MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

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

Vacantiecursus 1950

Waarschijnlijkheidsrekening

Waarschijnlijkheidsrekening en Statistiek.

door

J. Hemelrijk

1e druk: 1950
2e druk: 1955

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

1. Inleiding.

1.1. In deze voordracht zullen wij het gebruik van de waarschijnlijkheidsrekening in de mathematische statistiek aan enkele elementaire voorbeelden toelichten. Wij hebben deze voorbeelden gekozen uit de meest voorkomende statistische methoden met de bedoeling een illustratie te geven van de methodologie van de mathematische statistiek.

Wij zullen de waarschijnlijkheidsrekening beschouwen als een axiomatisch opgezet deel der zuivere wiskunde, met het begrip "waarschijnlijkheid" als abstract grondbegrip; de axiomas stellen ons in staat uit gegeven waarschijnlijkheden andere af te leiden. De mathematische statistiek beschouwen wij eveneens als een onderdeel der zuivere wiskunde, dat echter in hoge mate gericht is op de toepassing der theorie op diverse ervaringswetenschappen.

Deze toepassing komt bij een gegeven experiment of onderzoek op de volgende wijze tot stand:

A. Men kiest een mathematisch model, dat zo goed mogelijk bij het experiment of onderzoek aansluit en tevens niet zo ingewikkeld is, dat het mathematisch niet meer te beheersen is.

B. In dit model past men de theorie der waarschijnlijkheidsrekening en mathematische statistiek toe.

C. De verkregen wiskundige resultaten worden in conclusies of voorspellingen vertaald met behulp van het principe, dat met het optreden van een uitkomst met zeer geringe waarschijnlijkheid bij het verrichten van één enkel experiment geen rekening gehouden wordt.

Wij hopen deze punten door middel van de voorbeelden nader toe te lichten.

Men kan met evenveel recht de statistiek als een tak van toegepaste wiskunde beschouwen. Daar wij echter hoofdzakelijk aan

het zuiver mathematische gedeelte (punt B) aandacht willen besteden, prefereren wij voor het moment een scheiding tussen dit gedeelte, dat wij de mathematische statistiek zouden kunnen noemen, en de door de punten A en C aangegeven methode van toepassing.

1.2. Het model.

Bij ieder statistisch experiment of onderzoek wordt, ter bewerking van de gevonden resultaten, een mathematisch model gekozen. Bij deze keuze tracht men steeds met twee eisen zo goed mogelijk rekening te houden: enerzijds moet het model een zo getrouw mogelijk beeld zijn van het werkelijke experiment, anderzijds zal men trachten het zo eenvoudig te maken, dat het mathematisch beheerst kan worden.

Ter wille van een grotere aanschouwelijkheid kiest men als model vaak een geluksspel, zoals het werpen met een munt of dobbelsteen, het trekken van ballen uit een vaas of een ander soort loterij. Hierdoor wordt wel eens de indruk gewekt, alsof statistici zich speciaal met deze "gokspelletjes" zouden ophouden. Men ziet echter zelden een statisticus ballen uit een vaas trekken. Daar de waarde van een aanschouwelijke voorstelling niet te onderschatten is, zullen wij ook hier van een dergelijk model gebruik maken, al kan men dan, strikt genomen, niet van een zuiver mathematisch model spreken. In eenvoudige gevallen, zoals de hier te bespreken methoden, schuilt er in het gebruik van een dergelijk aanschouwelijk model weinig gevaar voor verlies van mathematische strengheid en precisie.

Het model, waarover wij zullen spreken is het werpen met één of twee (in het algemeen "onzuivere") munten, waarbij wordt ondersteld, dat het resultaat van iedere worp onafhankelijk is van het bij vroegere worpen verkregene. Dit model wordt verderop nog gepreciseerd.

Als voorbeeld van onderzoekingen, waarbij dit model vaak wordt gebruikt, noemen wij:

1. De bestudering van de verhouding van het aantal jongens- en meisjesgeboorten.

2. De bestudering van de kwaliteit van een massaproduct, indien deze kwaliteit kan worden uitgedrukt als het percentage ondeugdelijke exemplaren in een afgeleverde partij.

3. De bestudering van het percentage verzekerden van een gegeven categorie, die voor een bepaalde leeftijd overlijden.

4. Het vergelijken van twee fabricagemethoden of geneesmiddelen.

5. Ook bij vele theoretische problemen van de statistiek worden het model, of de binnen het model geldende stellingen, veelvuldig gebruikt.

De vraag, in hoeverre het gerechtvaardigd is bij deze en andere problemen het genoemde model te gebruiken, zullen wij niet in het geding brengen.

