel waarnemingen wij gemiddeld nodig zullen hebben. Wij zien ook dat dit gemiddelde aantal in dit geval maximaal wordt bij $\mu = 0,66$ en dan resp. ongeveer 20 en 32 is. Dit zijn echter gemiddelde waarden. Wij moeten er rekening mee houden, dat het niet uitgesloten is, dat de proef eerst tot een einde komt, wanneer driemaal zoveel waarnemingen zijn verricht!

Voorbeeld I

De vraagstelling van het te verrichten experiment luidde: Heeft inspuiting van de stof alloxaan bij ratten tot gevolg, dat het bloedsuikergehalte daalt in het eerste daarop volgende halve uur? Om dit te onderzoeken werd bij een aantal vrouwelijke ratten onmiddellijk voor en een half uur na de inspuiting het bloedsuikergehalte bepaald. Als er overwegend een daling bestaat zullen de verschillen tussen de 30 minuten- en 0-waarde overwegend een negatief teken hebben. Met behulp van het gegeven sequente schema kan worden onderzocht of de kans op een negatief verschil $\cong 0,8$ dan wel $\leq 0,5$ is. Een voordeel van de toepassing van de sequente analyse is, dat wij per dag niet al teveel ratten tegelijk behoeven te behandelen. Wij verrichten de proef iedere dag bij zoveel ratten als ons op die dag goed uitkomt. De proef loopt af, zodra een beslissing genomen kan worden, die aan onze tevoren gestelde eisen voldoet.

De uitvoering werd verricht met het schema $\mu_0 = 0,5$, $\mu_1 = 0,8$, $\alpha = 0,05$, $\beta = 0,01$. De uitkomsten waren (gegroepeerd naar de op een dag verkregen gegevens): - + + - - - - - /
/- - - - + + - - / + - - - - - . Na deze 22 waarnemingen werd de bovenste grenslijn bereikt (zie figuur 2). De veronderstelling, dat de kans op een daling $\leq 0,5$ is, wordt met een onbetrouwbaarheid van 0,05 verworpen. Bij herhaling van de proef kregen wij na 13 uitkomsten hetzelfde resultaat. De uitkomsten waren toen: - + + - - - - - / - - - - - // - - - /
Op de tweede dag waren er 3 waarnemingen teveel, doordat de waarnemingen hier niet stuk voor stuk, doch in groepjes werden verricht.

Voorbeeld II

Na toediening van 3 ml arachisolie ontstaat bij ratten in het bloed een troebeling. Het interesseerde ons te weten, of deze z.g. lipemische troebeling bij diabetische ratten sterker is dan bij normale ratten. De proef werd bij paren ratten verricht. Iedere dag werd de proef bij zoveel paren ratten uitgevoerd als ons het beste uitkwam. Twee uur na de toediening van de olie werden de ratten verbloed. De troebeling van het serum van de beide partners werd vergeleken. Dikwijls was het slechts mogelijk om de troebeling op het oog te vergelijken, omdat er te weinig serum aanwezig was om het in de cuvette van de colorimeters te brengen. In sommige gevallen, waarin het verschil niet duidelijk was, en waarin er voldoende was, werd dit echter wèl gedaan. Wij kenden een plusteken toe, wanneer de troebeling bij het diabetische dier van het paar het sterkste was, en een minteken bij het tegendeel. Het ging er dus om of de kans op een plusteken $\leq 0,5$ of $\geq 0,8$ was en dit toetsten wij weer met de toets met $\mu_0 = 0,5$, $\mu_1 = 0,8$, $\alpha = 0,05$, $\beta = 0,01$.

De resultaten waren: I + + + + + - + + +
 II + + + + + + +

Na 10, resp. 7 waarnemingen bleek $\mu \leq 0,5$ verworpen te worden met een onbetrouwbaarheid van 0,05 (tabel 5). Terwijl het bij de vorige proef ging om verschillen met een bepaalde waarde, waren de exacte waarden bij de verschillen waarover wij nu spraken niet bekend. Dat dit niet nodig was betekende een grote vereenvoudiging bij de uitvoering van de proef.

Voorbeeld III

Dit voorbeeld sluit direct aan bij het vorige. Wij hebben onderzocht, of bij normale ratten de lipemie na toediening van ACTH sterker was dan na toediening van fysiologisch zout. De proef werd bij paren ratten verricht. Wij kenden een plusteken toe als de troebeling het sterkste was bij het met ACTH behandelde dier. Wij wilden dus weer weten, of de kans op een plusteken $\geq 0,8$ of $\leq 0,5$ was. De resultaten waren

- - + - - + + - + - - - // - + + - + .

De veronderstelling, dat de kans op een sterkere troebeling bij de met ACTH behandelde ratten $\geq 0,8$ is bleek na 12 waarnemingen verworpen te moeten worden. Men kan dit verifiëren door de verkregen resultaten zelf in tabel 4 in te vullen. Omdat wij de kans om deze uitspraak ten onrechte te doen $< 0,01$ gemaakt hebben, behoeven wij niet te aarzelen, om het onderzoek in deze richting af te breken.

Door waarnemingen in groepen te doen loopt men altijd de kans, dat men iets meer waarnemingen verricht dan nodig is om tot een beslissing te komen. Van de zuinigheid van de sequente methode gaat dan iets verloren. In dit geval verrichtten wij vijf waarnemingen teveel.

Voorbeeld IV

In voorbeeld I liet ik zien, hoe wij onderzochten, of toediening van alloxaan een daling van het bloedsuikergehalte ten gevolge heeft. Wij hadden echter geen rekening gehouden met het bestaan van een dagritme van het bloedsuikergehalte bij ratten. Wij hebben daarom op iets andere wijze later de proef herhaald en vergeleken of de daling bij met alloxaan ingespoten ratten groter was dan bij met fysiologisch zout ingespoten controles. De proef werd weer bij paren ratten uitgevoerd. Een plusteken werd toegekend als de daling bij de met alloxaan voorbehandelde rat groter was en een minteken bij het tegendeel. Het ging er dus weer om te onderzoeken, of de kans op een plusteken $\geq 0,8$ of $\leq 0,5$ is. De resultaten waren + + - - - - +/- - + - - // + -/+ - + - + -/- - + + - - -// . Na 12 waarnemingen bleek de veronderstelling, dat de kans groter is dan 0,8 dat een daling na alloxaan groter is dan na fysiologisch zout, met een onbetrouwbaarheid van 0,01 verworpen moest worden. De resterende twee waarnemingen van de uitkomsten op de tweede dag van het experiment konden meteen gebruikt worden voor de tweede toetsing. Na 15 waarnemingen leverde deze hetzelfde resultaat op. Ondanks de kleine onbetrouwbaarheid van de uitkomst van het eerste experiment, hebben wij dit toch herhaald, om een nog grotere zekerheid te verkrijgen, dat deze voor ons onverwachte uitkomst juist was.

Opmerking

De proeven werden iedere dag bij ongeveer zes alloxaan en zes controle-ratten verricht, waarbij de uitkomsten paarsgewijs werden bestudeerd. Vroeger zouden wij de uitkomsten van beide groepen van zes ratten in een toets van Wilcoxon of in een Student-toets hebben vergeleken. Thans wordt hier dus een sequente binomiale toets gebruikt ter vervanging van een toets voor de vergelijking van twee steekproeven. Het is jammer, dat er eigenlijk nog geen goed hanteerbare sequente χ^2 -toets of toets van Wilcoxon bestaat, die zonder heel erg veel rekenwerk kan worden uitgevoerd. Het spreekt vanzelf, dat in het besproken geval sequente toetsen, die op de een of andere manier van de

waarde van het verschil gebruik zouden maken, efficiënter zouden kunnen zijn.

