MATHEMATISCH CENTRUM
AMSTERDAM

Augustus 1951

Statistische Afdeling
Rapport S 66 (P12),
door Dr J.Hemelrijk Jr.

De behandeling van onbesliste experimenten bij de tekentoets.

§ 1. Inleiding.

De tekentoets, waarover o.a. door W.J.DIXON en A.M.MOOD [1] in 1946 een publicatie geschreven is, kan, in de vorm van een model, het eenvoudigst beschreven worden als een toets voor de zuiverheid van een munt.
Op grond van de resultaten van een aantal onafhankelijke worpen met de munt wordt de hypothese getoetst, dat de kans op "kruis" bij iedere worp gelijk is aan de kans op "munt". Deze toets, die berust op de binomiale verdeling van het aantal maal "kruis", dat bij een serie worpen verkregen wordt, is van een zeer eenvoudig karakter, maar bezit slechts een gering onderscheidingsvermogen; zij wordt vaak gebruikt ter oriëntering bij uitgebreid materiaal. Van tijd tot tijd ontmoet men echter problemen, waarbij men, op grond van de omstandigheden of de inrichting van een experiment, geheel op deze toets is aangewezen. Voor dergelijke gevallen is een nadere beschouwing van de toets van belang.

Een aan de practijk ontleend voorbeeld van een dergelijke proef vindt men b.v. bij de keuring van wasmiddelen. Om twee wasmiddelen A en B te vergelijken verricht men een serie experimenten, ieder bestaande uit de behandeling van twee ongeveer even vuile partijen wasgoed met middel A resp. B. Het resultaat van ieder experiment wordt visueel beoordeeld en kan 3 uitkomsten geven, te weten:

F, voor zoverre het schoonwassing (vuil- en vlekkenverwijdering) betreft,

[A]: wasmiddel A heeft het beste resultaat gegeven;
[B]: " B " " " " ;
[C]: er is geen verschil geconstateerd.

Indien men bij deze proef geen kans ziet tot een meer gedifferentieerde beoordeling te komen, is men bij de statistische verwerking wel aangewezen op de tekentoets. Daarbij doet zich dan de vraag voor, wat men met de experimenten, die de met [C] aangeduide uitkomst gegeven hebben, zal doen. Dit probleem treedt ook op, wanneer men over paren quantitatieve waarnemingen (x_i, y_i) $(i=1,\dots,n)$ beschikt, waarvan men de verschillen $x_i - y_i$ beschouwt, om te toetsen, of voor deze verschillen de kans om negatief te zijn gelijk is aan de kans om positief te zijn. In dat geval treden soms gelijke waarnemingen ($x_i=y_i$) op, zodat het verschil gelijk aan 0 wordt. Deze waarnemingsparen komen overeen met de experimenten [C] in bovenstaande proefopzet. Wij zullen verder de terminologie van het genoemde voorbeeld blijven gebruiken en de uitkomsten aangeven als [A], [B] en [C], terwijl we deze laatste ook "onbesliste experimenten" zullen noemen. Verder zullen we steeds het aantal experimenten met uitkomst [A], [B] resp. [C] aangeven door n_1, n_2 resp. n_3, zodat

(1) $$n=n_1+n_2+n_3$$

het aantal bij de proef verrichte experimenten voorstelt.

Er zijn twee voor de hand liggende antwoorden te geven op de vraag, wat men met de onbesliste experimenten zal doen: men kan deze experimenten bij de statistische verwerking weglaten, of men kan ze voor de helft bij de uitkomsten [A] en voor de andere helft bij [B] tellen. De tweede methode werd in 1946, zonder nader commentaar, voorgeschreven door DIXON en MOOD [1], terwijl DIXON en MASSEY [2], pag. 248, in 1951, eveneens zonder commentaar, de eerste methode voorschrijven. De eerste methode bezit, zoals verderop wordt aangetoond, een groter onderscheidingsvermogen dan de tweede en is dus in dit opzicht te prefereren, te meer daar het onderscheidingsvermogen van de tekentoets toch al niet groot is. Er is echter ook aan deze methode een bezwaar verbonden, gelegen in het feit dat de uitslag van de proef bij deze handelwijze slechts van n_1 en n_2 afhangt en niet meer van n_3. Twee proeven b.v., waarvan de ene als uitkomst geeft:

$n_1=8$ $\qquad n_2=1$ $\qquad n_3=0$ $\qquad (n=9)$

en de tweede

$n_1=8$ $\qquad n_2=1$ $\qquad n_3=1000$ $\qquad (n=1009)$,

leiden beide tot hetzelfde resultaat; bij de wasmiddelenproef zou (bij een onbetrouwbaarheidsdrempel 0,05) in beide gevallen de conclusie zijn, dat middel A beter is dan middel B en in het tweede geval is dit resultaat niet erg acceptabel, daar men bij verreweg de meeste der experimenten geen verschil heeft gevonden.

Naast een beschrijving van de beide genoemde methoden en een vergelijking van hun merites, zullen wij daarom in dit artikel een derde methode aangeven, die bestaat uit een combinatie van deze twee en die de voordelen van beide tot op zekere hoogte verenigt en de nadelen vermijdt.

Bij de mathematische beschouwingen worden de begindelen van de toetsingstheorie bekend ondersteld (zie b.v. [3] en [4]). Verder wordt de binomiale verdeling bekend ondersteld, evenals het gemiddelde en de spreiding daarvan en de asymptotische normaliteit van deze verdeling voor $n \to \infty$.

De bewijzen van de stellingen zijn voor het merendeel in een appendix samengevat.

§ 2. Notaties en definities.

Een stochastische grootheid geven wij aan door een onderstreepte letter, terwijl dezelfde letter, niet onderstreept, vaak gebruikt wordt voor een door een dergelijke grootheid aangenomen waarde.

De voorwaardelijke waarschijnlijkheid van een gebeurtenis G , onder de voorwaarde V en (eventueel) onder hypothese H, geven wij aan door

(2) $$P\left[G \mid V ; H \right].$$

Ook een voorwaardelijke verwachting, d.i. de mathematische verwachting (het gemiddelde) van een stochastische grootheid met een voorwaardelijke waarschijnlijkheidsverdeling, geven wij op deze wijze aan: zo stelt b.v.

(3) $$\mathcal{E}\left\{ g(\underline{x}) \mid \underline{y}=y \,;\, H \right\} \quad {}^{1)}$$

de voorwaardelijke verwachting van de functie $g(\underline{x})$ voor, onder voorwaarde $\underline{y}= y$ en hypothese H. Deze uitdrukking heeft slechts betekenis, indien de stochastische grootheden $\underline{x}$ en $\underline{y}$, onder aanname van H, een simultane waarschijnlijkheidsverdeling bezitten. Voor definitie en een aantal eigenschappen van voorwaardelijke waarschijnlijkheidsverdelingen vergelijke men b.v. CRAMER [5] p. 157-159 en 267-270.

Wij gebruiken steeds de letter α voor de onbetrouwbaarheidsdrempel (Eng.: "level of significance") van een toets; α is de bovengrens, die men oplegt aan de kans op verwerping van de getoetste hypothese, indien deze juist is.

De genoemde kans zelf noemt men de onbetrouwbaarheid β_0 ("real level of significance"), die men zo dicht mogelijk onder α neemt. De waarde α kan ten gevolge van het optreden van discrete waarschijnlijkheidsverdelingen in het algemeen niet bereikt worden.