1.5. De axiomas.

Het begrip waarschijnlijkheid of kans kunnen wij opvatten als een mathematisch analogon (of, zo men wil, een mathematische idealisering) van het begrip frequentiequotiënt. Het frequentiequotiënt van het kenmerk "kruis" (K) bij een serie worpen met een munt is het aantal malen, dat K optreedt, gedeeld door het totale aantal worpen. Een dergelijk frequentiequotiënt kan worden opgevat als een meting (of zo men wil: schatting) van de waarschijnlijkheid van het kenmerk K bij het werpen van deze munt, waarbij dan deze waarschijnlijkheid een abstracte mathematische grootheid binnen het mathematisch model is (vergelijk het begrip "lengte" van een staaf).

Heeft men een eindige verzameling Λ van N elementen, die ieder één of meer van de kenmerken A, B, C enz. dragen, dan is het frequentiequotiënt van A (en analoog van B, C enz.) gedefinieerd als het aantal elementen, n(A), dat het kenmerk A draagt, gedeeld door het totale aantal elementen, N.

Notatie:

$$(1) \qquad fq_{\Lambda}(A) = \frac{n(A)}{N} \quad .$$

Het voorwaardelijk frequentiequotiënt van B onder de voorwaarde A is per definitie het aantal elementen, dat zowel kenmerk A als kenmerk B bezit, gedeeld door n(A); notatie:

$$(2) \qquad fq_{\Lambda,A}(B) = \frac{n(A \wedge B)}{n(A)}$$

waarin "$A \wedge B$" gelezen wordt als "A en B".

Deze frequentiequotiënten bezitten een aantal eenvoudige eigenschappen, die ogenblikkelijk uit hun definitie volgen; wij noemen er slechts enkele:

(3) $0 \leqq f_q(A) \leqq 1$ voor iedere A.

(4) $f_q(A) = 0$ als A niet voorkomt.

(5) $f_q(A) = 1$ als ieder element van A kenmerk A bezit.

De axiomas van de waarschijnlijkheidsrekening zijn nu gekozen naar analogie van deze eigenschappen van frequentiequotiënten. Dit is, gezien de wijze van invoering van het begrip waarschijnlijkheid, een voor de hand liggende gang van zaken, waarvan men goede resultaten mag verwachten. Geven wij de waarschijnlijkheid van het optreden van een kenmerk A, binnen een gegeven model, aan met $P[A]$, voorwaardelijke waarschijnlijkheid van B onder de voorwaarde A met $P_A[B]$, het optreden van de kenmerken A en B met $A \wedge B$, en het optreden van minstens één der kenmerken A of B met $A \vee B$, dan gelden de volgende axiomas:

(6) $0 \leqq P[A] \leqq 1$ voor iedere A.

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

(8) $P[A \wedge B] = P[A] . P_A[B] = P[B] . P_B[A]$ (vermenigvuldigingsregel)

terwijl voorts naar analogie van (4) en (5) de onmogelijkheid van het optreden van A per definitie aangegeven wordt door

(9) $P[A] = 0$

en de zekerheid van het optreden van A door

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

De axiomas (6), (7) en (8) volgen, indien men "P" door "fq" vervangt, direct uit de definitie van frequentiequotiënt.

Een laatste axioma, dat een generalisatie van de optelregel voor een oneindig aantal elkaar uitsluitende kenmerken inhoudt, laten wij buiten beschouwing, daar wij het niet nodig zullen hebben. Wij beperken ons nl. tot het geval van een eindig aantal kenmerken. Indien men dit niet doet geven (9) en (10) ook niet meer precies onmogelijkheid en zekerheid aan; bij een eindig aantal kenmerken is dit wel het geval.

Indien de kenmerken A en B elkaar uitsluiten, d.w.z. indien hun gezamelijk optreden onmogelijk is, is volgens (9): $P[A \wedge B] = 0$,

zodat (7) overgaat in

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

Zowel (7) als (11) kunnen gemakkelijk tot meer dan twee kenmerken worden uitgebreid.

Twee kenmerken A en B heten <u>stochastisch onafhankelijk</u>, indien aan de relatie:

(12) $$P_A[B] = P[B]$$

voldaan is. In dat geval gaat (8) over in:

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

terwijl tevens uit (12) en (8) volgt, dat

(14) $$P_B[A] = P[A]$$

is, indien tenminste $P[B] \neq 0$ is. Deze onafhankelijkheidsdefinitie is het mathematische analogon van de reeds in 1.2 voor het model van het werpen met een munt met de woorden "dat het resultaat van iedere worp onafhankelijk is van het bij vroegere worpen verkregen resultaat" aangeduide eigenschap. De <u>onafhankelijkheid der worpen</u> betekent in dat geval, dat de kans op kruis bij iedere worp dezelfde is.

1.4. <u>Vertaling van een uitkomst</u>.