Voorbeeld V

In het volgende voorbeeld wordt een bijzondere toepassing gedemonstreerd van de sequente binomiale toets. Deze toets kan namelijk ook gebruikt worden om twee kansen te vergelijken. Bij de klassieke toetsingsmethode gebruikt men hier de 2x2-tabel.

Het interesseerde ons of bij ratten, die tevoren met hydergine waren ingespoten, de kans op een daling van het bloedsuikergehalte groter was na alloxaan toediening dan na toediening van fysiologisch zout.

De proef werd steeds bij paren ratten verricht, waarvan één alloxaan en één fysiologisch zout kreeg. Op één dag werden zoveel paren achter elkaar behandeld, als ons goed uitkwam. Meestal waren er dat ongeveer zeven. Het bloedsuikergehalte werd steeds vóór, en 30 minuten na inspuiting van alloxaan of fysiologisch zout bepaald.

Als bij beide ratten uit het paar het bloedsuikergehalte stijgt (of bij beide ratten daalt) geeft een dergelijk paar geen aanwijzing of onze veronderstelling juist is, dat de kans op een daling het grootst is bij met alloxaan ingespoten ratten. Dergelijke aanwijzingen krijgen wij wel, wanneer het bloedsuikergehalte daalt bij de met alloxaan behandelde rat en stijgt bij de met fysiologisch zout behandelde rat. Een dergelijke uitkomst pleit namelijk voor onze hypothese. Daarom kennen wij hieraan een plusteken toe. Wanneer het bloedsuikergehalte stijgt na alloxaan en daalt na fysiologisch zout pleit dat tegen onze hypothese. Hiervoor plaatsen wij een minteken.

Het probleem is nu gereduceerd tot de vraag van het type: is de kans op een plusteken kleiner dan 0,5 of groter dan 0,8? Dat is het probleem, dat wij al behandeld hebben. Tabel 6 laat de uitkomsten zien, die wij bij onze proeven hebben verkregen.

De reeks tekens + + + + - - + + + + - - - + - - - - - - - -
blijkt in de toets met $\mu_0 = 0,5$, $\mu_1 = 0,8$, $\alpha = 0,05$, $\beta = 0,01$ na 19 waarnemingen tot verwerping van $\mu \geq 0,8$ (met een onbetrouwbaarheid 0,01) te voeren (men verifiëre dit met behulp van tabel 4). Ook hier geldt weer, dat één reeks waarnemingen voldoende kan worden geacht, daar $\beta = 0,01$ is gekozen.

Bij deze sequente toets voor het vergelijken van twee kansen hebben wij gebruik gemaakt van het tevoren ingevoerde schema voor een sequente binomiale toets. Het zou op deze plaats te ver voeren, de gehele wiskundige achtergrond van deze toets uiteen te zetten. Hiervoor worde naar de boeken WALD of MORONEY verwezen (zie literatuurlijst)

Het moet worden opgemerkt, dat een toets van de sterkte $p_0 = 0,5$, $p_1 = 0,8$ voor dit probleem tamelijk ongevoelig is. Men vindt hiermee niet snel een verschil tussen twee kansen. Wij zouden voor dit doel liever een toets met $p_0 = 0,5$, $p_1 = 0,75$, $\alpha = 0,05$ en $\beta = 0,01$ willen aanbevelen. In tabel 7 kunnen de kritieke waarden van deze toets gevonden worden.

Bij grafische uitvoering van de toets kan men gebruik maken van het feit dat de onderste lijn door de punten $x = 0, y = -11,23$ en $x = 10, y = 5,76$ gaat, en de bovenste door de punten $x = 0, y = 7,36$ en $x = 10, y = 24,46$. In deze meer gevoelige toets wordt na 21 waarnemingen de veronderstelling, dat $p \geq 0,75$ is met een onbetrouwbaarheid 0,01 verworpen.

Opmerking

In principe kan van de sequente binomiale toets met $p_0 = 0,5$, $p_1 = 0,75$, $\alpha = 0,05$, $\beta = 0,01$ in alle tot dusverre genoemde gevallen gebruik gemaakt worden. Het spreekt echter vanzelf dat de grotere gevoeligheid van deze toets een toename van de gemiddelde steekproefgrootte met zich meebrengt.

Voorbeeld VI (Drie beslissingsmogelijkheden)

Alle voorafgaande voorbeelden hadden gemeen, dat men bij toepassing van overeenkomstige klassieke methoden eenzijdig had kunnen toetsen. Op grond van theoretische veronderstellingen was immers steeds een vraag ontstaan, waarbij slechts een verandering in één bepaalde richting van belang was. In de praktijk is dit echter lang niet altijd het geval. Dikwijls staat men voor vraagstukken, waar, bij toepassing van de klassieke methoden, een tweezijdige toetsing noodzakelijk is. Bij de sequente analyse kunnen wij in deze gevallen toetsen met drie beslissingsmogelijkheden toepassen.

Het gaat er nu weer om, geschikte waarden voor p_0, p_1, p_0' en p_1' te kiezen. Het liefst zien wij de toets symmetrisch om $p = 0,5$. Bij de keuze van voor ons geschikte waarden kunnen wij uitgaan van door de Boer (zie literatuurlijst) gemaakte berekeningen voor deze toetsen. Bij de toets met $p_0 = 0,20$, $p_1 = 0,40$, $p_0' = 0,60$ en $p_1' = 0,80$ is de gemiddelde steekproefgrootte aan de hoge kant. Bij een onbetrouwbaarheid 0,05 bedraagt de

steekproefgrootte bij $\mu = 0,5$ gemiddeld 23 en bij $\mu = 0,3$ of $0,7$ gemiddeld 57 waarnemingen.

Bij de toets $\mu_0 = 0,10$, $\mu_1 = 0,40$, $\mu'_0 = 0,60$, $\mu'_1 = 0,90$ geldt dit in mindere mate, maar het onderscheidingsvermogen ligt voor de randgebieden niet erg gunstig, omdat de waarden $\mu_0 = 0,10$ en $\mu'_1 = 0,90$ nogal aan de extreme kant zijn.

Wanneer men gaat toetsen met $\mu_0 = 0,15$, $\mu_1 = 0,45$, $\mu'_0 = 0,55$ en $\mu'_1 = 0,85$ zijn de randgebieden iets groter en het middengebied is iets kleiner. De G.S.G. is bij deze toets iets groter dan bij de vorige. Men kan de toets echter afknotten, dat wil dus zeggen, na een bepaald aantal waarnemingen in elk geval een beslissing nemen.

Door de Boer is berekend, dat bij afknopping van de laatstgenoemde toets na 25 waarnemingen de bovengrens van de kans op een foutieve beslissing $0,07$ bedraagt. Wij kunnen veilig aannemen, dat deze kans na dertig waarnemingen kleiner dan $0,05$ zal zijn. Binnen dertig waarnemingen komt men dus met deze toets in ieder geval tot een uitspraak, waarbij de kans op een foutieve conclusie $< 0,05$ is. Ook deze toets kan weer in een tabel (tabel 8) of grafisch worden uitgevoerd (zie fig. 7 in het vorige artikel).