Het onderscheidingsvermogen β van een toets is de kans op

---

1) Voorwaarden van de vorm $\underline{y}=y$ worden in (2) en (3) vaak kort aangegeven door y alleen:

$$\mathcal{E}\left\{ g(\underline{x}) \mid y ; H \right\}.$$

verwerping van de getoetste hypothese; β is afhankelijk van de juiste hypothese H (notatie: $\beta(H)$). Indien H de getoetste hypothese is, is $\beta=\beta_o$.

§ 3. Het model.

Aan iedere waarschijnlijkheidstheoretische beschouwing ligt een mathematisch model ten grondslag. Wij zullen hier het eenvoudigst mogelijke model gebruiken, door te onderstellen dat alle experimenten onafhankelijk zijn en dat bij ieder daarvan de kansen op de uitkomsten [A], [B] resp. [C] gegeven worden door:

(4) $$p_1 = P[A]; \quad p_2 = P[B] \quad \text{resp.} \quad p_3 = P[C].$$

In § 10. bespreken wij een verfijning van dit model.

§ 4. Methode I: Weglaten der onbesliste experimenten.

De te toetsen hypothese H_o is:

(5) $$H_o : p_1 = p_2.$$

De waarde van p_3 wordt hierbij buiten beschouwing gelaten. De toets berust nu op het feit, dat bij gegeven waarde van $\underline{n}_3$ (dus bij een gegeven aantal onbesliste experimenten) ~~één~~ onder aanname van H_o geldt:

(6) $$P[\underline{n}_1=n_1 \mid n_3; H_o] = 2^{n_3-n}\binom{n-n_3}{n_1}.$$

Dit betekent dus, dat $\underline{n}_1$, indien H_o juist is en indien er n_3 onbesliste experimenten zijn, een binomiale waarschijnlijkheidsverdeling bezit met gemiddelde en spreidingskwadraat:

(7) $$\begin{cases} \mathcal{E}(\underline{n}_1 \mid n_3; H_o) = \frac{1}{2}(n-n_3); \\ \sigma^2(\underline{n}_1 \mid n_3; H_o) = \frac{1}{4}(n-n_3). \end{cases}$$

Bij toepassing van de toets neemt men voor n_3 de bij de proef gevonden waarde, terwijl men als kritieke zône voor $\underline{n}_1$ bij tweezijdige toetsing een zône gebruikt van de vorm:

(8) $$Z_{I,n_3} : \left|\underline{n}_1 - \tfrac{1}{2}(n-n_3)\right| \geq a_I$$

en bij éénzijdige toetsing [2]) één van de beide volgende zônes:

---

[2]) Zie voor de keuze tussen één- en tweezijdige toetsing b.v. [4], § 6.

$$(9) \quad \begin{cases} Z'_{I,n_3} : & \underline{n}_1-\frac{1}{2}(n-n_3) \leqq -b_I \\ Z''_{I,n_3} : & \underline{n}_1-\frac{1}{2}(n-n_3) \geqq b_I \end{cases}$$

Hierin zijn de grootheden a_I en b_I afhankelijk van de waarde van n_3 en van de onbetrouwbaarheidsdrempel α. De afhankelijkheid van n_3 is bij de kritieke zônes aangegeven door de index n_3 bij het symbool Z. Het verband tussen a_I en α wordt gegeven door de relatie:

$$(10) \qquad P\left[\underline{n}_1 \in Z_{I,n_3} \mid n_3;\ H_0\right] \leqq \alpha,$$

waarbij bovendien a_I het ~~grootste~~ kleinste gehele getal is, dat aan (10) voldoet. Een dergelijke relatie geldt voor b_I, voor beide door (9) gegeven zônes.

Tot hier toe is de toets beschreven als een voorwaardelijke toets (met als voorwaarde $\underline{n}_3=n_3$). Deze voorwaarde kan echter gemakkelijk geëlimineerd worden. Daartoe nemen wij als onvoorwaardelijke kritieke zône Z_I de vereniging van alle kritieke zônes Z_{1,n_3}, dus:

$$(11) \qquad Z_1 = \bigcup_{n_3=0}^{n} Z_{I,n_3}.$$

Dan geldt (zie (10)):

$$(12) \qquad P\left[\underline{n}_1 \in Z_I \mid H_0\right] =$$

$$= \sum_{\nu=0}^{n} P\left[\underline{n}_3=\nu \mid H_0\right] P\left[\underline{n}_1 \in Z_{I,n_3} \mid \underline{n}_3=\nu\ ;\ H_0\right] \leqq$$

$$\leqq \alpha \sum_{\nu=0}^{n} P\left[\underline{n}_3=\nu \mid H_0\right] = \alpha\ ,$$

daar $\sum_{\nu=0}^{n} P\left[\underline{n}_3=\nu \mid H_0\right] = 1$ is. Wij zien dus, dat deze toets ook onvoorwaardelijk α als onbetrouwbaarheidsdrempel bezit.

De grootheden a_I en b_I kunnen voor verschillende waarden van α gevonden worden in [1] voor $n-n_3 \leqq 100$, en in VAN WIJNGAARDEN [6] voor iedere α en voor $n-n_3 \leqq 200$. Men kan ook gebruik maken van de normale benadering van de door (6) gegeven binomiale verdeling. De kritieke zône Z_{I,n_3} wordt dan vervangen door

$$(8') \quad Z^*_{I,n_3} : \left|\underline{n}_1-\tfrac{1}{2}(n-n_3)\right| \geqq \tfrac{1}{2}\, \xi_{\frac{1}{2}\alpha} \sqrt{n-n_3},$$

waarin $\xi_{\frac{1}{2}\alpha}$ voldoet aan de vergelijking

$$(13) \qquad \frac{1}{\sqrt{2\pi}} \int_{\xi_u}^{\infty} e^{-\frac{1}{2}x^2}\, dx = u,$$

met $u=\frac{1}{2}\alpha$. Bij gegeven u kan ξ_u in tabellen van de normale verdeling worden "terug"-gezocht.
De benadering van Z_{I,n_3} die men op deze wijze verkrijgt, is voor $\alpha=0,05$ reeds bij kleine $n-n_3$ bevredigend. Een nog nauwkeurigere benadering verkrijgt men door toepassing van de z.g. continuïteitscorrectie, die bestaat uit het vermeerderen van het rechterlid van de ongelijkheid (8') met $\frac{1}{2}$. Noemen wij de aldus verkregen kritieke zône Z^{**}_{I,n_3}, dan blijken bij numeriek narekenen de volgende eigenschappen te gelden, voor $n-n_3=1, 2, \ldots, 100$ en $\alpha=0,05$:

1ᵉ $Z^{**} \subset Z \subset Z^*$; [3])

2ᵉ $Z^* = Z$ of $Z^{**}=Z$ (of beide);

3ᵉ $Z^{**}=Z$ behalve voor $n-n_3=17$ en 94; in die gevallen behoren de twee het dichtst bij $\frac{1}{2}(n-n_3)$ gelegen waarden van n_1, die in Z^{**} liggen, niet tot Z.

4ᵉ Z^* is vaak groter dan Z, maar het verschil bestaat hoogstens uit één waarde van n_1 aan beide uiteinden van Z.[4])

Op analoge wijze kan men de beide éénzijdige kritieke zônes benaderen; daarbij treedt dan ξ_α in de plaats van $\xi_{\frac{1}{2}\alpha}$.