Het is van belang ons af te vragen, wat een uitkomst van de vorm

(15) $$P[K] = p$$

(de kans op kruis is gelijk aan p) nu voor de toepassing betekent. Daartoe moeten wij ons bedienen van het vertalingsprincipe <u>C</u> van § 1.1. Dit kan niet zonder meer worden toegepast, tenzij p zeer klein is, in welk geval dit principe zegt, dat bij één maal werpen met het optreden van K geen rekening gehouden behoeft te worden (is p groot, d.w.z. bijna gelijk aan 1, dan is de kans op munt, P[M], zeer klein, zodat <u>C</u> op analoge wijze kan worden toegepast). De vraag, bij welke waarde van p men precies moet beginnen rekening te houden met het optreden van K valt niet algemeen te beantwoorden. Dit hangt van allerlei factoren af, waarvan de belangrijkste wel is, welke gevolgen het optreden van K zou bezitten, indien daarmee geen rekening gehouden is. Ook een subjectief element is zeker aanwezig. Daar het principe <u>C</u> ons echter vanuit de zuivere wiskunde in de practijk verplaatst, mag dit geen bezwaar heten.

Willen wij het vertaalprincipe C ook op (15) toepassen, indien p niet in de buurt van 0 of 1 ligt, dan zullen wij eerst binnen het model een mathematisch procédé moeten uitvoeren, waardoor (15) in een andere formule van een dergelijke vorm wordt omgezet, waarin het rechterlid wèl een zeer kleine waarschijnlijkheid is. Wij zullen hier dit procédé niet uitvoeren, daar dit in een andere voordracht van deze cursus wordt behandeld, maar slechts het resultaat (de theoretische wet der grote getallen) vermelden, uitgedrukt in de terminologie van het door ons gekozen model:

Indien P[K]= p is, en de munt N maal geworpen wordt, nadert de kans, dat fq(K), het quotient van het aantal worpen.dat kruis oplevert, en het totale aantal worpen, meer dan een willekeurig klein positief getal ε van p verschilt, tot 0, indien men N onbegrensd laat aangroeien.

Hierop kan C worden toegepast; zelfs kan men de grootte van de kans, die men bereid is te verwaarlozen, van tevoren bepalen, evenals de grootte van ε . Zijn deze twee bepaald, dan kan een ondergrens voor N berekend worden en men kan dan zeggen: "Wordt de munt N maal geworpen, dan zal fq(K) liggen tussen $p-\varepsilon$ en $p+\varepsilon$. De kans, dat dit niet zo is, is zo klein, dat wij er geen rekening mee behoeven te houden".

Iets minder accuraat drukt men zich gewoonlijk als volgt uit: Indien P[K]= p is, zal onder een groot aantal (N) worpen, ongeveer N.p maal het kenmerk K optreden.

Wij wijzen erop, dat dit secundaire vertalingsprincipe uit C is afgeleid met behulp van het kruis en munt-model zelf. Dit is één van de belangrijkste toepassingen van dat model.

1.5. Programma.

Wij zullen in het volgende trachten na te gaan, hoe men conclusies kan trekken uit een gegeven waarnemingsresultaat, dat bestaat uit:

I. Het aantal maal, dat K opgetreden is bij n worpen van één munt; dit aantal geven wij aan met n_o.

II. Het aantal maal, dat K is opgetreden in twee onafhankelijke series van n resp. m worpen, met twee verschillende munten; deze aantallen geven wij aan met n_o resp. m_o.

Hierbij zullen wij ons tot de bespreking van een drietal methoden en conclusies moeten beperken.

2. Toetsing van de hypothese P[K] = p.

2.1. De verdeling van Bernoulli.

Indien men n maal werpt met een munt, waarbij bij iedere worp de kans op K gelijk aan p is, kan het aantal malen K, dat we met $\underline{x}$ aangeven, de waarden 0,1,2,...,n aannemen, terwijl ieder van deze waarden een bepaalde waarschijnlijkheid bezit. Men zegt in een dergelijk geval, dat $\underline{x}$ een waarschijnlijkheidsverdeling bezit en noemt $\underline{x}$ een stochastische variabele. Wij geven een dergelijke variabele aan door een onderstreepte letter, terwijl een door $\underline{x}$ aangenomen waarde x zonder onderstreping wordt aangegeven. De verdeling van $\underline{x}$, bij n worpen, wordt gegeven door

$$(16) \qquad \boxed{P[\underline{x} = x] = \binom{n}{x} p^x q^{n-x}} \qquad (\underline{\text{verdeling van Bernoulli}})^{1)}$$

Bewijs:

Dit kan op vele wijzen bewezen worden. Wij geven een bewijs door volledige inductie.

Voor n = 1 geldt: $P[\underline{x} = 0] = q$ en $P[\underline{x} = 1] = p$, zodat aan (16) voldaan is.