Opmerking

Bij grafische uitvoering van de toets met $\mu_0 = 0,15$, $\mu_1 = 0,45$, $\mu'_0 = 0,55$, $\mu'_1 = 0,85$, $\alpha = 0,05$ kan de grafiek getekend worden met behulp van de volgende gegevens. De lijnen zijn aangeduid, zoals in het artikel van Drs van Elteren (zie fig. 7).

L_0 loopt door $x = 8,35$, $y = 0$ en $x = 20$, $y = 4,61$

L_1 " " $x = 0$, $y = 2,74$ en $x = 20$, $y = 10,63$

L'_0 " " $x = 2,74$, $y = 0$ en $x = 10,63$, $y = 20$

L'_1 " " $x = 0$, $y = 8,35$ en $x = 4,61$, $y = 20$

De aanbevolen afknoppingslijn wordt weergegeven door de formule $x + y = 30$. Het deel van de afknoppingslijn dat boven het snijpunt van deze lijn en een lijn door de oorsprong evenwijdig met L_0 gelegen is, wordt bij L_1 gerekend; het gedeelte daaronder bij L_0 . De afknoppingslijn van T' wordt op overeenkomstige wijze in twee delen verdeeld.

Voorbeeld A

Op grond van bepaalde veronderstellingen is het waarschijnlijk, dat reserpine een invloed uitoefent op de bloedplaatjes. Wij hebben daarom onderzocht, of hun aantal door toediening van reserpine veranderd wordt, zonder enige veronderstelling te maken over de richting van die beïnvloeding.

De proef werd bij paren konijnen gedaan. Bij beide dieren werd bloed afgenomen om het aantal plaatjes te tellen; daarna werd de ene ingespoten met reserpine en de andere diende als controle. Vier uur later werd bij beide dieren wéér het aantal bloedplaatjes geteld.

Wij kenden een plusteken toe als de stijging van het aantal bloedplaatjes bij de met reserpine behandelde konijnen groter was dan bij de controles, en een minteken als het tegendeel het geval was. Het ging er dus om, te weten, in welk interval de kans op een plusteken gelegen was. De volgende uitkomsten werden verkregen: + - + + + + + + + + . Na proeven bij 12 paren konijnen kwam er een uitspraak. Het bleek dat de veronderstelling, dat de kans op een plusteken kleiner dan 0,55 was verworpen moest worden met een onbetrouwbaarheid van 0,05. De gevolgde weg is getekend in figuur 7 in het artikel van Drs van Elteren.

Voorbeeld B

Een uitspraak werd gevraagd, welke van twee stoffen, A dan wel B, het geschiktst was om als grondstof in een cosmetisch preparaat te verwerken. Het gevaar bestond nl., dat deze stoffen bij applicatie per ongeluk in het oog zouden komen. Welke stof zou in dat geval het minst prikkelend werken?

Om dat te onderzoeken hebben wij beide stoffen verdund tot de hoogste concentratie, die praktisch in het oog zou kunnen komen. Vervolgens hebben wij bij een aantal konijnen in het ene oog een druppel van de oplossing van stof A en in het andere een druppel van die van stof B gebracht. De volgende dag hebben wij onderzocht, welke van beide ogen het sterkst gereageerd had. Het bleek gelukkig bij geen van beide ogen erg veel te zijn, maar het was met enige goede wil toch wel mogelijk steeds een uitspraak te doen, welke van beide het meest geïrriteerd was. Het zou hier ondoenlijk zijn geweest enige reactie in maat of getal uit te drukken. Op deze wijze konden echter subjectieve indrukken gekwantificeerd worden.

Als A de meeste reactie veroorzaakt had, kenden wij een plusteken toe, en als dit bij B het geval was een minteken. De proef werd met nieuwe konijnen voortgezet totdat een uitspraak kon worden gedaan. Na de resultaten + + - - + - - - + + bleek dit het geval te zijn. De veronderstelling, dat de kans op een plusteken $\leq 0,15$ of $\geq 0,85$ is bleek te moeten worden verworpen met een onbetrouwbaarheid 0,05. In tabel 9 is aangegeven, hoe de tabel met kritieke waarden bij de uitkomsten van het laatste voorbeeld kan worden gebruikt. Allereerst wordt verworpen, dat $p \geq 0,85$ is. Bij voortzetting van de toetsing blijkt dan, dat dit ook geldt voor $p \leq 0,15$.

Literatuur:

De sequente analyse is door een groep van Amerikaanse statistici onder leiding van A. WALD in de tweede wereldoorlog ontwikkeld. De theorie van de methode vindt men uitvoerig in het boek:

A. WALD: Sequential Analysis; John Wiley & Sons, Chapman and Hall. New York, London (1947).

Een korte beschrijving van de sequente binomiale toets vindt men in het eenvoudige boekje van:

M.J. MORONEY: Facts from figures (Pelican book) p. 192-215.

Voor de toepassing is een uitvoerige verzameling van recepten en tabellen samengesteld in:

STATISTICAL RESEARCH GROUP, Columbia University, Sequential analysis of statistical data, applications (1945).

De in par. 5 beschreven sequente toets met 3 beslissingsmogelijkheden vindt men in:

J. DE BOER: Sequential test with three possible decisions for testing an unknown probability, Appl. Scientific Research B (1953) p. 249-259.

Voor een sequente toets voor de vergelijking van twee kansen zie men:

C. VAN EEDEN: Sequential test with three possible decisions for the comparison of two unknown probabilities. Revue de l'institut international de statistique 1956 (moet nog verschijnen).

Lijst van formules

Sequente binomiale toets met twee beslissingsmogelijkheden.

Bepalende grootheden: p_0, p_1, α en β .

Grenslijnen: (x = aantal witte, y = aantal rode ballen)

$$(1) L_0 : y = h_0 + s x$$

$$(2) L_1 : y = h_1 + s x$$

met

$$(3) s = \frac{\log \frac{1-p_0}{1-p_1}}{\log \frac{p_1}{p_0}},$$

en

$$(4) h_0 = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{p_1}{p_0}},$$

en

$$(5) h_1 = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{p_1}{p_0}}$$

Bijvoorbeeld: $\alpha = 0,05$; $\beta = 0,01$; $p_0 = 0,5$; $p_1 = 0,8$ geeft

$$(6) \left\{ \begin{array}{l} s = \frac{\log 2.5}{\log 1.6} = 1.94954 \approx 1.95 \\ h_0 = \frac{\log 1/95}{\log 1.6} = -9.68902 \approx -9.69 \\ h_1 = \frac{\log 19.8}{\log 1.6} = +6.35247 \approx +6.35 \end{array} \right. \quad \text{Vgl. fig. 2}$$

L_0 gaat door de punten: $(0; -9,69)$ en $(4,97; 0)$

L_1 " " " " : $(0; +6,35)$ en $(10; 25,85)$

Bij $\alpha = 0,05$; $\beta = 0,05$; $p_0 = 0,5$; $p_1 = 0,8$:

$$s = 1,95; h_0 = -6,26; h_1 = +6,26$$

$$(7) \left\{ \begin{array}{l} L_0 \text{ gaat door de punten: } (0; -6,26) \text{ en } (3,21; 0) \\ L_1 \text{ " " " " : } (0; +6,26) \text{ en } (10; 25,76) \end{array} \right. \quad \text{Vgl. fig. 2}$$