Opmerking:
In § 1 hebben wij reeds een aan methode I verbonden bezwaar besproken, dat de toepassing van deze methode in sommige gevallen niet wenselijk maakt.

§ 5. Methode II: Verdeling van de waarnemingen [C] over de groepen [A] en [B].

Men tracht vaak rekening te houden met de onbesliste experimenten, door deze voor de helft bij de uitkomsten [A] en voor de helft bij de uitkomsten [B] te tellen. Als toetsingsgrootheid gebruikt men dan dus in plaats van $\underline{n}_1$ de grootheid:

$$\underline{n}_1' = \underline{n}_1 + \tfrac{1}{2}\underline{n}_3, \tag{14}$$

terwijl de getoetste hypothese H_0 dezelfde is als bij methode I (zie (5)). Men voert nu de toets uit, alsof $\underline{n}_1'$ een binomiale verdeling bezat van de vorm:

$$P[\underline{n}_1' = n_1' \mid H_0] = 2^{-n}\binom{n}{n_1'}, \tag{15}$$

met als gemiddelde en spreidingskwadraat:

---

[3]) $\subset$ betekent: is bevat in. De indices 1 en n_3 zijn voor de eenvoud weggelaten.

[4]) Deze berekeningen zijn uitgevoerd door J. VAN KLINKEN.

(16) $\frac{1}{2}n$ resp. $\frac{1}{4}n$.

Bij tweezijdige toetsing gebruikt men dus een kritieke zône van de vorm:

(17) Z_{II}: $|\underline{n}_1'-\frac{1}{2}n| = |\underline{n}_1-\frac{1}{2}(n-\underline{n}_3)| \geq a_{II}$,

en bij éénzijdige toetsing twee met (9) analoge kritieke zônes Z_{II}' resp. Z_{II}''. Bij gebruik van de normale benadering wordt Z_{II} vervangen door Z_{II}^* :

(17') Z_{II}^*: $|\underline{n}_1-\frac{1}{2}(n-\underline{n}_3)| \geq \frac{1}{2}\xi_{\frac{1}{2}\alpha}\sqrt{n}$,

terwijl bij het rechterlid, als continuïteitscorrectie vaak $\frac{1}{2}$ wordt opgesteld.

Opmerking:
In het volgende zal blijken, dat (15) in werkelijkheid niet geldt; dit volgt trouwens reeds uit het feit, dat $\underline{n}_1'$ ook halve waarden aan kan nemen, waarmee bij (15) geen rekening gehouden wordt. Wij zullen de volgens (15) berekende onbetrouwbaarheid van methode II de nominale onbetrouwbaarheid noemen.

§ 6. Eigenschappen van de methoden I en II.

In de appendix (zie § A.2 en A.3) worden enkele stellingen bewezen, die wij hier laten volgen.

Stelling 1: Indien $p_1/p_2=c$, worden gemiddelde en spreidingskwadraat van $\underline{n}_1$ resp. $\underline{n}_1'$ gegeven door:

(18) $$\begin{cases} \mathcal{E}(\underline{n}_1|\, n_3;c) = (n-n_3)\dfrac{c}{1+c}, \\ \sigma^2(\underline{n}_1|\, n_3;c) = (n-n_3)\dfrac{c}{(1+c)^2}, \end{cases}$$

resp.

(19) $$\begin{cases} \mathcal{E}(\underline{n}_1'|\,c) = n\left\{\dfrac{c}{1+c} + p_3\left(\frac{1}{2}-\dfrac{c}{1+c}\right)\right\}, \\ \sigma^2(\underline{n}_1'|\,c) = n(1-p_3)\left\{\frac{1}{4}p_3+\dfrac{c}{(1+c)^2}(1-p_3)\right\}. \end{cases}$$

Gevolgtrekkingen:
Voor c=1 zou (19) eigenlijk moeten overgaan in (16), daar c=1 overeenkomt met H_0. Voor het gemiddelde is dit inderdaad het geval, maar voor het spreidingskwadraat slechts indien $p_3=0$ is.

Immers wij vinden voor c=1

$$\sigma^2(\underline{n}_1'|c=1) = \tfrac{1}{4}n(1-p_3)$$

en dit is voor $p_3 > 0$, dus kleiner dan de in (16) opgegeven waarde $\frac{1}{4}n$. Daar ook de grootheid $\underline{n}_1'$ asymptotisch normaal verdeeld is, volgt hieruit, dat voor voldoende grote n de werkelijke onbetrouwbaarheid kleiner is dan de nominale, volgens (15) berekende, onbetrouwbaarheid. Het is echter duidelijk, dat de werkelijke onbetrouwbaarheid β_o van p_3 afhankelijk is, daar p_3 in (19) voorkomt. Dit heeft ten gevolge, dat β_o onbekend blijft en dat slechts de nominale onbetrouwbaarheid als bovengrens daarvan bekend is. Daar een vermindering van de onbetrouwbaarheid gewoonlijk gepaard gaat met een vermindering van het onderscheidingsvermogen, moet dit als een nadeel van de methode worden gezien. Dit nadeel blijkt duidelijk uit de volgende stelling:

<u>Stelling</u> 2:
<u>Bij iedere gegeven waarde van n en $\alpha < \frac{1}{2}$ en bij iedere waarde van</u> $p_3 > 0$ <u>zijn het onderscheidingsvermogen en de onbetrouwbaarheid van methode II kleiner dan die van methode I</u>. <u>Het verschil neemt toe met toenemende</u> p_3.

<u>Opmerking</u>:
Bij deze stelling moet een kleine restrictie worden gemaakt, als gevolg van het feit, dat bij gegeven waarde van α de werkelijke onbetrouwbaarheid van methode I en de nominale ~~werkelijke~~ onbetrouwbaarheid van methode II meestal niet precies gelijk aan α kunnen worden gemaakt. De aard van deze weinig belangrijke restrictie blijkt duidelijk uit het in § A.3 gegeven exacte bewijs en blijft hier onvermeld.

<u>Gevolgtrekkingen</u>:
Uit deze stelling volgt, dat uit theoretisch oogpunt methode I verre te verkiezen is boven methode II. Daar voor toenemende p_3 de kans op grotere waarden van $\underline{n}_3$ bij gegeven n toeneemt, neemt het aantal $(n-\underline{n}_3)$ waarnemingen, dat voor toetsing volgens methode I overblijft, af. Hier~~onder~~door neemt, zoals bekend, ook het onderscheidingsvermogen af. Uit de laatste zin van de stelling blijkt, dat dit voor het onderscheidingsvermogen van methode II nog sterker het geval is. Tevens volgt uit deze laatste zin nogmaals de gevolgtrekking, die wij voor grote n, bij de vorige stelling reeds hebben gemaakt. Voor die gevallen, waarbij men methode I niet wenst te gebruiken, ten gevolge van het in § 1 genoemde bezwaar, wordt in § 8 eng een derde methode gegeven, die dit bezwaar vermijdt.

§ 7. Een soms voorgestelde foutieve methode.