Is voor n = k aan (16) voldaan, dan is, indien wij het aantal trekkingen aangeven door een index bij het symbool P:

$$P_{k+1}[\underline{x} = x] = p.P_k[\underline{x} = x-1] + q.P_k[\underline{x} = x]$$

zoals uit (11) en (13) volgt. Dus is:

$$P_{k+1}[\underline{x} = x] = p.\binom{k}{x-1} p^{x-1} q^{k+1-x} + q.\binom{k}{x} p^x q^{k-x} =$$

$$= \left\{\binom{k}{x-1} + \binom{k}{x}\right\} p^x q^{k+1-x} = \binom{k+1}{x} p^x q^{k+1-x}$$

daar

$$\binom{k}{x-1} + \binom{k}{x} = \frac{k!}{(x-1)!(k+1-x)!} + \frac{k!}{x!(x-k)!} = \frac{(k+1)!}{x!(k+1-x)!} = \binom{k+1}{x}$$

is. Hiermee is de stelling bewezen.

---

1) Notatie: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$; $k! = 1.2.3. \ldots .k$; $0! = 1$.
De verdeling van Bernoulli wordt ook de binomiale verdeling genoemd.

Ter illustratie geven wij in fig. 1 een voorbeeld van een dergelijke verdeling voor n = 10 en p = 0,4; de waarden van $P[\underline{x} = x]$ vindt men in tabel I.

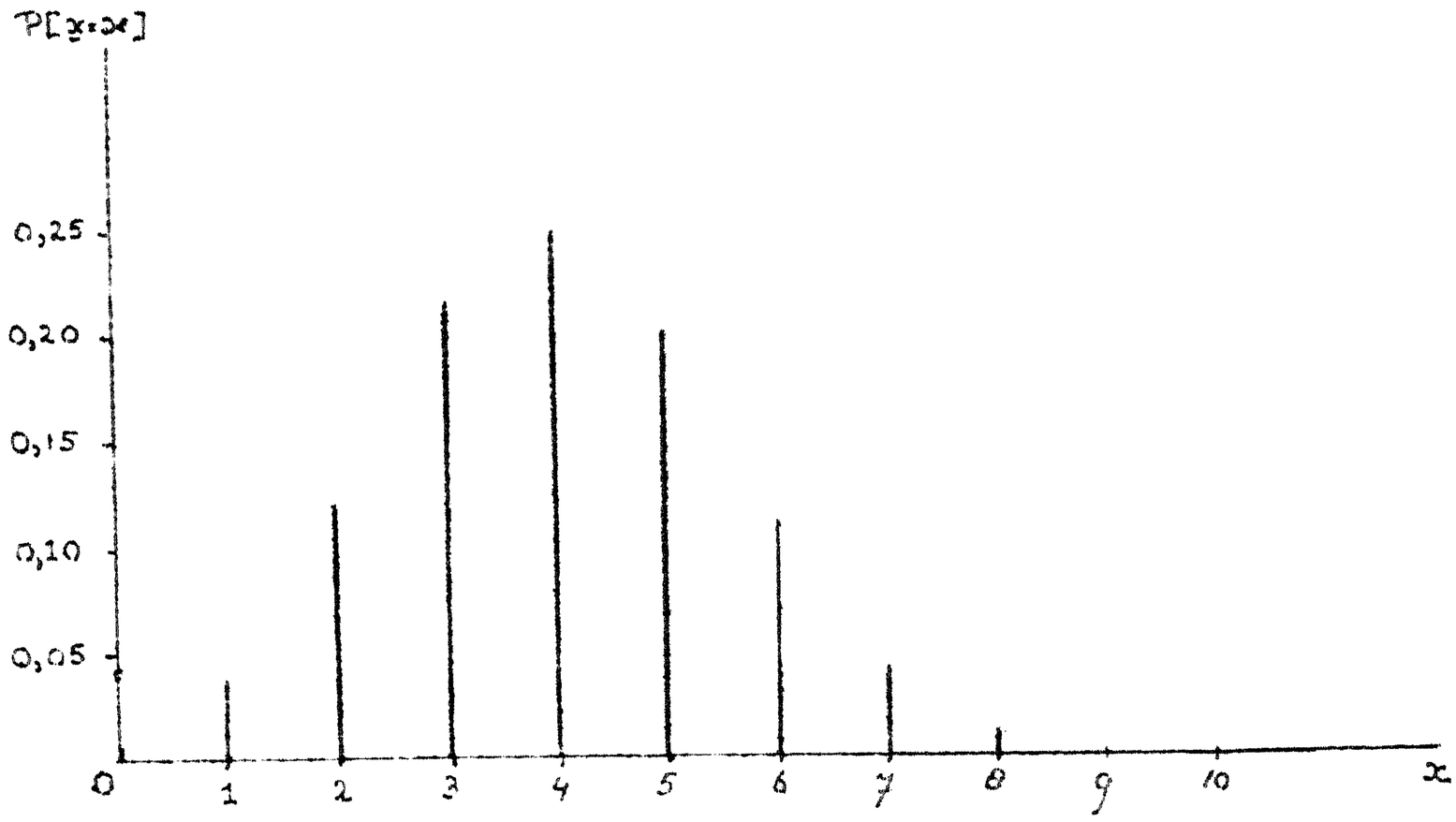

Fig. 1

TABEL I

Bernoulliverdeling voor n = 10; p = 0,4.