Grenslijnen als y = aantal rode, n = totaal aantal getrokken ballen:

$$(8) L_0 : y = \frac{h_0}{1+s} + \frac{s}{1+s} n$$

$$(9) L_1 : y = \frac{h_1}{1+s} + \frac{s}{1+s} n$$

Vgl. fig. 3

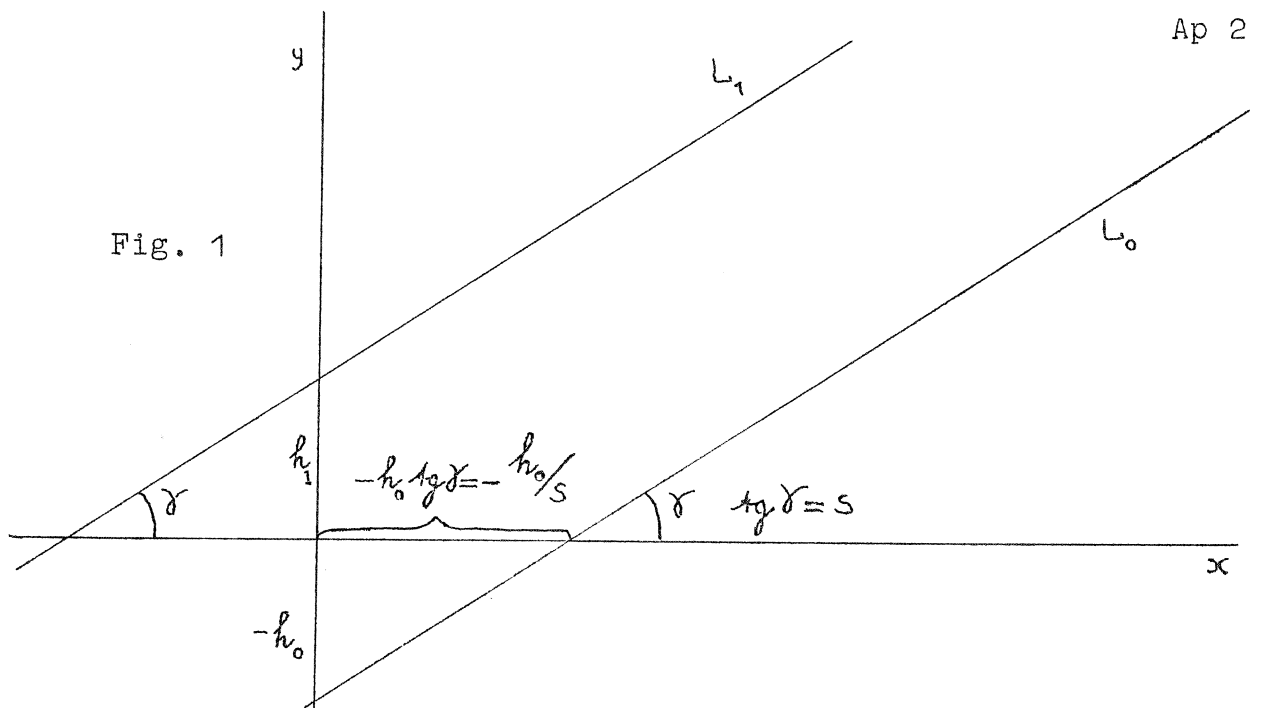


Fig. 1: Sequente binomiale toets

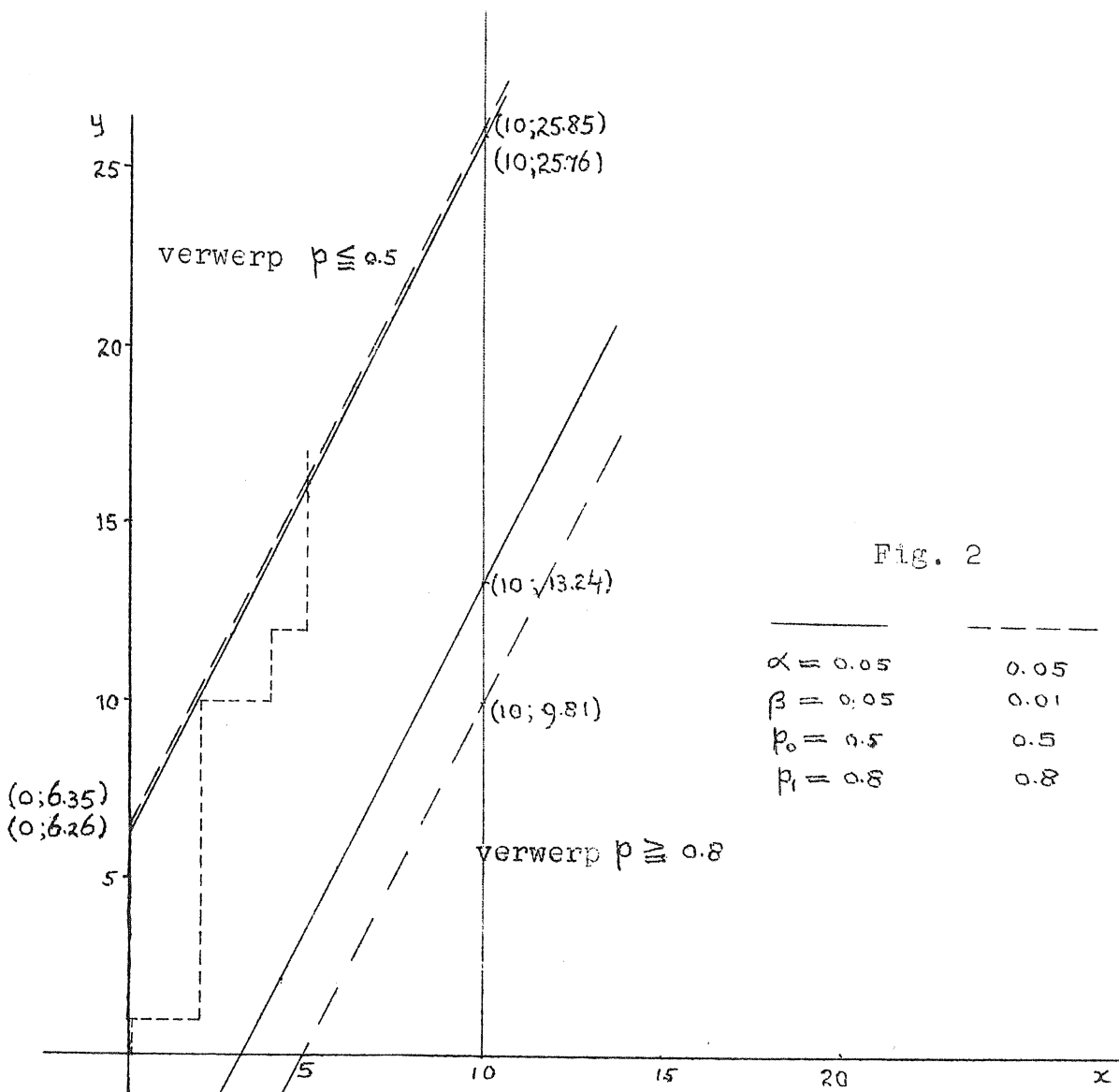


Fig. 2

Fig. 2: Sequente binomiale toets met ingetekende waarnemingsreeks

(0; 6.26)