De vraag wordt wel eens gesteld, of men er niet goed aan zou doen de onbesliste experimenten in de verhouding $\underline{n}_1/\underline{n}_2$ te verdelen over de groepen met uitkomst [A] resp. [B], en dan verder de weg van methode II te volgen. Bij deze methode, die we met II' zullen aangeven, zou dus als toetsingsgrootheid de grootheid

$$(20) \qquad \underline{n}_1'' = \underline{n}_1 + \frac{\underline{n}_1}{\underline{n}_1+\underline{n}_2}\,\underline{n}_3 = \frac{n\,\underline{n}_1}{n-\underline{n}_3}$$

gebruikt worden, terwijl men dan verder zou handelen, alsof $\underline{n}_1''$ de in (15) voor $\underline{n}_1'$ aangegeven verdeling bezat. Wij zullen (in § A.2) omtrent deze methode de volgende stelling aantonen:

<u>Stelling</u> 3:
<u>Indien</u> $p_1/p_2=c$ <u>is, geldt voor gemiddelde en spreidingskwadraat van</u> $\underline{n}_1''$:

$$(21) \qquad \begin{cases} \mathcal{E}(\underline{n}_1''\mid c) = \dfrac{nc}{1+c}, \\ \sigma^2(\underline{n}_1''\mid c) \geqq \dfrac{nc}{(1+c)^2}\,\dfrac{1}{1-p_3}, \end{cases}$$

<u>waarbij in de ongelijkheid het gelijkteken dan en slechts dan geldt, als</u> $p_3=0$ <u>is</u>.

Voor $c=1$ (d.w.z. als H_0 geldt) komt het gemiddelde overeen met (16). Het spreidingskwadraat is dan echter groter dan $\frac{1}{4}n$, zodra $p_3 \neq 0$ is. Dit betekent, dat dan, voor voldoende grote n, de kans op het vinden van een waarde in de kritieke zône, indien H_0 juist is, in werkelijkheid groter is dan de nominale waarde, die men verkrijgt door de berekening op (15) te basere[n]. De onbetrouwbaarheid van deze methode zal dus veelal groter zijn dan de opgegeven onbetrouwbaarheidsdrempel α, hetgeen niet acceptabel is.

§ 8. Methode III: Combinatie van methode I met toetsing van p_3.

Men kan vermijden, dat bij een groot aantal onbesliste experimenten toch tot een verschil in kwaliteit tussen de middelen A en B besloten wordt, en tevens, dat de onbesliste experimenten, indien het er niet teveel zijn, het onderscheidingsvermogen van de toets nadelig beïnvloeden, door het toepassen van de volgende methode:

Men kiest een getal k tussen 0 en 1 (over de keuze van k later meer) en toetst in de eerste plaats de hypothese:

(22) $$H_o' : p_3 \geq k.$$

Kan deze hypothese niet verworpen worden, dan is de toets afgelopen en luidt de conclusie, dat er geen verschil van practische betekenis tussen de beide middelen gevonden is. Wordt H_o' echter verworpen, dus concludeert men, dat $p_3 < k$ is, dan toetst men nu volgens methode I (dus met weglating van de onbesliste experimenten), de hypothese H_o, dat $p_1=p_2$ is. Wordt deze niet verworpen, dan is de conclusie dezelfde, als wanneer H_o' niet verworpen wordt. Wordt echter H_o ook verworpen, dan concludeert men tot een verschil in kwaliteit tussen A en B (en wijst men van deze twee de beste aan naar aanleiding van de grootteverhouding van n_1 en n_2).

De toetsing van H_o' berust op $\underline{n}_3$ als toetsingsgrootheid. Indien $p_3=k$ is, bezit $\underline{n}_3$ de volgende waarschijnlijkheidsverdeling:

(23) $$P\left[\underline{n}_3=n_3 \mid p_3=k\right] = \binom{n}{n_3} k^{n_3}(1-k)^{n-n_3},$$

met als gemiddelde een spreidingskwadraat:

(24) $$\begin{cases} \mathcal{E}(\underline{n}_3 \mid p_3=k)= n k ; \\ \sigma^2(\underline{n}_3 \mid p_3=k)= n k(1-k). \end{cases}$$

De toetsing geschiedt uiteraard éénzijdig (zie b.v. [4] § 6.) en wel linkszijdig, d.w.z. met een kritieke zône Z'_{III} van de vorm:

(25) $$Z_{III}' : \underline{n}_3-k n \leq -b_{III}.$$

De toetsing van H_o (indien deze aan de beurt komt), geschiedt volgens de beschrijving, die in § 4. gegeven is. Nu geldt, zoals verderop (§ A.4) bewezen wordt:

<u>Stelling 4:</u>

<u>Indien men H_o' linkszijdig toetst met onbetrouwbaarheidsdrempel α en, bij verwerping van H_o', H_o toetst volgens methode I eveneens met onbetrouwbaarheidsdrempel α; en indien men de hypothese, dat de middelen A en B geen verschil in kwaliteit vertonen dan en slechts dan verwerpt, indien zowel H_o' als H_o verworpen worden; dan bezit de aldus verkregen toets, waarbij de hypothese</u>

(26) $$H_o'' = H_o' \text{ of } H_o$$

<u>getoetst wordt, de onbetrouwbaarheidsdrempel</u> α.

Opmerkingen:
Het is duidelijk, dat nu inderdaad bij grote p_3 in de regel niet tot een kwaliteitsverschil besloten zal worden. Tevens garandeert het gebruik van methode I een groot onderscheidingsvermogen als n_3 niet groot is. Het spreekt echter tevens vanzelf, dat methode III minder vaak tot verwerping van H_o leidt, dan methode I. Dit is te wijten aan het verschil van de getoetste hypothesen. Verwerping van H_o" houdt verwerping van H_o in, maar niet andersom.

Opmerkelijk is, dat de onbetrouwbaarheidsdrempel van deze methode dezelfde is als van beide erin verwerkte toetsingen apart. Dit betekent, indien men met overschrijdingskansen werkt, dat de overschrijdingskans van methode III gelijk is aan de grootste der overschrijdingskansen verkregen bij de toetsing van H_o' resp. H_o.

De volgorde waarin H_o' en H_o worden getoetst is van geen belang en kan dus ook worden omgedraaid. Dit is de practijk wellicht wenselijk, daar er voor de toetsing van H_o volgens methode I meer tabellen beschikbaar zijn dan voor toetsing van H_o'. Men kan bij deze laatste toetsing gebruik maken van de tafels der onvolledige Bèta-functie van K.Pearson of, indien n niet te klein is en k niet te dicht bij 1 ligt, van de normale benadering van verdeling (23).

§ 9. De keuze van k; combinatie van methode I en II.

Methode III berust op de gedachte, dat men slechts dan rekening kan houden met de onbesliste experimenten, indien men in de definitie van "gelijke-" en "ongelijke"kwaliteit" met de mogelijkheid van het optreden van deze experimenten rekening houdt. Dit wordt nu gedaan door k zo te bepalen, dat men alleen indien p_3 duidelijk kleiner dan k is, de aanwezigheid van een kwaliteitsverschil van enige betekenis mogelijk acht. De keuze van k kan niet op theoretische gronden worden verricht; men zal hier proeven op wasmiddelen, waarvan bekend is, dat zij van kwaliteit verschillen, te hulp moeten roepen. Hier raken wij een heel complex van practische en experimentele problemen aan, daar de grootte van k b.v. ook van de experimenteertechniek afhankelijk zal zijn, terwijl de keuze van k toch op zekere hoogte arbitrair blijft. Men kan daarom, bij wijze van compromis, de keuze van k op de volgende wijze vastleggen:

Indien methode II, bij de gevonden waarde van n_3 en de gegeven waarde van n en α, nog tot verwerping van H_0 zou kunnen leiden, d.w.z. indien de uiterste waarde, die n_1' in dit geval kan aannemen in Z_{II} ligt, gaat men over tot toetsing van H_0 volgens methode I.
Verwerping van H_0 leidt dan tot de conclusie, dat A en B verschillend van kwaliteit zijn, terwijl hiertoe niet geconcludeerd wordt, indien ~~aan~~ één van beide genoemde voorwaarden niet voldaan is.