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

2.2. De toetsingsmethode.

Indien wij nu, op grond van de in 1.6.I beschreven gegevens, de hypothese willen toetsen, dat $P[K] = p$ is, waarin p een gegeven waarde bezit, gaan wij na, wat de verdeling van $\underline{x}$ zou zijn, indien deze hypothese juist is. Deze verdeling wordt gegeven door (16). Vervolgens berekenen wij $P[\underline{x} = n_o]$, waarin n_o de bij het experiment gevonden waarde van $\underline{x}$ is en zoeken al die waarden x bijeen, waarvoor

$$P[\underline{x} = x] \leq P[\underline{x} = n_o] \tag{17}$$

Wij berekenen dan

(18) $$\gamma = \sum P[\underline{x} = x]$$

waarbij over de genoemde waarden van x gesommeerd wordt. In woorden uitgedrukt, is γ de kans, dat, indien $P[K] = p$ is, bij een serie van n worpen een waarde van $\underline{x}$ gevonden zal worden, die even onwaarschijnlijk of nog onwaarschijnlijker is dan de gevonden waarde n_o (dat dit zo is volgt uit de optelregel in de vorm (11)). Is deze kans kleiner dan of gelijk aan een van tevoren vastgesteld positief getal α, dan verwerpen wij de hypothese $P[K] = p$; is $\gamma > \alpha$, dan wordt de hypothese niet verworpen. Hierbij dient te worden opgemerkt, dat "niet-verwerpen" niet aequivalent is met "aanvaarden". Indien de waarde p niet verworpen kan worden, zijn er ook steeds naburige waarden van p, die evenmin verworpen kunnen worden. Men noemt γ wel de <u>overschrijdingskans</u> van het gevonden resultaat met betrekking tot de te toetsen hypothese, en α de <u>onbetrouwbaarheidsdrempel</u> (Eng.: level of significance).

Indien b.v. in het geval van figuur 1 voor n_o de waarde 8 gevonden was, zou de overschrijdingskans, bij toetsing van de hypothese $P[K] = 0,4$, gelijk aan 0,019 zijn; dit is de som der waarschijnlijkheden $P[\underline{x} = x]$ met x = 0; 8; 9 en 10.

2.3. <u>Betekenis van deze methode</u>.

De betekenis van deze methode is daarin gelegen, dat men een bovengrens kent van de kans op het verwerpen van een <u>juiste</u> hypothese. Immers, stel dat de hypothese juist is, dus dat $P[K] = p$ is, dan is de kans, dat $\underline{x}$ bij n worpen een waarde aanneemt zo, dat $\gamma \leqq \alpha$ is, zelf $\leqq \alpha$. Daartoe moet de door $\underline{x}$ aangenomen waarde nl. behoren tot een verzameling Z van waarden, die tezamen een waarschijnlijkheid $\leqq \alpha$ bezitten, terwijl tevens voor iedere x, die tot Z behoort $P[\underline{x} = x]$ kleiner is dan voor iedere niet tot Z behorende waarde. Z wordt de <u>kritieke zône</u> genoemd (Eng.: critical region).

De keuze van α hangt van verschillende factoren af. Enerzijds zal men de gevolgen overwegen van het verwerpen van de hypothese, indien deze juist is, anderzijds moet men rekening houden met het (hier niet bewezen) feit, dat, naarmate α, (dus de kans op ten onrechte verwerpen van de hypothese, terwijl deze juist is) afneemt, de kans op terecht verwerpen van de hypothese, indien deze onjuist is, eveneens afneemt. Veelal neemt men $\alpha = 0,05$ of 0,01. Vertaald volgens 1.4 betekent dit, dat wij ongeveer 1 op de 20 resp. 1 op de 100 van de juiste hypothesen, die getoetst worden, zullen verwerpen.

2.4. Voorbeeld.

In 1930 werden in Amsterdam 6848 jongens en 6374 meisjes geboren ("De bevolking van Amsterdam", deel I: Loop der bevolking tot 1931, Amsterdam 1933, p. 47). Met behulp van bovenstaande methode kunnen wij nu de hypothese toetsen, dat de kans op een jongens-geboorte even groot is als die op een meisjes-geboorte, d.w.z. de hypothese $p = \frac{1}{2}$. Met behulp van een benaderingsmethode voor grote aantallen vinden wij, dat de overschrijdingskans van deze uitkomst ongeveer 0,0002 is. De kans op het gevonden resultaat of op een nog onwaarschijnlijker is dus slechts ongeveer 1/5000 Dit leidt tot verwerping van de getoetste hypothese ten gunste van de hypothese dat de kans op een jongens-geboorte groter is dan die op een meisjes-geboorte. Andere jaren en plaatsen bevestigen dit verschijnsel.