(0; 9.69)

- + + - - - - - - - - + + - - + - - - -

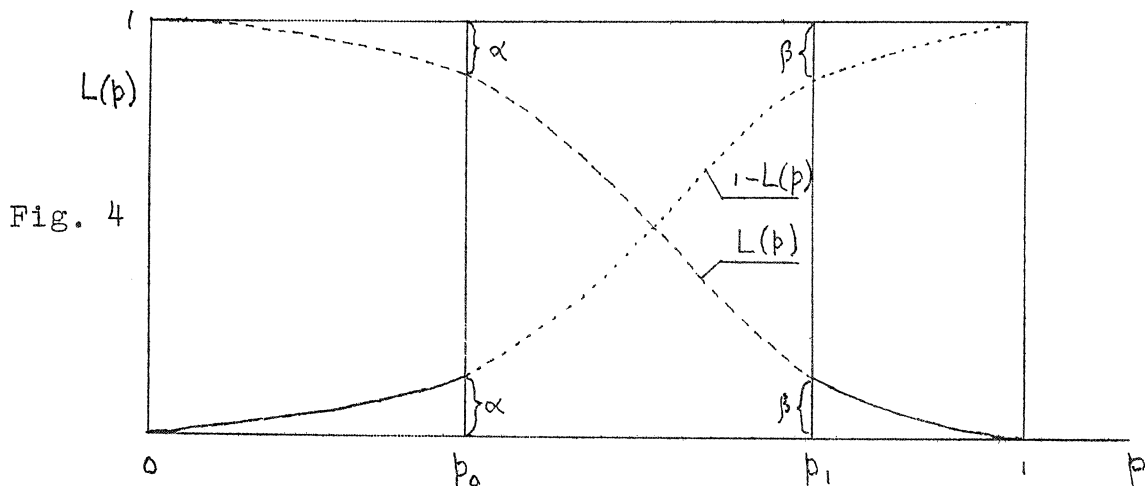


Fig. 4

Fig. 4. Schets van de werkingskromme (---) en van de kans op een foutieve conclusie (—) bij een sequente binomiale toets.

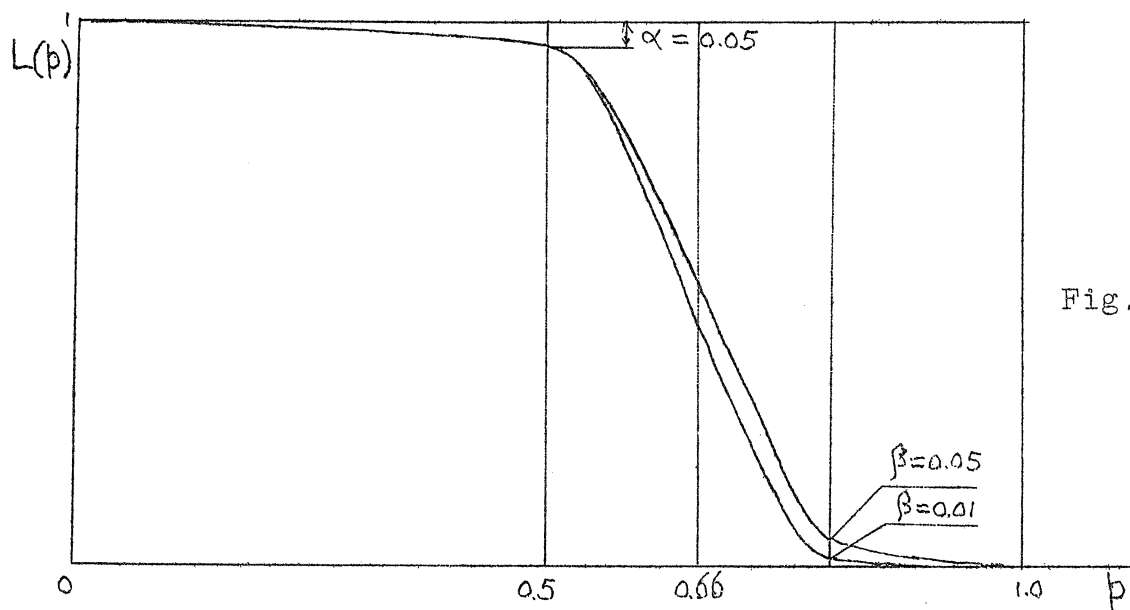


Fig. 5

Fig. 5. Werkingskrommen van twee bepaalde sequente binomiale toetsen (zie tabel 1) $p_0 = 0,5$, $p_1 = 0,8$, $\alpha = 0,05$, $\beta = 0,01$ en $0,05$.

Tabel 1

Bepaling van de werkingskrommen

| p | | L(p) | |
|---------|-----------------|----------------|----------------|
| | | $\beta = 0,01$ | $\beta = 0,05$ |
| 0 | 1 | 1 | 1 |
| p_0 | $1 - \alpha$ | 0,95 | 0,95 |
| $s/s+1$ | $r_1/r_1 - r_0$ | 0,396 | 0,5 |
| p_1 | β | 0,01 | 0,05 |
| 1 | 0 | 0 | 0 |

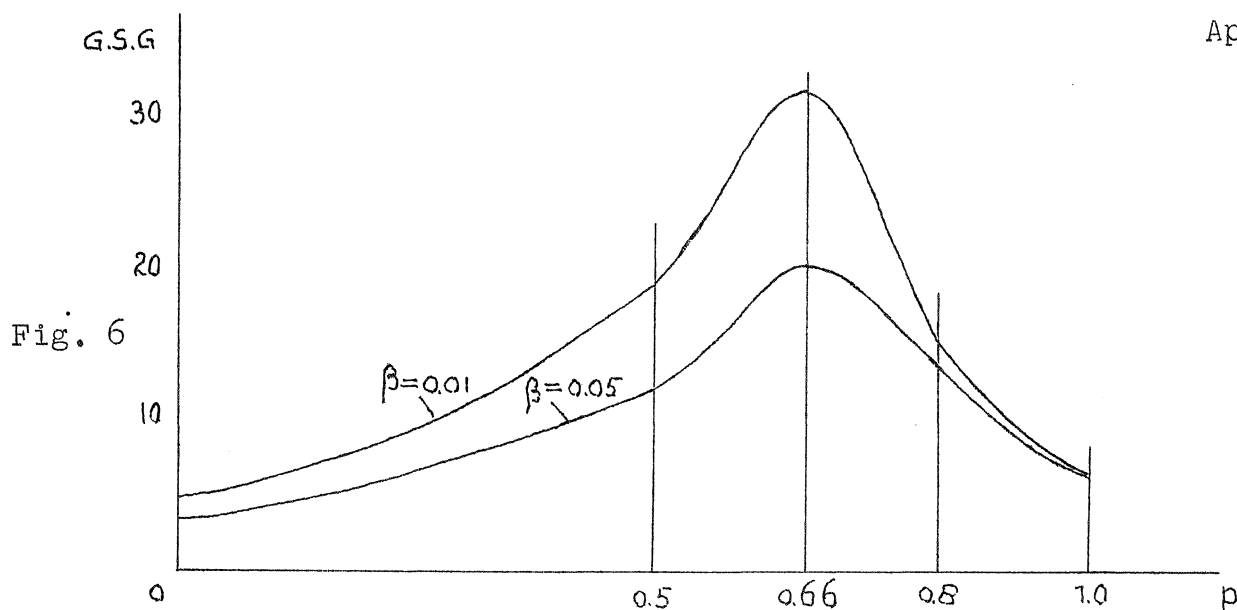


Fig. 6: G.S.G. voor $p_0 = 0,5$; $p_1 = 0,8$; $\alpha = 0,05$; $\beta = 0,01$ én $0,05$