Door deze regel wordt voor iedere n en α een waarde van k vastgelegd, die gemakkelijk te bepalen is, maar die voor toepassing van de methode in het geheel niet bepaald hoeft te worden. Immers de eerste stap van de toets bestaat nu hieruit, dat men (vgl. § 5.) nagaat of de waarde $n_1'=\frac{1}{2}n_3$ nog in Z_{II} (of bij links-éénzijdige toetsing in Z_{II}' resp. bij rechts-éénzijdige toetsing $n-\frac{1}{2}n_3$ in Z_{II}'') ligt. Alleen indien dit zo is, gaat men over tot toepassing van methode I, daarbij dezelfde onbetrouwbaarheidsdrempel gebruikende als die van Z_{II} (resp. Z_{II}' en Z_{II}'').

De grootheid k is nu, behalve van n, ook van α afhankelijk en wij zullen zien (zie § A.4), dat

$$(27) \qquad \lim_{n \to \infty} k = 1$$

is voor iedere α. Dit betekent enerzijds, dat de mogelijkheid van onderscheiding tussen twee wasmiddelen met toenemende n stijgt (hetgeen geheel in de lijn van een statistische analysemethode ligt), anderzijds zou men echter kunnen zeggen, dat aan het tegen methode I uitgebrachte bezwaar toch niet volledig tegemoet gekomen is. Indien men deze laatste mening is toegedaan kan men de k bepalen, die volgens het hier gegeven principe behoort bij een waarde van n, die men bij een bepaalde proefopzet gewoonlijk (ongeveer) gebruikt en vervolgens voor die proeven, waarbij men een ander aantal waarnemingen heeft, met de zo gevonden waarde van k de in de vorige paragraaf uiteengezette methode volgen.

Wij kunnen hier niet verder op deze kwestie ingaan. Het is echter wel duidelijk, dat er vele mogelijkheden zijn, om langs een weg als de hier beschreven tot een oplossing te komen, die aan een bepaalde proefopzet en aan het doel daarvan op bevredigende wijze ~~aanpast.~~ is aangepast.

Tenslotte zij nog opgemerkt, dat afhankelijkheid van k en n de geldigheid van methode II niet stoort, daar stelling 4 voor iedere k en n geldt.

§ 10. Een verfijning van het model.

In sommige gevallen ligt het voor de hand het optreden van onbesliste experimenten aan een onvoldoend nauwkeurige meting of beoordeling van het experimentele resultaat te wijten, of ook aan het gebruik van afgeronde meetresultaten (dit is vooral het geval, indien men met verschillen x_i-y_i werkt, zoals in § 1. beschreven is). Verder kan men ook de mogelijkheid van foutieve beoordeling van een resultaat in de beschouwingen betrekken.

Geven wij de "werkelijke" uitkomst aan met $A > B$ resp. $A < B$ indien middel A het beste resp. het slechtste resultaat heeft gegeven en laten wij de mogelijkheid van verkeerde beoordeling toe, dan lijkt het b.v. redelijk het volgende model te kiezen:

$$\left.\begin{array}{l} P\,[A > B]=p' \\ P\,[A < B]=q' \end{array}\right\} \qquad p'+q'=1,$$

terwijl voor het tot stand komen der beoordelingen, die weer met $[A]$, $[B]$ resp. $[C]$ worden aangegeven, geldt:

$$\left.\begin{array}{l} P\,[A \mid A > B]=P\,[B \mid B > A]=d_1 \\ P\,[B \mid A > B]=P\,[A \mid B > A]=d_2 \\ P\,[C \mid A > B]=P\,[C \mid B > A]=d_3 \end{array}\right\} \qquad d_1+d_2+d_3=1.$$

De te toetsen hypothese is nu: $p'=q'$.
Daar nu echter:

$$p_1=P\,[A]=P\,[A > B].P\,[A \mid A > B]+P\,[B > A].P\,[A \mid B > A]=p'd_1+q'd_2$$

en analoog

$$p_2=P\,[B]=p'd_2+q'd_1$$

is, volgt uit $p'=q'$ ook: $p_1=p_2$, zodat de toetsing bij dit model op dezelfde wijze plaats kan vinden als bij het eenvoudige in § 3. beschreven model.

Verder geldt uiteraard:

$$p_3=P\,[C]=d_3\ ,$$

zodat volgens stelling 2 de beoordeling volgens methode II s steeds slechter wordt naarmate d_3 toeneemt. Wij kunnen daarbij opmerken, dat dit ook voor methode I (en derhalve ook voor III) het geval is, daar met stijgende p_3 de kans op grotere waarden van $\underline{n}_3$ stijgt, waardoor het aantal waarnemingen beschikbaar voor de toetsing volgens methode I afneemt.
Uit stelling 2 blijkt echter, dat dit effect bij methode I kleiner is dan bij II. Verder valt op te merken. dat methode III, door de keuze van k, aan de grootte van d_3 kan worden

aangepast, indien men door proefexperimenten iets van d_3 te weten kan komen.

§ 11. Enkele opmerkingen.

Het ligt, bij niet te kleine n, voor de hand de minimum-χ^2-methode (zie b.v. [5] pag. 424 e.v.) toe te passen. Wij vermelden hier zonder bewijs, dat deze methode asymptotisch aequivalent is met methode I.

Tenslotte een opmerking van technische aard: het is van groot belang, dat de experimentator bij de beoordeling van de proefresultaten er, voor ieder stuk wasgoed, dat hij in handen krijgt, niet van op de hoogte is, welk wasmiddel gebruikt is. Psychologische onzuiverheden in de beoordeling zijn anders moeilijk te vermijden.

§ 12. Samenvatting.

a. Methode I (weglaten der onbesliste experimenten) bezit een groter onderscheidingsvermogen dan methode II (half om half verdelen van de onbesliste experimenten).

b. Verdeling in de verhouding n_1/n_2 is onjuist.

c. De beide methoden kunnen gecombineerd worden tot een toetsingsmethode voor een uit twee delen bestaande hypothese, die een andere definitie van kwaliteitsverschi inhoudt en die aan de practijk kan worden aangepast.

d. Het onderscheidingsvermogen van de methoden I, II en III neemt af, als de beoordeling in scherpte afneemt. Deze afname is het sterkst bij methode II.

Appendix.

A.1. Een hulpstelling.

Voor het berekenen van de gemiddelden en spreidingskwadraten der verschillende toetsingsgrootheden maken wij gebruik van een hulpstelling, ontleend aan het college Capita Selecta van Prof. Dr D. VAN DANTZIG.