3. Schatting van P[K] door een betrouwbaarheidsinterval.

3.1. Schatting door één waarde.

Op grond van de gevonden waarde n_o van $\underline{x}$, bij n worpen kan men P[K] schatten op n_o/n . Deze schatting bezit bepaalde optimale eigenschappen, waaruit men kan besluiten, dat het een zeer bruikbare en zelfs wel de beste schatting is, die men, door het aangeven van één enkele waarde, kan verkrijgen. Het is echter duidelijk, dat n_o/n in het algemeen niet de juiste waarde van P[K] zal zijn; dit blijkt reeds uit de noemer n, die bepaald wordt op grond van feiten, die niets met P[K] te maken hebben. Bovendien, zelfs indien P[K] rationaal was en de noemer n bezat, is er nog slechts een vrij gering kans (die bij toenemende n afneemt) dat n_o/n precies de juiste waarde zal bezitten.

3.2. Definitie van een betrouwbaarheidsinterval.

Deze nadelen kan men ontgaan door niet slechts één waarde als schatting te geven, maar een interval J van waarden aan te geven, waar P[K] naar schatting in ligt. Dit interval J zal uiteraard afhankelijk zijn van het gevonden resultaat, dus van de gevonden waarde van $\underline{x}$. J is derhalve een stochastisch interval (d.w.z. de uiteinden ervan zijn stochastische variabelen); wij geven het daarom aan door $\underline{J}$ of $J(\underline{x})$. Het wordt nu zo bepaald (vgl. 3.3) dat

(19) $$P[\,P[K] \in \mathfrak{J}(\underline{x})\,] \geqq 1-\alpha$$ [2])

is, waarin α een van tevoren gekozen positief getal is. Ook nu heet α de onbetrouwbaarheidsdrempel (in de Engelse literatuur wordt $1-\alpha$ de "confidence level" genoemd). Het is van groot belang nadrukkelijk op te merken, dat in (19) niet $P[K]$, maar $\mathfrak{J}(\underline{x})$ stochastisch is. $P[K]$ is een onbekende, maar constante, waarschijnlijkheid.

Is nu bij een bepaald experiment de waarde n_o gevonden, dan betekent (19), vertaald met behulp van het in 1.4 beschreven vertalingsprincipe, dat de uitspraak "$P[K]$ ligt in $\mathfrak{J}(n_o)$", op deze wijze verkregen, ongeveer in een fractie α van de gevallen, waarin men de methode toepast, fout zal zijn. Evenals bij het toetsen van hypothesen neemt men α gewoonlijk 0,05 of 0,01, terwijl ook de overwegingen ter bepaling van α van soortgelijke aard zijn: enerzijds zijn de gevolgen van een foute conclusie van belang, anderzijds moet men overwegen, dat het betrouwbaarheidsinterval langer wordt, naarmate men α kleiner neemt.

3.3. Bepaling van het betrouwbaarheidsinterval.

Bij een gevonden waarde n_o van $\underline{x}$ in een serie van n worpen, nemen wij als betrouwbaarheidsinterval $\mathfrak{J}(n_o)$ de verzameling van die waarden p die, bij toetsing van de hypothese $P[K] = p$ met onbetrouwbaarheidsdrempel α, niet voor verwerping in aanmerking zouden komen. Het stochastische interval $\underline{\mathfrak{J}}$ wordt verkregen door in deze definitie $\underline{x}$ voor n_o te substitueren. Aangezien de toetsingsmethode de eigenschap bezit, dat de kans, dat de juiste waarde p voor verwerping aan aanmerking komt, $\leqq \alpha$ is, is de kans, dat deze juiste waarde niet verworpen wordt, dus in $\mathfrak{J}$ ligt, $\geqq 1-\alpha$.

Het principe der bepaling van $\underline{\mathfrak{J}}$ is hiermee gegeven. De verdere uitwerking heeft enige haken en ogen en ook de numerieke uitwerking ervan is niet zo eenvoudig, dat wij deze in korte tijd in details uiteen kunnen zetten. Wij laten het daarom bij bovenstaande methodologische beschouwingen, die het algemene verband tussen de toetsingstheorie en de theorie der betrouwbaarheidsintervallen aangeven.

---

[2]) $P[K] \in \mathfrak{J}(\underline{x})$ betekent: $P[K]$ ligt in het interval $\mathfrak{J}(\underline{x})$.

4. De vergelijking van twee waarschijnlijkheden.

4.1. Het probleem.

Het probleem, waarvan in deze paragraaf een oplossing besproken zal worden, is, uit de resultaten van twee onafhankelijke series worpen met twee munten A en B na te gaan, of de hypothese dat bij beide de kans op kruis gelijk is, verworpen kan worden. Indien dit het geval is, kan men bovendien gemakkelijk aangeven bij welke van de twee munten deze kans het grootst is.