Tabel 2

| p | G.S.G. | | |
|---------|---|----------------|-------|
| | $\beta = 0,01$ | $\beta = 0,05$ | |
| 0 | $-\frac{h_0}{s}$ | 4,97 | 3,21 |
| p_0 | $\frac{(1-\alpha)h_0 + \alpha h_1}{(1+s)p_0 - s}$ | 18,72 | 11,88 |
| $s/1+s$ | $-\frac{h_0 - h_1}{s}$ | 31,57 | 20,13 |
| p_1 | $\frac{\beta h_0 + (1-\beta)h_1}{(1+s)p_1 - s}$ | 15,10 | 13,75 |
| 1 | h_1 | 6,35 | 6,26 |

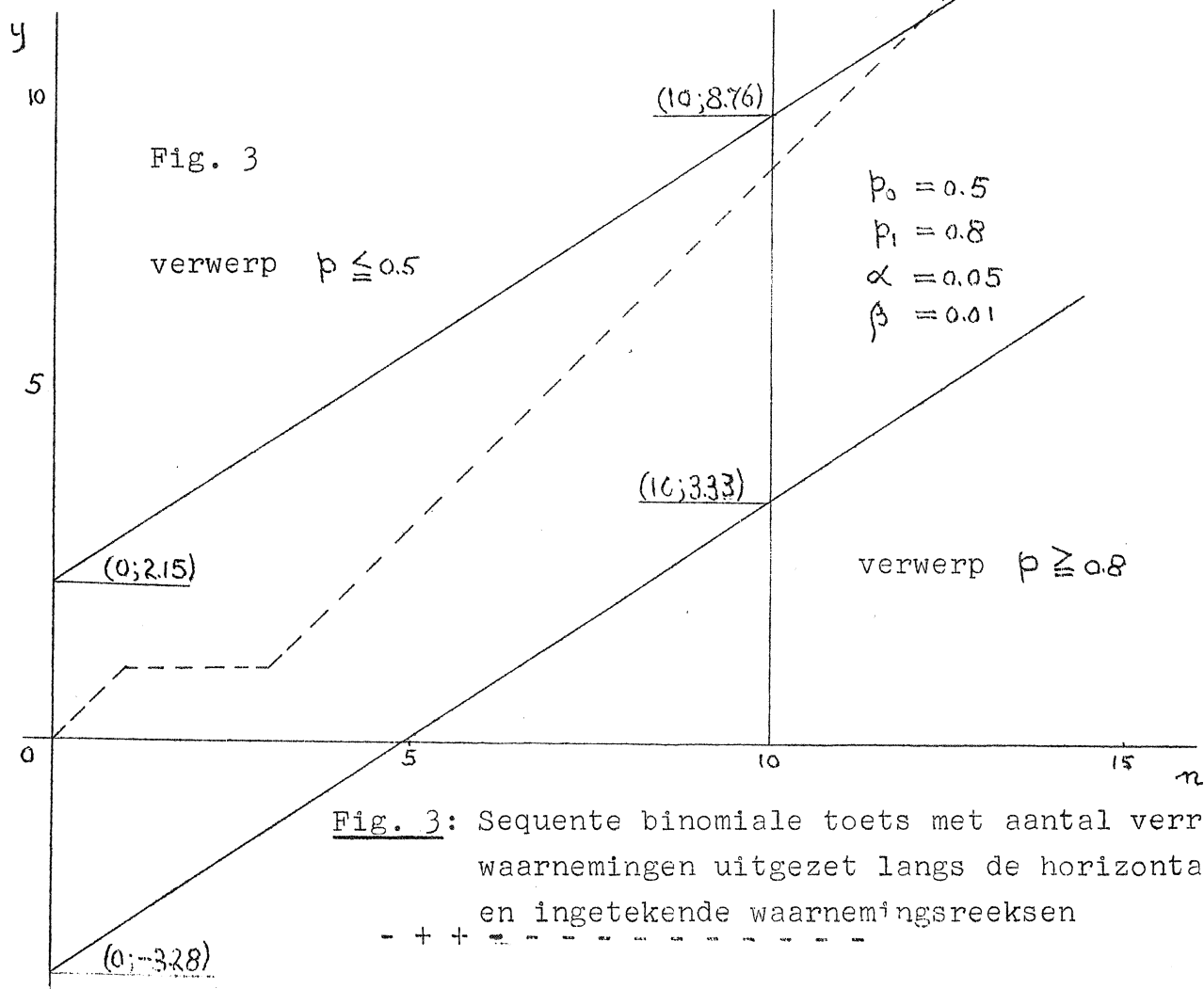


Fig. 3: Sequentie binomiale toets met aantal verrichte waarnemingen uitgezet langs de horizontale as en ingetekende waarnemingsreeksen