Indien $\underline{x}$ en $\underline{y}$ twee stochastische grootheden zijn, die een simultane waarschijnlijkheidsverdeling bezitten, en indien

$$\varphi(y) = \mathcal{E}(\underline{x}|y) = \mathcal{E}(\underline{x}|\underline{y}=y) \tag{A,1}$$

en

$$\psi(y) = \sigma^2(\underline{x}|y) = \sigma^2(\underline{x}|\underline{y}=y), \tag{A,2}$$

het voorwaardelijke gemiddelde resp. spreidingskwadraat van $\underline{x}$ zijn, onder de voorwaarde $\underline{y}=y$, dan zijn $\varphi(\underline{y})$ en $\psi(\underline{y})$ zelf weer stochastische grootheden. Nu geldt:

Stelling A.I:

$$\mathcal{E}\,\underline{x}=\mathcal{E}\,\varphi(\underline{y});$$
$$\sigma^2(\underline{x})=\sigma^2\{\varphi(\underline{y})\}+\mathcal{E}\{\psi(\underline{y})\}. \tag{A.3}$$

Het bewijs van deze stelling zullen wij hier niet geven. Men vindt het in compacte vorm in [7], pag.59.

A.2. Bewijs van de stellingen 1 en 3.

Voor $p_1/p_2=c$ is:

$$\frac{p_1}{p_1+p_2}=\frac{c}{1+c} \quad \text{en} \quad \frac{p_2}{p_1+p_2}=\frac{1}{1+c}\,. \quad (\text{en } p_1+p_2=1-p_3) \tag{A.4}$$

Hieruit volgt (18) onmiddellijk; immers $\underline{n}_1$ heeft, onder de voorwaarde $\underline{n}_3=n_3$, een binomiale verdeling met als parameters $n-n_3$ (het aantal niet onbesliste experimenten) en $\frac{p_1}{p_1+p_2}$ (de kans op [A] bij ieder der experimenten).

Voor het bewijs van de tweede vergelijking van (19) maken wij gebruik van de stelling A.I, met $\underline{n}_1'$ als $\underline{x}$ en $\underline{n}_3$ als $\underline{y}$. Steeds wordt $p_1/p_2=c$ ondersteld; voor de eenvoud der notatie wordt dit niet steeds vermeld.

Nu is:

$$\mathcal{E}\,\underline{n}_1'=\mathcal{E}(\underline{n}_1+\tfrac{1}{2}\underline{n}_3)=\mathcal{E}\,\underline{n}_1+\tfrac{1}{2}\,\mathcal{E}\,\underline{n}_3=np_1+\tfrac{1}{2}np_3,$$

waaruit met behulp van (A.4) gemakkelijk de eerste der twee vergelijkingen (19) gevonden wordt.

Verder is (vgl. (A.1) en (A.2) en (18)):

$$\varphi(n_3)=\mathcal{E}(\underline{n}_1'\,|\,n_3)=\frac{c}{1+c}(n-n_3)+\tfrac{1}{2}n_3,$$

en

$$\psi(n_3)=\sigma^2(\underline{n}_1'\,|\,n_3)=\frac{c}{(1+c)^2}(n-n_3),$$

dus is volgens (A.3):

$$\sigma^2(\underline{n}_1')=\sigma^2\left\{\frac{c}{1+c}(n-\underline{n}_3)+\tfrac{1}{2}\underline{n}_3\right\}+\mathcal{E}\left\{\frac{c}{(1+c)^2}(n-\underline{n}_3)\right\}=$$

$$=\sigma^2\left\{\underline{n}_3\left(\tfrac{1}{2}-\frac{c}{1+c}\right)\right\}+\frac{c}{(1+c)^2}n(1-p_3)=$$

$$=\left(\tfrac{1}{2}-\frac{c}{1+c}\right)^2 np_3(1-p_3)+\frac{c}{(1+c)^2}n(1-p_3)=$$

$$=n(1-p_3)\left\{\tfrac{1}{4}p_3+\frac{c}{(1+c)^2}(1-p_3)\right\}.$$

Hiermee is stelling 1 bewezen. Het bewijs van stelling ~~4~~ 3 verloopt als volgt:

Neem in (A.3) $\underline{n}_1''$ voor $\underline{x}$ en $\underline{n}_3$ voor $\underline{y}$, dan is (vgl.(20) en (18)):

$$\varphi(n_3)=\mathcal{E}(\underline{n}_1''|n_3)=\frac{n}{n-n_3}\mathcal{E}(\underline{n}_1|n_3)=$$

$$=\frac{n}{n-n_3}(n-n_3)\frac{c}{1+c}=\frac{nc}{1+c}.$$

Deze grootheid is onafhankelijk van n_3, dus is ook

$$\mathcal{E}\varphi(\underline{n}_3)=\frac{nc}{1+c}.$$

Verder is:

$$\psi(n_3)=\sigma^2(\underline{n}_1''|n_3)=\frac{n^2}{(n-n_3)^2}\sigma^2(\underline{n}_1|n_3)=$$

$$=\frac{n^2}{(n-n_3)^2}(n-n_3)\frac{c}{(1+c)^2}=\frac{n^2c}{(n-n_3)(1+c)^2}.$$

Daar $\varphi(n_3)$ onafhankelijk is van n_3, is $\sigma^2\{\varphi(n_3)\}=0$, dus wordt (A.3) nu:

$$\sigma^2(\underline{n}_1'')=\mathcal{E}\left\{\frac{n^2c}{(1+c)^2}\frac{1}{n-\underline{n}_3}\right\}=\frac{n^2c}{(1+c)^2}\mathcal{E}\left(\frac{1}{n-\underline{n}_3}\right).$$

Daar $0\leq n_3\leq n$ is, is $\frac{1}{n-n_3}$ een positief-convexe functie van n_3. Volgens de ongelijkheid van JENSEN [8] (zie b.v. [9] pag. 94), geldt daarom:

$$\mathcal{E}\frac{1}{n-\underline{n}_3}\geq\frac{1}{n-\mathcal{E}\underline{n}_3},$$

waarin alleen dan het gelijkheidsteken geldt, indien $\underline{n}_3$ met waarschijnlijkheid 1 slechts één waarde aanneemt, d.w.z. indien $p_3=0$ is. Voor $p_3>0$ hebben wij dus, daar $\mathcal{E}\underline{n}_3=np_3$ is:

$$\sigma^2(\underline{n}_1'') > \frac{n^2c}{(1+c)^2} \quad \frac{1}{n-np_3} = \frac{nc}{(1+c)^2} \quad \frac{1}{1-p_3} \; .$$

A.3. Bewijs van Stelling 2.

Daar dit bewijs nogal ingewikkeld is, laten wij een eenvoudiger bewijs vooraf gaan, dat asymptotisch voor $n \to \infty$ geldig is en de stelling dus slechts "voor voldoende grote n" bewijst.

a. Asymptotisch bewijs.

Zijn zekere waarden n_1, n_2 en n_3 gevonden, en toetsen wij tweezijdig, dan leidt methode I volgens de normale benaderingsmethode tot verwerping van H_0, indien (vgl. (8')):

$$|n_1 - \tfrac{1}{2}(n-n_3)| \geq \tfrac{1}{2} \xi_{\frac{1}{2}\alpha} \sqrt{n-n_3}$$

is. Methode II echter slechts dan (vgl. (17')), indien

$$|n_1 - \tfrac{1}{2}(n-n_3)| \geq \tfrac{1}{2} \xi_{\frac{1}{2}\alpha} \sqrt{n}$$

is. Het is duidelijk, dat aan de tweede ongelijkheid slechts voldaan kan zijn, indien aan de eerste voldaan is, terwijl het omgekeerde niet noodzakelijk is. Daar bovendien voor $p_3 \neq 0$ de stochastische grootheid $\underline{n}_3/n$ in waarschijnlijkheid tot p_3 convergeert, is er bij iedere waarde van c en bij voldoende grote n een positieve waarschijnlijkheid, dat methode I wel, maar methode II niet tot verwerping van H_0 leidt. Zowel het onderscheidingsvermogen als de onbetrouwbaarheid van methode I zijn dan groter dan die van methode II. Dit geldt ook bij éénzijdige toetsing, zoals geheel analoog bewezen wordt. Dat de onbetrouwbaarheid van methode II onbekend is volgt reeds uit het feit, dat in (19) de onbekende grootheid p_3 voorkomt. Tevens ziet men direct, dat het verschil in onderscheidingsvermogen tussen beide methoden met p_3 toeneemt, daar dan de kans op grotere waarden van n_3 toeneemt.

b. Exact bewijs.