Dit probleem kan zich bijvoorbeeld voordoen bij het bepalen van een keuze tussen twee geneesmethoden of twee productiemethoden.

Indien bij een serie van n worpen met munt A het aantal maal, dat K verkregen is, gelijk is aan n_o en de overeenkomstige aantallen bij B gelijk zijn aan m en m_o, kunnen wij deze gegevens samenvatten in de vorm van tabel II:

TABEL II

	K	M	totaal
A	n_o	$n-n_o$	n
B	m_o	$m-m_o$	m
totaal	r	s	N

waarin $n_o + m_o = r$, $n+m = N$ en $r+s = N$ is.

4.2. De toetsingsmethode.

Beschouwen wij de aantallen n en m als gegeven constanten dan is de algemene vorm van deze tabel in tabel III weergegeven.

TABEL III

	K	M	
A	$\underline{x}$	$n-\underline{x}$	n
B	$\underline{y}$	$m-\underline{y}$	m
	$\underline{r}$	$\underline{s}$	N

waarin $\underline{x}$, $\underline{y}$, $\underline{r}$ en $\underline{s}$ stochastisch zijn en de betrekkingen $\underline{x}+\underline{y}=\underline{r}$ en $\underline{r}+\underline{s} = N = \underline{m}+\underline{n}$ gelden. Geven wij de kans op K bij munt A en B resp. aan met p_1 en p_2, dan moet de hypothese $p_1 = p_2$ getoetst worden.

Bij de toetsing van deze hypothese wordt nu ondersteld, dat $\underline{r}$ de (in het experiment gevonden) waarde r (en derhalve $\underline{s}$ de waarde s) aanneemt. Een dergelijke toets wordt een voorwaardelijke toets genoemd. Onder alle mogelijke uitkomsten voor de stochastische variabelen in tabel III worden alleen die uitkomsten in de beschouwingen betrokken, waarbij $\underline{r} = r$ is. Dit lijkt misschien een enigszins merkwaardige procedure; in het volgende zal echter blijken, waartoe deze maatregel dient. Verder merken wij (zonder bewijs) op, dat de opgelegde voorwaarde geen wezenlijke beperking betekent: een voorwaardelijke toets als de hier besprokene kan steeds tot een onvoorwaardelijke worden herleid, zonder dat de toetsingsmethode in een gegeven geval (d.w.z. bij gegeven n, m en r) enige wijziging ondergaat.

Wij leiden nu, onder de hypothese, dat $p_1 = p_2$ is, de voorwaardelijke waarschijnlijkheidsverdeling van $\underline{x}$, onder de voorwaarde $\underline{r} = r$, af. Bij gegeven r en gegeven waarde x van $\underline{x}$ is de waarde y, die $\underline{y}$ aanneemt, bepaald door $y = r-x$.

Volgens (16) geldt dan, als we $p_1 = p_2 = p$ en $q = 1-p$ stellen:

$$P[\underline{x} = x] = \binom{n}{x} p^x q^{n-x}$$

en

$$P[\underline{y} = y] = \binom{m}{y} p^y q^{m-y}$$

dus volgens (13):

$$P[\underline{x} = x \wedge \underline{y} = y] = \binom{n}{x}\binom{m}{y} p^{x+y} q^{m+n-x-y}$$

Wegens de relatie $\underline{x} + \underline{y} = \underline{r}$ kunnen wij voor het linkerlid van deze vergelijking ook schrijven:

$$P[\underline{x} = x \wedge \underline{r} = r]$$

waarin $r = x + y$ is.

Verder is

$$P[\underline{r} = r] = \binom{N}{r} p^r q^{N-r}$$

Volgens de vermenigvuldigingsregel (8) vinden wij tenslotte voor de voorwaardelijke verdeling van $\underline{x}$:

$$P_{\underline{r}=r}[\underline{x}=x] = \frac{P[\underline{x}=x \wedge \underline{r}=r]}{P[\underline{r}=r]} = \frac{\binom{n}{x}\binom{m}{y}}{\binom{N}{r}} \quad \frac{p^{x+y} q^{m+n-x-y}}{p^r q^s}$$

dus

$$(20) \qquad \boxed{P_{\underline{r}=r}\,[\underline{x}=x] = \frac{\binom{n}{x}\binom{m}{r-x}}{\binom{m+n}{r}}}$$

Deze verdeling wordt een <u>hypergeometrische verdeling</u> genoemd. De onbekende parameter p is hieruit geëlimineerd met behulp van de relatie x + y = r. Dit is de reden voor het stellen van de voorwaarde $\underline{r} = r$.