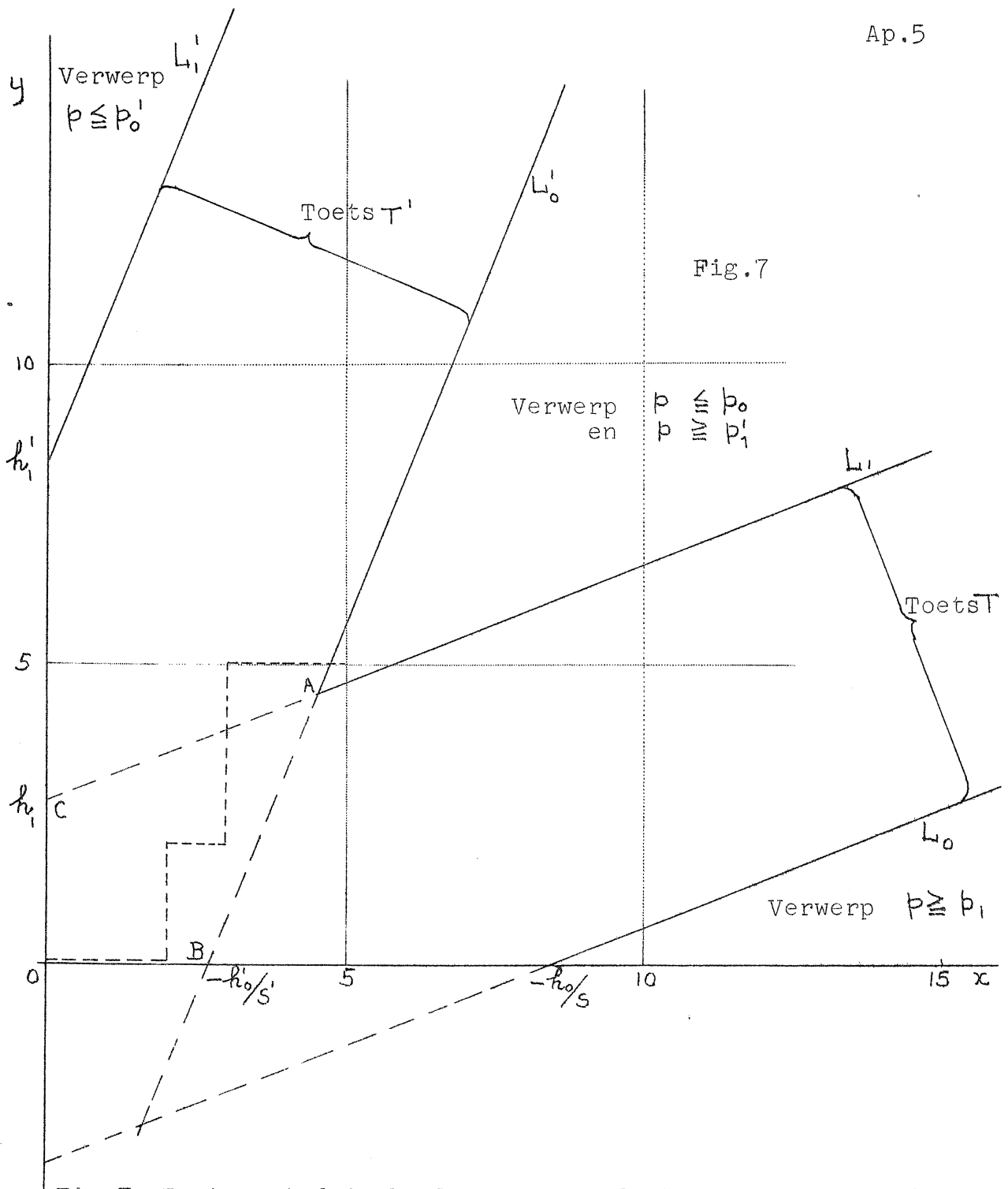


Fig.7. Toets met drie beslissingsmogelijkheden met $p_0=0.15; p_1=0.45$
 $p_0'=0.55; p_1'=0.8$
 $\alpha=\beta=0.05; \alpha'=\beta'=0.2$
 en ingetekende waarnemingsreeks + + - - + - - - + + .

Tabel 3

Gedrag van de éézijdige tekentoets indien $p=p_1$ (onbetrouwbaarheidsdrempel is 0,05)

| p_1 | noodzakelijke omvang van de steekproef n | kritieke waarde r |
|--------------|--|---------------------|
| 0.45 of 0.55 | 1080 | 512 |
| 0.40 0.60 | 297 | 119 |
| 0.35 0.65 | 118 | 40 |
| 0.30 0.70 | 67 | 26 |
| 0.25 0.75 | 42 | 15 |
| 0.20 0.80 | 28 | 9 |
| 0.15 0.85 | 18 | 8 |
| 0.10 0.90 | 13 | 3 |
| 0.05 0.95 | 11 | 2 |

Tabel 4

Kritieke waarden voor de sequente binomiale toets met
 $p_0 = 0,5$; $p_1 = 0,8$ en $\alpha = 0,05$

| n | $\beta = 0,01$ | | | $\beta = 0,05$ | | |
|----|----------------|-------|-------|----------------|-------|-------|
| | a_n | d_n | b_n | a_n | d_n | b_n |
| 1 | - | | - | - | | - |
| 2 | - | | - | - | | - |
| 3 | - | | - | - | | - |
| 4 | - | | - | 0 | | - |
| 5 | 0 | | - | 1 | | - |
| 6 | 0 | | - | 1 | | - |
| 7 | 1 | | 7 | 2 | | 7 |
| 8 | 2 | | 8 | 3 | | 8 |
| 9 | 2 | | 9 | 3 | | 9 |
| 10 | 3 | | 9 | 4 | | 9 |
| 11 | 3 | | 10 | 5 | | 10 |
| 12 | 4 | | 11 | 5 | | 11 |
| 13 | 5 | | 11 | 6 | | 11 |
| 14 | 5 | | 12 | 7 | | 12 |
| 15 | 6 | | 13 | 7 | | 13 |
| 16 | 7 | | 13 | 8 | | 13 |
| 17 | 7 | | 14 | 9 | | 14 |
| 18 | 8 | | 15 | 9 | | 15 |
| 19 | 9 | | 15 | 10 | | 15 |
| 20 | 9 | | 16 | 11 | | 16 |
| 21 | 10 | | 17 | 11 | | 17 |
| 22 | 11 | | 17 | 12 | | 17 |
| 23 | 12 | | 18 | 13 | | 18 |
| 24 | 12 | | 19 | 13 | | 18 |
| 25 | 13 | | 19 | 14 | | 19 |
| 26 | 13 | | 20 | 15 | | 20 |
| 27 | 14 | | 20 | 15 | | 20 |
| 28 | 15 | | 21 | 16 | | 21 |
| 29 | 15 | | 22 | 17 | | 22 |
| 30 | 16 | | 22 | 17 | | 22 |
| 31 | 17 | | 23 | 18 | | 23 |
| 32 | 17 | | 24 | 19 | | 24 |
| 33 | 18 | | 24 | 19 | | 24 |
| 34 | 19 | | 25 | 20 | | 25 |
| 35 | 19 | | 26 | 21 | | 26 |
| 36 | 20 | | 26 | 21 | | 26 |
| 37 | 21 | | 27 | 22 | | 27 |
| 38 | 21 | | 28 | 22 | | 28 |
| 39 | 22 | | 28 | 23 | | 28 |
| 40 | 23 | | 29 | 24 | | 29 |
| 41 | 23 | | 30 | 24 | | 30 |
| 42 | 24 | | 30 | 25 | | 30 |
| 43 | 25 | | 31 | 26 | | 31 |
| 44 | 25 | | 32 | 26 | | 32 |
| 45 | 26 | | 32 | 27 | | 32 |

n = aantal waarnemingen

d_n = aantal plustekens na n waarnemingen als $p = p[+]$

a_n = kritieke waarden voor verwerping $p \geq 0,8$

b_n = kritieke waarden voor verwerping $p \leq 0,5$

Tabel 5

Toepassing van tabel 4 voor $\beta = 0,01$ voor twee waarnemingsreeksen d_n en d'_n .

| n | a_n | d_n | d'_n | b_n |
|-----|-------|----------------------|----------------------|-------|
| 1 | - | 1 | 1 | - |
| 2 | - | 2 | 2 | - |
| 3 | - | 3 | 3 | - |
| 4 | - | 4 | 4 | - |
| 5 | 0 | 5 | 5 | - |
| 6 | 0 | 6 | 6 | - |
| 7 | 1 | 6 | <u>7</u> - - - - - 7 | 7 |
| 8 | 2 | 7 | | 8 |
| 9 | 2 | 8 | | 9 |
| 10 | 3 | <u>9</u> - - - - - 9 | | 9 |
| 11 | 3 | | | 10 |
| 12 | 4 | | | 11 |

Reeks d_n : + + + + + - + + + /

d'_n : + + + + + + /

Tabel 6

Gedrag van het bloedsuikergehalte bij paren ratten

| | | |
|----------|---------------|---------------|
| alloxaan | d d d d d d d | s s d s d s |
| fys.zout | d s d s s d s | d d d s s s |
| uitkomst | + + + + | - - + |
| alloxaan | d d d d d | s d s d s d s |
| fys.zout | d d d s s | s s d d s d d |
| uitkomst | + + | + - - |
| alloxaan | s s d s s s s | s s s s s s d |
| fys.zout | d s s s d d d | d d d s s d s |
| uitkomst | - + - - - | - - - - + |

Stijging (s) of daling (d) bij ratten, die na toediening van hydergine met alloxaan of fysiologisch zout zijn ingespoten.