Wij geven het bewijs voor éénzijdige toetsing. Daar het bewijs voor links- en rechtszijdige toetsing geheel gelijk verloopt (men behoeft slechts $\underline{n}_1$ en $\underline{n}_2$ te verwisselen), behoeft het slechts voor één van beide gegeven te worden en volgt de stelling dan ook direct voor tweezijdige toetsing, waarbij niet noodzakelijk van een symmetrische kritieke zône gebruik behoeft te worden gemaakt. De waarde $\frac{1}{2}n$ mag echter niet in de kritieke zônes voorkomen, hetgeen overeenkomt met $\alpha < \frac{1}{2}$.

Wij beschouwen een linker-kritieke zône voor methode I, bij gegeven waarde van n_3. Deze is van de vorm

$$Z'_{I,n_3} : \underline{n}_1 \leq d$$

met

$$P[\underline{n}_1 \leq d \mid n_3 ; H_0] = 2^{n_3 - n} \sum_{\nu=0}^{d} \binom{n-n_3}{\nu} = \beta_0 \leq \alpha .$$

Bij methode II is de overeenkomstige kritieke zône van de vorm

$$Z'_{II} : \underline{n}_1 + \tfrac{1}{2}\underline{n}_3 \leq d' ,$$

met een (nominale) onbetrouwbaarheid gelijk aan

$$\beta_0' = 2^{-n} \sum_{\nu=0}^{d'} \binom{n}{\nu} \leq \alpha .$$

Wij bewijzen nu, dat voor iedere gegeven positieve waarde van n_3 de onbetrouwbaarheid van Z'_{I,n_3}, berekend volgens de methode, waarmee de nominale onbetrouwbaarheid van methode II bepaald wordt, groter is dan β_0. Dit betekent dan, dat $d' < d$ is, indien wij afzien van toevallige gevallen, te wijten aan het feit, dat men β_0 gewoonlijk niet precies gelijk aan α ~~kunnen~~ kan maken. Wij hebben dan bewezen, dat verwerping van H_0 met methode II slechts dan op kan treden, indien dit met methode I het geval is, dus juist wat wij boven voor voldoende grote n ~~berekend~~ bewezen hebben.

Voor $n_1 = d$ is $n_1' = d + \frac{1}{2}n_3$, dus is $[d + \frac{1}{2}n_3 + \frac{1}{2}]$[5] de grootste waarde van de met Z'_{I,n_3} corresponderende kritieke zône van methode II. Wij moeten nu dus bewijzen:

$$\text{(A.5)} \qquad 2^{n_3 - n} \sum_{\nu=0}^{d} \binom{n-n_3}{\nu} < 2^{-n} \sum_{\nu=0}^{[d+\frac{1}{2}n_3+\frac{1}{2}]} \binom{n}{\nu} .$$

Hierbij beperken wij ons tot $\alpha < \frac{1}{2}$, hetgeen betekent, dat wij ons slechts bezig houden moeten met waarden van d en n_3, waarvoor

$$\text{(A.6)} \qquad [d + \tfrac{1}{2}n_3 + \tfrac{1}{2}] < \tfrac{1}{2}n - \tfrac{1}{2}$$

is.

Immers indien hier niet aan voldaan is wordt het rechterlid van (A.5) $\geq \frac{1}{2}$, zodat de ongelijkheid, daar het ~~rechter~~ Linkerlid $< \frac{1}{2}$ is, dan reeds bewezen is.

Wij voeren nu de volgende notatie in:

---

5) $[x]$ is per definitie het grootste gehele getal $\leq x$

$$(A.7)\quad \begin{cases} L_k = 2^k \sum\limits_{\nu=0}^{d} \binom{n-k}{\nu} \\ R_k = \sum\limits_{\nu=0}^{[d+\frac{1}{2}k+\frac{1}{2}]} \binom{n}{\nu}, \end{cases}$$

zodat dus $R_{2h+1}=R_{2h+2}$ is.

Wij moeten dan bewijzen:

(A.8) $L_k < R_k$ voor $[d+\frac{1}{2}k+\frac{1}{2}] < \frac{1}{2}n-\frac{1}{2}$ (A.9)

Het bewijs vindt plaats door inductie, waarbij de index k bij iedere stap 2 opvolgende waarden aanneemt. Wij bewijzen dus eerst, dat (A.8) geldt voor k=1 en k=2 (voor k=0 is $L = R$).

Voor k=1 geldt:

$$L_1 - R_1 = 2\sum_{\nu=0}^{d}\binom{n-1}{\nu} - \sum_{\nu=0}^{d+1}\binom{n}{\nu} = \sum_{\nu=0}^{d}\binom{n-1}{\nu} + \sum_{\nu=0}^{d}\left\{\binom{n-1}{\nu}-\binom{n}{\nu}\right\} - \binom{n}{d+1} =$$

$$= \sum_{\nu=0}^{d}\binom{n-1}{\nu} - \sum_{\nu=1}^{d}\binom{n-1}{\nu-1} - \binom{n}{d+1} = \binom{n-1}{d} - \binom{n}{d+1},$$

en dit is <0 als $d+1 < n$ is, hetgeen volgens (A.9) het geval is.

Voor k=2 geldt (daar $R_1 = R_2$ is):

$$L_2 - R_2 = L_2 - L_1 + L_1 - R_2 =$$

$$= 4\sum_{\nu=0}^{d}\binom{n-2}{\nu} - 2\sum_{\nu=0}^{d}\binom{n-1}{\nu} + L_1 - R_1 =$$

$$= 2\left(\sum_{\nu=0}^{d}\binom{n-2}{\nu} + \sum_{\nu=0}^{d}\left\{\binom{n-2}{\nu}-\binom{n-1}{\nu}\right\}\right) + L_1 - R_1 =$$

$$= 2\left(\sum_{\nu=0}^{d}\binom{n-2}{\nu} - \sum_{\nu=1}^{d}\binom{n-2}{\nu-1}\right) + L_1 - R_1 =$$

$$= 2\binom{n-2}{d} + \binom{n-1}{d} - \binom{n}{d+1} = \frac{(n-2)!}{(d+1)!(n-2-d)!}(2d+3-n),$$

en dit is <0, indien $2d+3 < n$ is, waaraan volgens (A.9) inderdaad voldaan is.

Zij vervolgens gegeven

(A.10) $L_{2h+1} < R_{2h+1}$ en $L_{2h+2} < R_{2h+2}$ $(d+h+1 < \frac{1}{2}n-\frac{1}{2})$,

dan bewijzen wij nu:

(A.11) $L_{2h+3} < R_{2h+3}$ en $L_{2h+4} < R_{2h+4}$, indien $d+h+2 < \frac{1}{2}n-\frac{1}{2}$.