Overigens verloopt nu de toetsingsmethode geheel als in 2.2. Wij volstaan daarom met het geven van (aan de practijk ontleend) voorbeeld. In de Engelse literatuur wordt tabel III een "<u>2 x 2 table</u>" genoemd, terwijl de onderscheiding naar 2 paren kenmerken (K en M resp. A en B) ook aangeduid wordt met de uitdrukking "<u>double dichotomy</u>".

4.3. <u>Voorbeeld</u>.

Ter vergelijking van de merites van twee geneesmiddelen voor een ziekte werden 11 patiënten met het eerste en 16 met het ~~tweede~~ behandeld. Vervolgens werd nagegaan, of de kwaal door het geneesmiddel aanmerkelijk werd verlicht. Het verkregen resultaat is in tabel IV samengevat:

<u>TABEL IV</u>.

	succes	geen succes	totaal
therapie I	6	5	11
therapie II	14	2	16
totaal	20	7	27

Wij hebben dus: $n_o = 6$, $n = 11$, $m = 16$, $r = 20$. Volgens (20) is dus

$$P_{\underline{t}=20}\,[\underline{x}=n_o] = \frac{\binom{11}{6}\binom{16}{14}}{\binom{27}{20}} = 0{,}0624$$

De waarden, die $\underline{x}$, d.i. het aantal met succes met therapie I behandelde patiënten kan aannemen, zijn: x = 4; 5;...; 11. Hieruit zoeken wij, met behulp van (20), die waarden, waarvoor de waarschijnlijkheid $\leq 0{,}0624$ is.

Dit zijn:

x	$P_{\gamma=10}[\underline{x} = x]$
4	0,0004
5	0,0083
6	0,0624
11	0,0130
	0,084

Voor de overige waarden van x is de waarschijnlijkheid $\geq$ 0,07. De overschrijdingskans is in dit geval dus gelijk aan 0,084, hetgeen te groot is, om te concluderen, dat de getoetste hypothese onjuist is. Het experiment heeft dus niet aangetoond, dat er een verschil in werking is tussen de twee therapieën; mede in verband met het betrekkelijk geringe aantal behandelde patiënten, kan men echter wel zeggen, dat er een aanwijzing is in de richting, dat therapie II beter werkt dan therapie I en dat een voortzetting van het onderzoek door de uitkomst gerechtvaardigd wordt. Indien er nl. een klein verschil tussen de twee therapieën is, zal men dit met weinig proefnemingen niet, maar met een groot aantal wel kunnen aantonen.

4.4. Opmerkingen.

Een duidelijke scheiding tussen het medische en het statistische deel van een dergelijk onderzoek is van groot belang. De medicus beslist over de aan- of afwezigheid van succes van een therapie bij iedere patiënt en stelt de gegevens pas in handen van de statisticus, als zij in de vorm van tabel IV kunnen worden gebracht. De statistische conclusie wordt gegeven in de vorm van een overschrijdingskans met de daarbij behorende beschouwingen. De generalisatie van de conclusie (als die er is) van de onderzochte groep patiënten tot een grotere groep is in het algemeen ook een kwestie van medische aard. Alleen als de groep, waarvoor men de conclusie wil laten gelden van tevoren is aangegeven en de te onderzoeken patiënten daaruit op een bepaalde wijze zijn uitgezocht, valt de generalisatie van de conclusie tot <u>deze</u> groep onder de competentie van de statisticus.

Bij een onderzoek met grotere aantallen wordt de boven beschreven methode te omslachtig. Men kan dan echter gebruik maken van benaderingsmethoden, die eenvoudig zijn en een ruim voldoende nauwkeurigheid bezitten.

5. Literatuur.

Het spreekt vanzelf, dat wij hier slechts de allereerste beginselen van de theorieën, waaraan onze voorbeelden ontleend zijn, hebben besproken. Ter orientatie vermelden wij hieronder enkele titels uit de literatuur omtrent deze methoden. De toetsingstheorie en de theorie der betrouwbaarheidsintervallen zijn afkomstig van J. Neyman en E.S. Pearson , terwijl de in § 4 besproken toets door R.A. Fisher ontwikkeld is.

Enige oorspronkelijke publicaties:

R.A. Fisher, Statistical methods for research workers, Oliver & Boyd, Edinburgh 1925.

J. Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Phil.Trans. A 236 (1937) p. 333.

J. Neyman, Basic ideas and some recent results of the theory of testing statistical hypotheses, Jrn.Roy.Stat.Soc. 105 (1942) p. 292-327.

Overzichten:

D. van Dantzig, Kadercursus mathematische statistiek, Mathematisch Centrum, Amsterdam 1947-1950, hoofdstuk 5.

M.G. Kendall, The advanced theory of Statistics I and II, London 1943, 1946.

A. Wald, Statistical inference, Notre Dame, Indiana 1942.

Een aantal artikelen omtrent verschillende methoden ter oplossing van het in § 4 besproken probleem en daarmee verwante problemen, is verschenen in de jaargang 34 (1947) van het tijdschrift Biometrika.