Tabel 7

Kritieke waarden voor de sequente binomiale toets met

$$p_0 = 0,50; p_1 = 0,75; \alpha = 0,05; \beta = 0,01$$

| n | a_n | d_n | b_n |
|-----|-------|-------|-------|
| 1 | - | | - |
| 2 | - | | - |
| 3 | - | | - |
| 4 | - | | - |
| 5 | - | | - |
| 6 | - | | - |
| 7 | 0 | | - |
| 8 | 0 | | 8 |
| 9 | 1 | | 9 |
| 10 | 2 | | 10 |
| 11 | 2 | | 10 |
| 12 | 3 | | 11 |
| 13 | 4 | | 11 |
| 14 | 4 | | 12 |
| 15 | 5 | | 13 |
| 16 | 5 | | 13 |
| 17 | 6 | | 14 |
| 18 | 7 | | 15 |
| 19 | 7 | | 15 |
| 20 | 8 | | 16 |
| 21 | 9 | | 16 |
| 22 | 9 | | 17 |
| 23 | 10 | | 18 |
| 24 | 10 | | 18 |
| 25 | 11 | | 19 |
| 26 | 12 | | 20 |
| 27 | 12 | | 20 |
| 28 | 13 | | 21 |
| 29 | 14 | | 22 |
| 30 | 14 | | 22 |

n = aantal waarnemingen

d_n = aantal plustekens na n waarnemingen als $p = p[+]$

a_n = kritieke waarden voor verwerping $p \geq 0,75$

b_n = kritieke waarden voor verwerping $p \leq 0,50$.

Tabel 8

Kritieke waarden voor de sequente binomiale toets met drie beslissingsmogelijkheden, waarbij $p_0 = 0,15$; $p_1 = 0,45$; $p_2 = 0,55$; $p_3 = 0,85$ en $\alpha = 0,05$.

| n | a_n | b_n | d_n | a'_n | b'_n |
|-----|-------|-------|-------|--------|--------|
| 1 | - | - | | - | - |
| 2 | - | - | | - | - |
| 3 | - | 3 | | 0 | - |
| 4 | - | 4 | | 0 | - |
| 5 | - | 4 | | 1 | - |
| 6 | - | 4 | | 2 | - |
| 7 | - | 4 | | 3 | - |
| 8 | - | 5 | | 3 | - |
| 9 | 0 | 5 | | 4 | 9 |
| 10 | 0 | 5 | | 5 | 10 |
| 11 | 0 | 6 | | 5 | 11 |
| 12 | 1 | 6 | | 6 | 11 |
| 13 | 1 | 6 | | 7 | 12 |
| 14 | 1 | 6 | | 8 | 13 |
| 15 | 1 | 7 | | 8 | 14 |
| 16 | 2 | 7 | | 9 | 14 |
| 17 | 2 | 7 | | 10 | 15 |
| 18 | 2 | 8 | | 10 | 16 |
| 19 | 3 | 8 | | 11 | 16 |
| 20 | 3 | 8 | | 12 | 17 |
| 21 | 3 | 8 | | 13 | 18 |
| 22 | 3 | 9 | | 13 | 19 |
| 23 | 4 | 9 | | 14 | 19 |
| 24 | 4 | 9 | | 15 | 20 |
| 25 | 4 | 10 | | 15 | 21 |
| 26 | 5 | 10 | | 16 | 21 |
| 27 | 5 | 10 | | 17 | 22 |
| 28 | 5 | 10 | | 18 | 23 |
| 29 | 5 | 11 | | 18 | 24 |
| 30 | 6 | 11 | | 19 | 24 |

n = aantal waarnemingen

d_n = aantal plustekens na n waarnemingen als $p = p[+]$

a_n = kritieke waarden voor verwerping $p \geq 0,45$

b_n = " " " " $p \leq 0,15$
 a'_n = " " " " $p \geq 0,85$
 b'_n = " " " " $p \leq 0,55$

} na n waarnemingen

Tabel 9

Toepassing van tabel 8 voor een waarnemingsreeks

| n | a_n | b_n | d_n | a'_n | b'_n |
|-----|-------|-------|-------|--------|--------|
| 1 | - | - | 1 | - | - |
| 2 | - | - | 2 | - | - |
| 3 | - | 3 | 2 | 0 | - |
| 4 | - | 4 | 2 | 0 | - |
| 5 | - | 4 | 3 | 1 | - |
| 6 | - | 4 | 3 | 2 | - |
| 7 | - | 4 | 3 | 3 | - |
| 8 | - | 3 5 | | 3 | - |
| 9 | 0 | 4 5 | | 4 | 9 |
| 10 | 0 | 5 5 | | 5 | 10 |
| 11 | 0 | 6 | | 5 | 11 |
| 12 | 1 | 6 | | 6 | 11 |

Waarnemingsreeks d_n : + + - - + - - - + + .

Na 7 waarnemingen wordt $p \cong 0,85$ verworpen en vervolgens na 10 waarnemingen ook $p \leq 0,15$.

Lijst van errata behorende bij Rapport S 199 (V 14)

"De sequente binomiale toets"

I. De Statistische techniek

| | <u>staat:</u> | <u>moet zijn:</u> |
|-------------------------------------|-------------------|-------------------|
| <u>pag. 2.</u> regel 11 van boven | en zo voorts | enzovoorts |
| <u>pag. 4.</u> regel 16 van onderen | Werkingskromme | 3. Werkingskromme |
| regel 3 van onderen | aan 1, | aan 1. |
| <u>pag. 5.</u> regel 3 van boven | In tabel 5 | In tabel 1 |
| regel 14 van boven | charaderistic | characteristic |
| <u>pag. 6.</u> regel 12 van boven | steekproefgrootte | steekproefgrootte |
| regel 4 van onderen | $p \leq p_1$ | $p \geq p_1$ |
| <u>pag. 8.</u> regel 2 van onderen | eerste | eerst |

II. Enkele toepassingen op farmacologisch terrein

| | <u>staat:</u> | <u>moet zijn:</u> |
|---|-------------------------------------|-------------------------------------|
| <u>pag. 1.</u> regel 17 van boven | een bepaald
geneesmiddel | een bepaalde stof |
| <u>pag. 2.</u> regel 10 van onderen | trokken | trekken |
| <u>pag. 5.</u> regel 10 van boven | colorimeters | colorimeter |
| <u>pag. 6.</u> regel 13 van boven | physiologisch | fysiologisch |
| <u>pag. 7.</u> regel 6 van onderen | De reeks tekens
+ + - + - - enz. | De reeks tekens
+ + + + - - enz. |
| pag.10. regel 14, 15 en 16
van boven | De gevolgde
weg | vervalt |

Appendix

| | | |
|----------------------|------------|------------|
| pag. 6. regel n = 23 | $a_n = 12$ | $a_n = 11$ |
|----------------------|------------|------------|