Nu is, met $m=n-1$:

$$L_{2h+3} = 2^{2h+3}\sum_{\nu=0}^{d}\binom{n-2h-3}{\nu} = 2 . 2^{2h+2}\sum_{\nu=0}^{d}\binom{m-2h-2}{\nu},$$

hetgeen, wegens $L_{2h+2} < R_{2h+2}$, kleiner is dan

$$2\sum_{\nu=0}^{d+h+1}\binom{m}{\nu},$$

althans indien voldaan is aan $d+h+1 < \frac{1}{2}m-\frac{1}{2}=\frac{1}{2}n-1$; dit is volgens (A.10) een toegelaten beperking. Substitueren wij: $b=d+h+1$, dan geldt dus

$$L_{2h+3} < 2\sum_{\nu=0}^{b}\binom{n-1}{\nu},$$

en dit is, wegens $L_1 < R_1$, kleiner dan:

$$\sum_{\nu=0}^{b+1}\binom{n}{\nu} = \sum_{\nu=0}^{d+h+2}\binom{n}{\nu} = R_{2h+3},$$

indien tenminste $b+1 < n$, dus $d+h+1 < n$ is, waaraan weer is voldaan. Tenslotte is op analoge wijze, wegens $L_{2h+2} < R_{2h+2}$ en $L_2 < R_2$ en met $r=n-2$ en $b=d+h+1$:

$$L_{2h+4} = 2^{2h+4}\sum_{\nu=0}^{d}\binom{n-2h-4}{\nu} = 2^2 . 2^{2h+2}\sum_{\nu=0}^{d}\binom{r-2h-2}{\nu} <$$

$$< 2^2 . \sum_{\nu=0}^{d+h+1}\binom{r}{\nu} = 2^2\sum_{\nu=0}^{b}\binom{n-2}{\nu} < \sum_{\nu=0}^{b+1}\binom{n}{\nu} =$$

$$= \sum_{\nu=0}^{d+h+2}\binom{n}{\nu} = R_{2h+4},$$

waarbij wegens $d+h+2 < \frac{1}{2}n-\frac{1}{2}$ (zie (A.11)), weer aan de verschillende beperkende voorwaarden is voldaan. Waarmee het eerste deel der stelling bewezen is. De toename van het verschil van het onderscheidingsvermogen der beide methoden blijkt weer uit de toename van de waarschijnlijkheid van grote waarden van $\underline{n}_3$, tezamen met het feit, dat bij iedere stap der inductie de ongelijkheden scherper worden, doordat opnieuw kleiner-tekens optreden.

A.4. Bewijs van stelling ~~A~~.4. en (27).

a. Volgens methode III verwerpen wij $H_o'' = \{H_o$ of $H_o'\}$ dan en slechts dan, indien voldaan is aan

(A.12) $$n_3 \in Z'_{III}$$

en tevens aan

(A.13) $$n_1 \in Z_{I,n_3}.$$

Daarbij zijn Z'_{III} en Z_{I,n_3} zo bepaald, dat voldaan is aan:

(A.14) $$P[\underline{n}_3 \in Z'_{III} \mid H_o'] \leq \alpha,$$

en

(A.15) $$P[\underline{n}_1 \in Z_{I,n_3} \mid n_3; H_o] \leq \alpha.$$

Hierin is het linkerlid van (A.14) onafhankelijk van $c = p_1/p_2$, dus van de al of niet juistheid van H_o, en evenzo is het linkerlid van (A.15) onafhankelijk van het al of niet juist zijn van H_o'.

Wij moeten bewijzen:

(A.16) $$\alpha' = P[\underline{n}_3 \in Z'_{III} \text{ en } \underline{n}_1 \in Z_{I,n_3} \mid H_o \text{ of } H_o'] \leq \alpha,$$

waarin de index n_3 van Z_I de gevonden waarde n_3 aangeeft. Nu zijn er 3 mogelijkheden, indien $H_o'' = \{H_o$ of $H_o'\}$ juist is:

1e: H_o' is juist. Dan is (zie (A.14)):

$$\alpha' \leq P[\underline{n}_3 \in Z'_{III} \mid H_o] \leq \alpha,$$

daar de kans op het optreden van twee gebeurtenissen niet groter is dan de kans op één van die twee en daar de verdeling van $\underline{n}_3$ niet van p_1/p_2 afhangt.

2e: H_o is juist. Dan geldt analoog (zie (A.15)):

$$\alpha' \leq P[\underline{n}_1 \in Z_{I,n_3} \mid n_3; H_o] \leq \alpha.$$

3e: H_o en H_o' zijn beide juist, dan gelden 1e en 2e.

Hiermee is de stelling bewezen.

b. Tenslotte bewijzen wij (27).

In (27) is k zo gekozen, dat methode II nog juist significantie kan geven. Bij tweezijdige toetsing volgens methode II (voor éénzijdige verloopt het bewijs analoog), betekent dit, dat $\underline{n}_1'$ voor $\underline{n}_1 = 0$ nog juist in de kritieke zône Z_{II} ligt. Daar (27) een asymptotische betrekking is, mogen wij Z_{II} vervangen door Z^*_{II} (zie (17')), zodat wij vinden, dat $n_3 \in Z'_{III}$ ligt, indien voldaan is aan:

$$\left|\tfrac{1}{2}(n_3-n)\right| \geq \tfrac{1}{2}\xi_{\frac{1}{2}\alpha}\sqrt{n},$$

zodat dus Z'_{III} gegeven wordt door:

(A.17) $$n_3 \leq n-\xi_{\frac{1}{2}\alpha}\sqrt{n},$$

terwijl deze volgens (25) gegeven wordt door

(25) $$n_3 \leq k\,n-b_{III}.$$

Derhalve is

$$k\,n-b_{III}=n-\xi_{\frac{1}{2}\alpha}\sqrt{n}$$

en daar hierin $b_{III}=\mathcal{O}(\sqrt{n})$ is (immers $\underline{n}_3$ is bij iedere gegeven waarde van p_3 binomiaal verdeeld), is

$$k=1-\mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

waaruit (27) volgt.

Litteratuur.

[1]. W.J. Dixon and A.M.Mood, The statistical sign test, Journ.Amer.Stat.Ass. 41 (1946) p. 557-566.

[2]. W.J.Dixon and F.J.Massey Jr, An introduction to statistical analysis, N.Y. 1951.

[3]. A.M.Mood, Introduction to the theory of statistics, N.Y. 1950.

[4]. J.Hemelrijk en H.R.van der Vaart, Het gebruik van één- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica 4 (1950) p.54-66.

[5]. H.Cramér, Mathematical methods of Statistics, Princeton University Press, 1946.

[6]. A.van Wijngaarden, Table of the cumulative symmetric binomial distribution, Proc.Kon.Nederl.Akad.Wetensch. 53 (1950) pp. 857-868 en Indagationes Mathematicae 12 (1950) pp. 301-312.

[7]. J.Hemelrijk, Symmetrietoetsen en andere toepassingen van de theorie van Neyman en Pearson, Diss.,Den Haag 1950.

[8]. J.L.W.V.Jensen, Acta Mathematica 30 (1906).

[9]. D.van Dantzig, Capita waarschijnlijkheidsrekening, gestencilde college syllabus, Amsterdam 1946-1947.